RANDOM PHASE APPROXIMATION AND BEYOND: FROM THEORY TO REALISTIC MATERIALS

Dario Rocca

To cite this version:

Dario Rocca. RANDOM PHASE APPROXIMATION AND BEYOND: FROM THEORY TO REALISTIC MATERIALS. Theoretical and/or physical chemistry. Université de Lorraine, 2020. tel-02979044

HAL Id: tel-02979044

https://hal.science/tel-02979044

Submitted on 26 Oct 2020

HABILITATION À DIRIGER DES RECHERCHES
Date de soutenance: 14 janvier 2020

Dario Rocca

UNIVERSITÉ DE LORRAINE ET CNRS

Laboratoire de Physique et Chimie Théoriques, LPCT

Institut Jean Barriol

RANDOM PHASE APPROXIMATION AND BEYOND: FROM THEORY TO REALISTIC MATERIALS

APPROXIMATION DE LA PHASE ALÉATOIRE ET AU-DELÀ : DE LA THÉORIE AU MATÉRIAUX RÉELS

Composition du jury

RAPPORTEURS

Dr. Julien Toulouse, LCT, Sorbonne Université & CNRS
Prof. Mébarek Alouani, IPCMS, Université de Strasbourg & CNRS
Dr. Lucia Reining, LSI, École Polytechnique & CNRS

EXAMINATEURS

Dr. Sébastien Lebègue, LPCT, Université de Lorraine & CNRS
Prof. Emilie Gaudry, IJL, Université de Lorraine & CNRS
Prof. Paolo Umari, Università degli Studi di Padova
Prof. Stefano de Gironcoli, Scuola Internazionale Superiore di Studi Avanzati

Abstract

Despite the high computational cost the adiabatic connection fluctuation dissipation theorem (ACFDT) represents a promising approach to improve the description of the electronic correlation within density functional theory. The simplest approximation that can be applied in the context of the ACFDT is the random phase approximation (RPA). First, we show how the RPA can be improved by introducing a kernel containing an approximate electron-hole exchange term that leads to two different beyond-RPA methods. Then, we show that the RPA and beyond-RPA approaches can be efficiently computed within a plane-wave basis set implementation by using dielectric eigenpotentials as a compact auxiliary basis set and the Lanczos algorithm. A series of applications to molecules and solids are presented to demonstrate the efficiency and accuracy of these approximations. Importantly, it will be shown that the highly accurate beyond-RPA methods can be scaled to treat molecular systems with one hundred electrons requiring a basis set with hundreds of thousands of plane-waves. Finally, it is shown how the sophisticated and computationally expensive ACFDT methods can be used to compute finite-temperature properties of realistic materials (adsorption enthalpies of molecules in zeolites) by coupling molecular-dynamics simulations with machine learning algorithms.

Résumé

Malgré un coût de calcul élevé, le théorème de fluctuation-dissipation avec connexion adiabatique (ACFDT) représente une approche prometteuse pour améliorer la description de la corrélation électronique en théorie de la fonctionnelle de la densité. L'approximation pratique la plus simple dans le contexte de l'ACFDT est l'approximation de la phase aléatoire (RPA). D'abord, nous montrons comment la RPA peut être améliorée en introduisant un noyau contenant un terme approximatif d'échange "électron-trou" qui conduit à deux méthodes différentes au-delà de la RPA. Ensuite, nous montrons que les approches RPA et au-delà de la RPA peuvent être calculées d'une façon efficace dans le cadre d'une implémentation basée sur une base d'ondes planes en utilisant les vecteurs propres de la matrice diélectrique comme une base auxiliaire compacte et l'algorithme de Lanczos. Une série d'applications à des molécules et à des solides sont présentées pour démontrer l'efficacité et la précision de ces approximations. De plus, il sera démontré que les méthodes très précises au-delà de la RPA peuvent traiter des systèmes moléculaires avec des centaines d'électrons de valence et nécessitant une base avec des centaines de milliers d'ondes planes. Enfin, nous montrons comment les méthodes ACFDT, très couteuse en termes de temps de calcul, peuvent être utilisées pour calculer les propriétés à température finie de matériaux réalistes (spécifiquement les enthalpies d'adsorption de molécules dans des zéolites)

en couplant des simulations de dynamique moléculaire à des algorithmes d'apprentissage automatique.

Contents

1 Introduction 3

2 Correlation energy within the random phase approximation and beyond 7

2.1 Density Functional Theory 8
2.2 Correlation energy within the adiabatic connection fluctuation-dissipation theorem 10
2.3 Random phase approximation 11
2.4 Electron-hole time-dependent Hartree-Fock for correlation energies 12
2.5 Adiabatic connection second order screened exchange approximation 14

3 Numerical implementation within a plane-wave basis set using iterative algorithms 16

3.1 Implementation using an auxiliary basis set 17
3.2 Elimination of the empty states 20
3.3 Lanczos algorithm 21
3.4 Construction of a compact auxiliary basis set by diagonalizing an approximate response function 23
3.5 Gram-Schmidt orthogonalization 27

4 Numerical applications to molecules and solids 32

4.1 Applications to weakly bound molecular dimers 33
  4.1.1 A24 Test Set 33
  4.1.2 S22 Test Set 35
4.2 Binding curves 38
4.3 Application to reaction energies involving covalent bonds 40
4.4 Applications to solids 46

5 Auxiliary basis set convergence 52
5.1 Convergence of energy differences vs. complete basis set extrapolation 52
5.2 Numerical examples 59
5.2.1 A24 Test Set 59
5.2.2 Methane-Formaldehyde Complex 61
5.2.3 Reaction energies 62
5.3 Conclusions 64

6 Applications to realistic systems by machine learning thermodynamic perturbation theory 67
6.1 Methodological approach: Machine learning thermodynamic perturbation theory 68
6.1.1 Thermodynamic perturbation theory 69
6.1.2 Machine Learning approach 71
6.1.3 Statistical error evaluation 73
6.2 RPA enthalpies of adsorption in zeolites 76
6.3 Conclusions 81

7 Perspectives 82

Appendix A Curriculum Vitæ 86

Appendix B Publications 100

Appendix C Résumé étendu en français 105

Bibliography 111

Chapter 1

Introduction

This manuscript describes my main (but not exclusive) research activities developed starting from September 2012, when I was hired as a Maître de Conférences at the Université de Lorraine. My work has largely focused on methodological development to compute correlation energies based on the random phase approximation (RPA) and its variants.

Because of its reasonable compromise between accuracy and efficiency density functional theory (DFT)[70, 77] has become the most widespread computational method in first principles materials science. DFT is applicable to relatively large systems of up to thousands of atoms, mainly owing its original success to rather simple approximations for the exchange-correlation (xc) functional, such as the local-density (LDA) [77, 108] and generalized-gradient approximations (GGA) [106]. More recent classes of functionals include meta-generalized gradient approximations [143, 140] and hybrid functionals, which mix a certain percentage of Hartree-Fock exchange with semi-local contributions [7]. All of these approximations, involving an increasing level of accuracy but also of computational complexity, can be pictorically depicted by the Jacob's ladder scheme of DFT approximations proposed by John Perdew (see Fig. 1.1) [107]. While several approximations for the the unknown xc functional have been proposed, research in this field is still quite active, with the ultimate goal of obtaining a "final theory". Challenging classes of problems for DFT approximations include materials where strong electronic interactions or van der Waals forces play an important role. Additionally, traditional approximate xc functionals do not systematically reach the chemical accuracy threshold (1 kcal/mol).

This manuscript will focus on the fifth and highest rung of the ladder of DFT functionals in Fig. 1.1. Specifically, the ground-state electronic correlation energy within the random phase approximation (RPA) will be considered. The RPA was already introduced in the 1950's to describe the correlation energy in the degenerate electron gas [14] and its diagrammatic structure has long

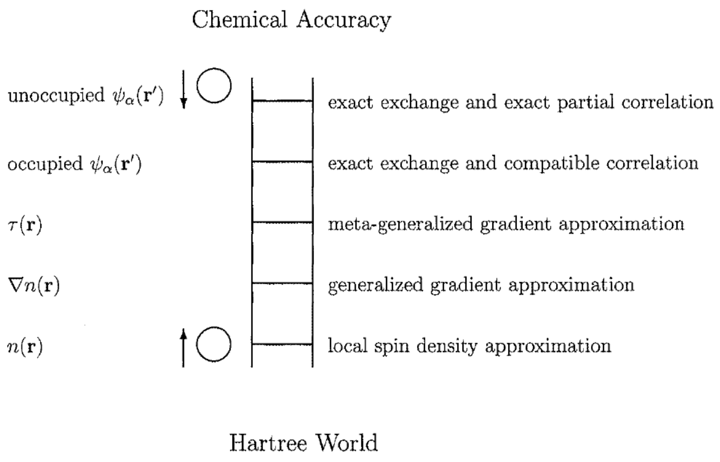


Figure 1.1: The Jacob's ladder of density functional approximations proposed by John Perdew to visualize the different levels of approximation in the exchange-correlation functionals. By climbing the ladder the accuracy increases at the expense of a higher computational cost (figure reproduced from Ref. 107).

been known (see Fig. 1.2) [45]. In the DFT context the theoretical basis for the RPA correlation energy is provided by the adiabatic-connection fluctuation-dissipation theorem (ACFDT) [80, 81], that will be introduced in Chapter 2. The ACFDT provides in principle an exact equation for the ground-state correlation energy in terms of linear response functions (i.e. the polarizability or the dielectric matrix). Within the ACFDT the random phase approximation (RPA) has provided a first practical scheme for applications to realistic systems, such as molecules [39, 144, 100, 153, 101, 35, 117, 124], solids [58, 85, 151, 60, 74], layered materials [92, 82], and molecules adsorbed on surfaces [118, 133]. In particular, for condensed matter/solid state applications the traditional quantum-chemical methods (e.g. the coupled-cluster theory) can be hardly applied and the RPA is probably the most sophisticated correlated approximation of practical interest. The rise of interest in the RPA is mostly, but not exclusively, related to the capability of the RPA to accurately include van der Waals (vdW) forces in a seamless way [32, 41]. Despite its growing success, the RPA is still a niche approach, especially if compared to other DFT approximations routinely applied in materials simulations. This is related to different reasons: (1) While being a highly sophisticated theory, shortcomings remain within the RPA: Binding curves of certain diatomic dimers present an unphysical "bump" at intermediate distances [39], total correlation energies are usually overestimated [78], and binding energies are often underestimated [41, 59]. (2) Implementations to compute the RPA correlation energies are rather complex and practical calculations involve several numer-

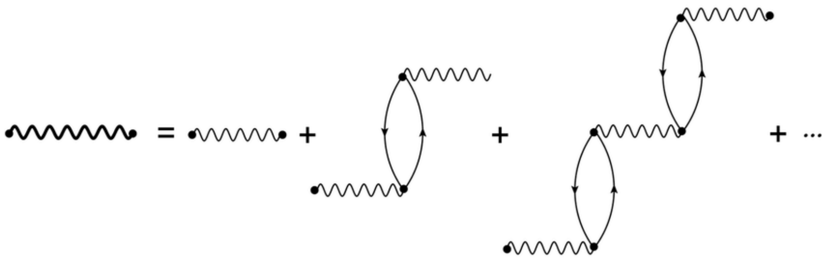


Figure 1.2: The effective interaction within the RPA is obtained as the sum to infinite order of ring diagrams (figure reproduced from Ref. 75).

ical bottlenecks, such as the full diagonalization of the starting-point Hamiltonian to obtain all the virtual (conduction) states, the manipulation of large response function matrices, and the evaluation of improper integrals over the frequency. (3) The computational time involved in RPA calculations is often several orders of magnitude higher than the time required by basic (semi-)local approximations; additionally, implementations to compute analytic interatomic forces are not trivial. For these reasons applications to realistic materials are highly demanding in terms of computational resources and typically based on frozen geometries optimized using different theories.

In this manuscript I will address these different issues. While a series of methods have already been proposed to improve the RPA in the literature [104, 84, 42, 120, 66, 25, 63, 6], in Chapter 2 I will discuss my work on the development of new approximations that avoid the self-correlation problem of the RPA by using a non-local exchange kernel. In Chapter 3 I will show how the previously introduced RPA and beyond-RPA methods can be efficiently implemented in an *ab initio* solid state code based on plane-waves by using a compact auxiliary basis set, the Lanczos algorithm, and density functional perturbation theory techniques. In Chapter 4 this numerical implementation is applied to a series of benchmark molecular test sets and to bulk solids. Chapter 5 will discuss in detail the (slow) numerical convergence of the (beyond-)RPA correlation energy with respect to the basis set size; this is a well known issue for quantum chemical methods that instead does not manifest itself when DFT is employed with more traditional (semi-)local or hybrid functionals. In Chapter 6 I address the issue of the applicability of RPA-based methods to realistic materials including temperature effects. In this context an impressive level of efficiency and accuracy is achieved by coupling machine learning algorithms with thermodynamic perturbation theory to provide the MLPT method. Applications of MLPT to compute molecular enthalpies of adsorption in zeolites shows the feasibility and potential impact of this approach. The future perspectives of this work are presented in Chapter 7. In particular, future research will further address the development of MLPT and, more in general, the development of techniques to bridge highly sophisticated methods for the electronic correlation with applications to realistic materials.

Chapter 2

Correlation energy within the random phase approximation and beyond

This chapter is based on original work previously published in Refs. [99, 28, 31].

In this chapter we will discuss the theoretical formalism which is at the base of this manuscript. After a short presentation of the main ideas behind density functional theory (DFT), the adiabatic connection fluctuation dissipation theorem (ACFDT) is introduced. In principle, this approach provides an exact equation for the exchange-correlation (xc) energy functional of DFT in terms of response functions (i.e. the polarizability for molecules and the dielectric matrix for solids). In practice, approximations for the response functions are necessary and, in this context, the random phase approximation (RPA) is probably the simplest practical approach. The RPA provides a fully non-local xc functional which accurately includes van der Waals forces in a seamless way [32, 41]. Still, the RPA is characterized by some issues: Binding curves of certain diatomic dimers present an unphysical "bump" at intermediate distances [39], total correlation energies are usually overestimated [78], and binding energies are often underestimated [41, 59]. With the aim of decreasing the self-correlation of the RPA and alleviating these issues we introduce two new approximations that include exact exchange effects: The electron-hole time-dependent Hartree-Fock (eh-TDHF) and the adiabatic connection second order screened exchange approximation (AC-SOSEX). The eh-TDHF represents a fully original contribution of my work and the AC-SOSEX was not previously used in a dielectric matrix formulation. The accuracy of these RPA and beyond-RPA approaches will be established by the numerical applications in Chapter 4.

Before starting the discussion a few technical notes are useful to follow the development of the formalism:

- Atomic units will be used throughout this manuscript.
- The formalism will be developed in a spin-restricted way and the orbitals ϕ are considered as space orbitals.
- For the sake of simplicity, we will suppose that all the orbitals ϕ are real; this is a general assumption in the case of molecular systems with time-reversal symmetry. For the RPA the formalism has also been implemented for periodic systems by including the integration over the first Brillouin zone [74].
- The subscripts $_v$ will be used to denote valence (namely occupied) states and the subscripts $_c$ will be used to denote conduction (namely virtual or unoccupied) states.

2.1 Density Functional Theory

Density Functional Theory has become impressively popular in the condensed matter physics and quantum chemistry communities because of its good compromise between accuracy and computational cost. The main (formidable) achievement of this approach is the mapping of a many-body electronic problem onto a non-interacting one. DFT was developed in the work of Hohenberg and Kohn [70] and in the work of Kohn and Sham [77]. Hohenberg and Kohn (HK) established the two theorems which constitute the theoretical foundation of DFT:

- **Theorem I**: For any system of interacting electrons in an external potential $V_{\text{ext}}(\mathbf{r})$, there is a one-to-one (apart from an irrelevant additive constant) correspondence between $V_{\text{ext}}(\mathbf{r})$ and the ground-state particle density $n_0(\mathbf{r})$;
- **Theorem II**: The energy of an interacting-electron system in an external potential $V_{\text{ext}}(\mathbf{r})$ can be expressed as a functional of the electronic density $n(\mathbf{r})$ of the form

  $$E_{\text{HK}}[n] = F_{\text{HK}}[n] + \int V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r}, \tag{2.1}$$

  where $F_{\text{HK}}[n]$ is a universal functional of the density which does not depend on $V_{\text{ext}}(\mathbf{r})$. Furthermore, the minimum value of $E_{\text{HK}}[n]$ is reached for $n = n_0$ and corresponds to the exact ground-state energy of the system.

The meaning of the first theorem is that the density completely determines all the properties of a given many body system. The second theorem is simply a corollary of the first and of the variational principle of quantum mechanics. The HK theorems provide a quite general theoretical result but they do not contain any practical recipe on how to solve the quantum many-body problem in practice.

A practical formulation of DFT is provided by the Kohn-Sham (KS) ansatz, which results in rewriting Eq. (2.1) as:

$$E_{\text{HK}}[n] = T[n] + E_{\text{H}}[n] + E_{\text{xc}}[n] + \int V_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \tag{2.2}$$

In Eq. (2.2) T is the kinetic energy, $E_{\text{H}}[n]$ the Hartree energy, and E_{xc} is defined by Eq. (2.2) as the difference $F_{\text{HK}}[n] - T[n] - E_{\text{H}}[n]$. In order to apply DFT in practice, a good approximation for E_{xc} is necessary. In the Introduction we briefly mentioned some popular approaches, such as the local-density approximations (LDA) [77, 108], the generalized-gradient approximation (GGA) [106], the meta-GGA [143, 140], and the hybrid functionals [7]; in the next section we will introduce an in principle exact formula for E_{xc} that can be used in practice to develop new approximate xc functionals. By applying the variational principle of Theorem II to the HK energy functional in the form of Eq. (2.2) the Euler equation is obtained:

$$\frac{\delta T[n]}{\delta n(\mathbf{r})} + V_{\text{KS}}(\mathbf{r}) = \mu, \tag{2.3}$$

where

$$\begin{aligned} V_{\text{KS}}(\mathbf{r}) &= V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) + V_{\text{ext}}(\mathbf{r}) \qquad (2.4) \\ &= \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})} + V_{\text{ext}}(\mathbf{r}), \qquad (2.5) \end{aligned}$$

and μ is a Lagrange multiplier that enforces the conservation of the total number of particles. In this way the ground-state density of the interacting-electron system $n_0(\mathbf{r})$ can be obtained by solving the Schrödinger equation of a fictitious non-interacting system:

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{KS}}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}). \tag{2.6}$$

Indeed, the one-particle orbitals ϕ_i can be then used to construct the density

$$n(\mathbf{r}) = \sum_i^{N_v} \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}), \tag{2.7}$$

where N_v is the number of occupied states. Eq. (2.6) is named Kohn-Sham (KS) equation. Since V_{KS} is a functional of the exact ground-state density, Eq. (2.6) has to be solved self-consistently. It is worth to note that, strictly speaking, the eigenvalues ϵ_i and eigenvectors ϕ_i do not have any physical meaning, but they are just mathematical devices used to obtain the ground-state energy and charge density of the full many-body system. Within this approach it is possible to calculate many ground-state properties, such as atomic forces, equilibrium geometries, stress tensors and phonon dispersion curves.

2.2 Correlation energy within the adiabatic connection fluctuation-dissipation theorem

In this section we discuss the exact expression of the exchange-correlation energy E_{xc} that can be obtained using the adiabatic connection fluctuation-dissipation theorem (ACFDT) [80, 81].

In the context of density functional theory the starting point to develop the ACFDT formalism is the adiabatic connection Hamiltonian for an electronic system

$$\hat{H}^{\lambda} = \hat{T} + \hat{V}_{\mathrm{ext}}^{\lambda} + \lambda \hat{V}_{\mathrm{ee}}, \tag{2.8}$$

where

$$\hat{V}_{\mathrm{ee}} = \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{2.9}$$

contains the contribution from the electron-electron Coulomb interaction and the parameter λ can assume values between 0 and 1. When $\lambda = 0$ the electrons are not interacting and the corresponding system is the Kohn-Sham system with $\hat{V}_{\mathrm{ext}}^{\lambda=0} = \hat{V}_{\mathrm{ext}} + \hat{V}_{\mathrm{H}} + \hat{V}_{\mathrm{xc}} = \hat{V}_{\mathrm{KS}}$. When $\lambda = 1$ the full electron interaction $\hat{V}_{\mathrm{ee}}$ is turned on and $\hat{V}_{\mathrm{ext}}^{\lambda=1} = \hat{V}_{\mathrm{ext}}$. Because of the particular form of $\hat{H}^{\lambda}$ the density $n(\mathbf{r})$ remains constant by switching on the interaction going from $\lambda = 0$ to $\lambda = 1$; this introduces some simplifications in the the ACFDT formalism.

By using the Hellmann-Feynman theorem and by integrating over the parameter λ it can be shown that

$$E_{\mathrm{H}} + E_{\mathrm{xc}}^{\mathrm{ACFDT}} = \int_0^1 d\lambda \frac{d}{d\lambda} (\langle \psi^{\lambda} | \hat{H}^{\lambda} | \psi^{\lambda} \rangle - \langle \psi^{\lambda} | \hat{V}_{\mathrm{ext}}^{\lambda} | \psi^{\lambda} \rangle) \tag{2.10}$$

$$= \int_0^1 d\lambda \langle \psi^{\lambda} | \hat{V}_{\mathrm{ee}} | \psi^{\lambda} \rangle. \tag{2.11}$$

The expression in Eq. 2.11 can then be recast in terms of density fluctuation operators. By using the fluctuation-dissipation theorem we can arrive to an expression of the correlation energy in terms of response functions $\widetilde{\Pi}$ evaluated on the complex frequency axis $i\omega$:

$$E_{\mathrm{c}}^{\mathrm{ACFDT}} = -\frac{1}{2}\int_0^1 d\lambda \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\ \mathrm{Tr}\left\{\widetilde{\mathbf{\Pi}}^{\lambda}(i\omega)\,\widetilde{\boldsymbol{K}} - \widetilde{\mathbf{\Pi}}^{0}(i\omega)\,\widetilde{\boldsymbol{K}}\right\}. \tag{2.12}$$

The exchange term $E_{\mathrm{x}}^{\mathrm{ACFDT}}$ is formally equivalent to the Hartree-Fock exact-exchange but is evaluated using KS orbitals.

In the definition of $E_{\mathrm{c}}^{\mathrm{ACFDT}}$ the matrix $\widetilde{\boldsymbol{K}}$ is constituted by four identical blocks $\boldsymbol{K}$:

$$\widetilde{\boldsymbol{K}} = \begin{pmatrix} \boldsymbol{K} & \boldsymbol{K} \\ \boldsymbol{K} & \boldsymbol{K} \end{pmatrix}, \tag{2.13}$$

where the matrix elements

$$K_{vc,v'c'} = 2\langle \phi_v \phi_c | V | \phi_{v'} \phi_{c'} \rangle = 2\int \phi_v(\mathbf{r})\phi_c(\mathbf{r})V(\mathbf{r}-\mathbf{r}')\phi_{v'}(\mathbf{r}')\phi_{c'}(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{2.14}$$

correspond to the electron-hole representation of the Coulomb potential V; the independent electron polarizability $\widetilde{\mathbf{\Pi}}^0$ is given by the equation

$$\widetilde{\mathbf{\Pi}}^0(i\omega) = \begin{pmatrix} -(\mathbf{\Delta} - i\omega \boldsymbol{I}) & \mathbf{0} \\ \mathbf{0} & -(\mathbf{\Delta} + i\omega \boldsymbol{I}) \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{\Pi}_{+}^{0}(i\omega) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}_{-}^{0}(i\omega) \end{pmatrix}, \tag{2.15}$$

where $\boldsymbol{I}$ denotes the identity matrix and $\Delta_{vc,v'c'} = (\epsilon_c - \epsilon_v)\delta_{vv'}\delta_{cc'}$ contains the KS independent electron excitations, defined as differences between conduction state energies ϵ_c and valence state energies ϵ_v. In Eqs. 2.12-2.15 as well as in the rest of this manuscript the bold letters are used to denote matrices. An additional tilde $\widetilde{}$ has been added on top of $2N_{eh} \times 2N_{eh}$ matrices (where $N_{eh} = N_v N_c$ is equal to the product of the number of valence N_v and conduction N_c states) to distinguish them from the $N_{eh} \times N_{eh}$ blocks; this notation will be useful to avoid confusion in the equations of the following sections.

2.3 Random phase approximation

In Eq. 2.12, $\widetilde{\mathbf{\Pi}}^{\lambda}$ refers to the polarizability of a fictitious system where the electrons interact with a scaled Coulomb potential λV. This term is unknown, and approximations need to be made to compute the correlation energy expression. The interacting polarizability under the RPA is defined as:

$$\widetilde{\mathbf{\Pi}}_{\mathrm{RPA}}^{\lambda}(i\omega) = \left(\widetilde{\boldsymbol{I}} - \lambda\widetilde{\mathbf{\Pi}}^{0}(i\omega)\widetilde{\boldsymbol{K}}\right)^{-1}\widetilde{\mathbf{\Pi}}^{0}(i\omega), \tag{2.16}$$

which when plugged in Eq. 2.12 gives us the RPA correlation energy after integration over λ and simplification of the trace over the 4×4 block matrices:

$$E_c^{\mathrm{RPA}} = \frac{1}{2} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \,\mathrm{Tr}\left[\ln\left\{\boldsymbol{I} - \boldsymbol{\chi}^0(i\omega)\boldsymbol{K}\right\} + \boldsymbol{\chi}^0\boldsymbol{K}\right], \tag{2.17}$$

where we introduced

$$\boldsymbol{\chi}^0(i\omega) = \mathbf{\Pi}^0_+ + \mathbf{\Pi}^0_- = 2\Re\left(\frac{1}{i\omega + \epsilon_v - \epsilon_c}\right)\delta_{vv'}\delta_{cc'} \tag{2.18}$$

and $\Re$ represents the real part. The $\boldsymbol{I} - \boldsymbol{\chi}^0(i\omega)\boldsymbol{K}$ term is commonly known as dielectric matrix and will be denoted by ε. The absence of $\;\widetilde{}\;$ over the matrices in Eq. 2.17 indicates that the dimensionality has been reduced to $N_{eh} \times N_{eh}$. The expression of the correlation energy in Eq. 2.17 is often referred as direct-RPA (or d-RPA) in the quantum chemistry community to emphasize the absence of exchange effects.

2.4 Electron-hole time-dependent Hartree-Fock for correlation energies

In this section the response function will be approximated at the time-dependent Hartree-Fock (TDHF) level of theory [93, 16]:

$$\widetilde{\mathbf{\Pi}}^{\lambda}(i\omega) = (\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\mathbf{\Pi}}^0(i\omega)\,\widetilde{\Xi})^{-1}\widetilde{\mathbf{\Pi}}^0(i\omega), \tag{2.19}$$

where the kernel $\widetilde{\Xi}$ is defined as

$$\widetilde{\Xi} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{pmatrix}, \tag{2.20}$$

with

$$A_{vc,v'c'} = 2\langle\phi_v\phi_c|V|\phi_{v'}\phi_{c'}\rangle - \langle\phi_v\phi_{v'}|V|\phi_c\phi_{c'}\rangle \tag{2.21}$$

$$B_{vc,v'c'} = 2\langle\phi_v\phi_c|V|\phi_{v'}\phi_{c'}\rangle - \langle\phi_v\phi_{c'}|V|\phi_{v'}\phi_c\rangle. \tag{2.22}$$

In principle, TDHF denotes an approach developed from a Hartree-Fock reference ground-state. In practice, the approximations used in this manuscript will instead use a DFT reference state, which can be efficiently obtained within a PW basis set implementation. A HF ground-state could be also used as a starting point, by rederiving the formalism using the ACFDT based on a HF reference [34].

In order to simplify the formalism we introduce an approximation in the full TDHF response function. This approximation was introduced in Ref. 99 and shown to lead to very accurate results. Specifically, the kernel $\widetilde{\Xi}$ in Eq. 2.20 can be divided in two different components

$$\widetilde{\Xi} = \widetilde{\boldsymbol{B}} + \left(\widetilde{\Xi} - \widetilde{\boldsymbol{B}}\right) = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{B} \end{pmatrix} + \begin{pmatrix} \boldsymbol{A}-\boldsymbol{B} & 0 \\ 0 & \boldsymbol{A}-\boldsymbol{B}. \end{pmatrix} \tag{2.23}$$

and the identity

$$\begin{aligned} \left(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\Xi}\right)^{-1} &= \left(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\boldsymbol{B}}\right)^{-1} \\ &+ \lambda\left[\widetilde{\boldsymbol{I}} - \lambda(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\boldsymbol{B}})^{-1}\widetilde{\boldsymbol{\Pi}}^0(i\omega)\left(\widetilde{\Xi} - \widetilde{\boldsymbol{B}}\right)\right]^{-1} \times \\ &\times \left(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\boldsymbol{B}}\right)^{-1}\widetilde{\boldsymbol{\Pi}}^0(i\omega)\left(\widetilde{\Xi} - \widetilde{\boldsymbol{B}}\right)\left(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\boldsymbol{B}}\right)^{-1} \end{aligned} \tag{2.24}$$

can be introduced. By discarding the last term in Eq. 2.24 a TDHF equation with an approximate kernel is obtained:

$$\widetilde{\boldsymbol{\Pi}}^\lambda(i\omega) = (\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\boldsymbol{B}})^{-1}\widetilde{\boldsymbol{\Pi}}^0(i\omega); \tag{2.25}$$

this equation involves exclusively the evaluation of products of valence and conduction states (conduction-conduction and valence-valence products are contained only in the $\boldsymbol{A}$ matrix) and will be denoted as electron-hole time-dependent Hartree-Fock (eh-TDHF). In Ref. 99 it was shown that the eh-TDHF approximation leads to significantly more accurate correlation energies with respect to the RPA. Additional approximations based on the eh kernel (Eq. 2.25) have also been proposed in the literature. In Ref. 66, Heßelmann and Görling relied on this approximation to develop a local exact-exchange kernel and showed that the electron-hole kernel is equivalent to the full kernel for a two electron system. In Ref. 6, Bates and Furche used the $\widetilde{\boldsymbol{B}}$ kernel to develop the RPA renormalized many-body perturbation theory. In the context of the plane-wave basis set Eq. 2.25 presents important practical advantages with respect to Eq. 2.19 (see Chapter 3 for details).

It is important to stress that the discussion in this section is specifically meant for ground state correlation energies. Indeed, the $\widetilde{\boldsymbol{B}}$ kernel and the corresponding response function in Eq. 2.25 are good approximations within the ACFDT framework and are not intended, for example, to compute optical absorption spectra. For this kind of applications the $\widetilde{\boldsymbol{B}}$ kernel does not contain the contributions necessary to correctly describe the physics of charge-transfer excitations in molecules [127] and of excitons in solids [1, 105]. However, this approximate kernel would satisfy the f-sum rule for any possible reference state used to build the response function [89, 145, 128]. Since the ACFDT requires an integral between $-\infty$ and ∞ of the response functions (see Eq. 2.12), the fulfillment of the f-sum rule might play a role in making this kernel accurate for correlation energies.

We can now replace the eh-TDHF response function (Eq. 2.25) in the ACFDT equation for the correlation energy (Eq. 2.12). Because of the simplified structure of the electron-hole kernel $\widetilde{\boldsymbol{B}}$, the integration over the parameter λ can be carried out analytically and the trace can be expressed in terms of the $N_{eh} \times N_{eh}$ blocks only:

$$E_{\mathrm{c}}^{\text{eh-TDHF}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\mathrm{Tr}\left\{\ln\big(\boldsymbol{I} - \boldsymbol{\chi}^0(i\omega)\boldsymbol{B}\big)\,\boldsymbol{B}^{-1}\boldsymbol{K} + \boldsymbol{\chi}^0(i\omega)\boldsymbol{K}\right\}, \qquad (2.26)$$

where $\boldsymbol{\chi}^0(i\omega)$ has been defined in Eq. 2.18. If the exchange component of $\boldsymbol{B}$ in Eq. 2.26 is set to 0 (namely $\boldsymbol{B} = \boldsymbol{K}$) the direct RPA approximation is found. By considering the Taylor expansion of the logarithm in Eq. 2.26 to the second order the MP2 approximation is obtained [30]:

$$\begin{aligned} E_c^{\mathrm{MP2}} &= -\frac{1}{4}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\mathrm{Tr}\left\{\boldsymbol{\chi}^0(i\omega)\boldsymbol{B}\boldsymbol{\chi}^0(i\omega)\boldsymbol{K}\right\} \\ &= -\frac{1}{2}\sum_{vc,v'c'}\frac{K_{vc,v'c'}B_{v'c',vc}}{\epsilon_c + \epsilon_{c'} - \epsilon_v - \epsilon_{v'}}. \end{aligned} \qquad (2.27)$$

As already discussed in Ref. 99, the Taylor expansion of Eq. 2.26 is convergent only when all the eigenvalues of $\boldsymbol{\chi}^0(i\omega)\boldsymbol{B}$ are smaller than 1. Nevertheless, the expression in Eq. 2.27 is mathematically equivalent to MP2 and represents a response function reformulation of this approximation [30].

2.5 Adiabatic connection second order screened exchange approximation

The adiabatic connection second order screened exchange (AC-SOSEX) approximation has been first proposed in Ref. 2 and formulated within a dielectric matrix framework in Ref. 99. Within the AC-SOSEX approximation the correlation energy is expressed as

$$E_{\mathrm{c}}^{\text{AC-SOSEX}} = -\frac{1}{2}\int_0^1 d\lambda \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\mathrm{Tr}\left\{\widetilde{\boldsymbol{\Pi}}^{\lambda}_{RPA}(i\omega)\,\widetilde{\boldsymbol{\Xi}} - \widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\boldsymbol{\Xi}}\right\}. \qquad (2.28)$$

where $\widetilde{\boldsymbol{\Xi}}$ is defined by Eqs.2.20-2.22 and $\widetilde{\boldsymbol{\Pi}}^{\lambda}$ is obtained within the direct RPA:

$$\widetilde{\boldsymbol{\Pi}}^{\lambda}_{RPA}(i\omega) = (\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^0(i\omega)\,\widetilde{\boldsymbol{K}})^{-1}\widetilde{\boldsymbol{\Pi}}^0(i\omega). \qquad (2.29)$$

In Eq. 2.28 the antisymmetrization of the correlation energy, which is missing in the direct RPA, is introduced by the exchange part of $\widetilde{\boldsymbol{\Xi}}$. The trace in Eq. 2.28 can be more conveniently expressed as

$$\begin{aligned}
& \mathrm{Tr}\left\{\widetilde{\boldsymbol{\Pi}}^{\lambda}_{RPA}\,\widetilde{\widetilde{\boldsymbol{\Xi}}} - \widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\widetilde{\boldsymbol{\Xi}}}\right\} \\
= \; & \lambda\mathrm{Tr}\left\{(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}})^{-1}\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}}\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{B}}\right\} \\
& +\lambda\mathrm{Tr}\left\{(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}})^{-1}\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}}\widetilde{\boldsymbol{\Pi}}^{0}\,\left(\widetilde{\widetilde{\boldsymbol{\Xi}}} - \widetilde{\boldsymbol{B}}\right)\right\}, \qquad (2.30)
\end{aligned}$$

where in the last line $\widetilde{\widetilde{\Xi}}$ has been split as in Eq. 2.23 and the dependence on $i\omega$ is kept implicit. By exploiting the properties of the trace operator and matrix algebra, the two contributions to the trace in Eq. 2.30 can be rewritten in terms of reduced size $N_{eh} \times N_{eh}$ matrices:

$$\begin{aligned}
& \lambda\mathrm{Tr}\left\{(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}})^{-1}\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}}\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{B}}\right\} \\
= \; & \lambda\mathrm{Tr}\left\{(\boldsymbol{I} - \lambda\,\boldsymbol{\chi}^{0}\,\boldsymbol{K})^{-1}\boldsymbol{\chi}^{0}\,\boldsymbol{K}\boldsymbol{\chi}^{0}\,\boldsymbol{B}\right\} \qquad (2.31)
\end{aligned}$$

and

$$\begin{aligned}
& \lambda\mathrm{Tr}\left\{(\widetilde{\boldsymbol{I}} - \lambda\,\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}})^{-1}\widetilde{\boldsymbol{\Pi}}^{0}\,\widetilde{\boldsymbol{K}}\widetilde{\boldsymbol{\Pi}}^{0}\,\left(\widetilde{\widetilde{\boldsymbol{\Xi}}} - \widetilde{\boldsymbol{B}}\right)\right\} \\
= \; & \lambda\mathrm{Tr}\left\{(\boldsymbol{I} - \lambda\boldsymbol{\chi}^{0}\boldsymbol{K})^{-1}(\boldsymbol{\Pi}^{0}_{+}\boldsymbol{K}\boldsymbol{\Pi}^{0}_{+} + \boldsymbol{\Pi}^{0}_{-}\boldsymbol{K}\boldsymbol{\Pi}^{0}_{-})(\boldsymbol{A} - \boldsymbol{B})\right\}. \qquad (2.32)
\end{aligned}$$

In the following we will show that the contribution to the correlation energy coming from Eq. 2.32 is likely to be small. By using the matrix identity

$$\left(\boldsymbol{I} - \lambda\,\boldsymbol{\chi}^{0}\,\boldsymbol{K}\right)^{-1} = \boldsymbol{I} + \lambda\left(\boldsymbol{I} - \lambda\,\boldsymbol{\chi}^{0}\,\boldsymbol{K}\right)^{-1}\boldsymbol{\chi}^{0}\,\boldsymbol{K} \qquad (2.33)$$

we notice that the inverse of $\boldsymbol{I} - \lambda\,\boldsymbol{\chi}^{0}\,\boldsymbol{K}$ is given by the sum of the identity matrix and a smaller contribution. Indeed, since the eigenvalues of $-\boldsymbol{\chi}^{0}\boldsymbol{K}$ are positive and decay to 0 rapidly [150], all the eigenvalues of $\lambda(\boldsymbol{I} - \lambda\,\boldsymbol{\chi}^{0}\,\boldsymbol{K})^{-1}\boldsymbol{\chi}^{0}\,\boldsymbol{K}$ are < 1 and most of them are $\ll 1$. By assuming $(\boldsymbol{I} - \lambda\,\boldsymbol{\chi}^{0}\,\boldsymbol{K})^{-1} \approx \boldsymbol{I}$, the contribution of Eq. 2.32 to the correlation energy in Eq. 2.28 can be computed analytically and the final result is exactly 0 [99]. Accordingly, the term in Eq. 2.32 is discarded. The effect of this approximation on the correlation energy was discussed in Ref. [99] (see, for example, Figure 2 and Table II) and it was found that neglecting the contribution in Eq. 2.32 leads to more accurate correlation energies. From a practical point of view this approximation allows us to express the correlation energy in terms of electron-hole integrals only. Indeed, by replacing Eq. 2.31 in Eq. 2.28 and carrying out an analytical integration over λ, the following expression for the AC-SOSEX is obtained:

$$E_{\mathrm{c}}^{\text{AC-SOSEX}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\mathrm{Tr}\left\{\ln\!\big(\boldsymbol{I} - \boldsymbol{\chi}^{0}(i\omega)\boldsymbol{K}\big)\,\boldsymbol{K}^{-1}\boldsymbol{B} + \boldsymbol{\chi}^{0}(i\omega)\boldsymbol{B}\right\}. \qquad (2.34)$$

Similarly to Eq. 2.26, also Eq. 2.34 reduces to the RPA in the $\boldsymbol{B} = \boldsymbol{K}$ case and to MP2 by Taylor expanding the logarithm.

Chapter 3

Numerical implementation within a plane-wave basis set using iterative algorithms

This chapter is based on original work previously published in Refs. [124, 28, 31].

The equations of RPA and beyond-RPA methods presented in Chapter 2 are based on an electron-hole representation (only conduction-valence products are present in the formalism). A first "brute-force" implementation of this formalism has been presented in Ref. [99] within the Gaussian basis set code Molpro [148]. Results obtained with this implementation will be used in Chapter 4 for benchmark and comparison purposes. However, in this manuscript we will focus on algorithms and applications based on the PW basis set. This numerical approach is more flexible and allows us to treat molecules, bulk solids, and periodic nanostructures within the same framework. However, the size of the basis set can be large, easily reaching several hundreds of thousands for large supercells. In a similar way, the number of electron-hole pairs also becomes rapidly unmanageable. For this reason it is convenient to represent the response functions involved in RPA and beyond-RPA methods using a compact basis set. Here we use the eigenvectors of an approximate polarizability that includes only the kinetic energy contribution [124]. This method is an approximation of the projective dielectric eigenpotential technique (PDEP) of Refs. 150, 100. In order to evaluate efficiently the matrix elements of the polarizability on the auxiliary basis set it is useful to resort to additional numerical techniques: The resummation of conduction states to obtain a formalism that depends on only occupied states [3, 147, 126] and the use of the Lanczos algorithm to compute the resolvent of the Hamitonian [126, 146, 102]. These methodologies to compute RPA and beyond-RPA correlation energies have been implemented as a separate module in the Quantum

Espresso package, which uses plane-waves and pseudopotentials [47, 46].

For the sake of simplicity the formalism in this chapter will be discussed exclusively including only the Γ point in the integration over the first Brillouin zone and by considering the orbitals as real. This assumption is valid for calculations involving sufficiently large supercells. The same ideas presented here can be extended to periodic solids. In this case the evaluation of RPA and beyond-RPA correlation energies requires a double integration in the first Brillouin zone (over k- and q-points). Until now this has been done only for the RPA [74]. Results obtained with this implementation are presented in Sec. 4.4 of the next chapter.

The sections of this chapter are structured as follows:

- In Sec. 3.1 we recast the equations of Chapter 2 by introducing an auxiliary basis set. At this stage we suppose that an optimally small basis set for response functions exists but we do not define it explicitly.

- In Sec. 3.2 we show how the equations presented in Sec. 3.1 can be reformulated by avoiding any reference to unoccupied states. Within this formalism the response functions are recast as matrix elements of the resolvent of the Hamiltonian.

- In Sec. 3.3 we show how the resolvent of the Hamiltonian can be computed efficiently by using the Lanczos algorithm.

- In Sec. 3.4 we discuss the construction of the auxiliary basis set.

- In Sec. 3.5 we present an additional algorithm, based on the Gram-Schmidt orthogonalization, that is essential to further lower the numerical cost of beyond-RPA methods.

3.1 Implementation using an auxiliary basis set

To start the presentation of the formalism we suppose that a basis set exists that allows for a compact representation of the matrices involved in Eq. 2.17, 2.26, and Eq. 2.34 (where compact representation means considerably smaller than the $N_{eh} \times N_{eh}$ full size). The details on the generation of this basis set will be discussed later in Sec. 3.4. In the following we will denote the vectors belonging to this auxiliary basis set as Φ and their total number as N_{aux}; the discussion will be carried out for the eh-TDHF correlation energy but analogous equations can be introduced in the AC-SOSEX case. The equations for the RPA can be simply obtained from the eh-TDHF formalism by setting to 0 the exchange component. While the focus of this manuscript is mainly related to the PW implementation, the formalism presented in this section is general and could be also employed for a density fitting implementation based on localized basis sets [99].

By using the completeness relation $\boldsymbol{I} = \sum_G |\Phi_G\rangle\langle\Phi_G|$ and the definition (see Eq. 2.22)

$$B_{vc,v'c'} = K_{vc,v'c'} - K^x_{vc,v'c'} = 2\langle\phi_v\phi_c|V|\phi_{v'}\phi_{c'}\rangle - \langle\phi_v\phi_{c'}|V|\phi_{v'}\phi_c\rangle, \qquad (3.1)$$

we obtain

$$\begin{aligned} K_{vc,v'c'} &= 2\sum_G \langle\phi_v\phi_c|V^{\frac{1}{2}}|\Phi_G\rangle\langle\Phi_G|V^{\frac{1}{2}}|\phi_{v'}\phi_{c'}\rangle \\ &= \sum_G N_{vc,G}N_{G,v'c'} = \left[\boldsymbol{N}\boldsymbol{N}^T\right]_{vc,v'c'} \qquad (3.2) \\ K^x_{vc,v'c'} &= \sum_G \langle\phi_v\phi_{c'}|V^{\frac{1}{2}}|\Phi_G\rangle\langle\Phi_G|V^{\frac{1}{2}}|\phi_{v'}\phi_c\rangle \\ &= \sum_G M_{vc,G}M_{G,v'c'} = \left[\boldsymbol{M}\boldsymbol{M}^T\right]_{vc,v'c'}. \qquad (3.3) \end{aligned}$$

Since the Coulomb potential is diagonal in a PW representation, the calculation of $V^{\frac{1}{2}}$ in Eqs. 3.2-3.3 is trivial.

By using Eqs. 3.1-3.3 the first contribution to the trace in Eq. 2.26 can be written as

$$\begin{aligned} & \mathrm{Tr}\left\{\ln(\boldsymbol{I}-\boldsymbol{\chi}^0\boldsymbol{B})\boldsymbol{B}^{-1}\boldsymbol{K}\right\} \\ = \; & \mathrm{Tr}\left\{\ln(\boldsymbol{I}-\boldsymbol{\chi}^0\boldsymbol{K}+\boldsymbol{\chi}^0\boldsymbol{K}^x)(\boldsymbol{K}-\boldsymbol{K}^x)^{-1}\boldsymbol{K}\right\} \\ = \; & \mathrm{Tr}\left\{\ln(\boldsymbol{I}-\boldsymbol{\chi}^0\boldsymbol{K}+\boldsymbol{\chi}^0\boldsymbol{K}^x)(\boldsymbol{K}-\boldsymbol{K}^x)^{-1}(\boldsymbol{\chi}^0)^{-1}\boldsymbol{\chi}^0\boldsymbol{K}\right\} \\ = \; & \mathrm{Tr}\left\{f\left[\boldsymbol{\chi}^0\boldsymbol{K}-\boldsymbol{\chi}^0\boldsymbol{K}^x\right]\boldsymbol{\chi}^0\boldsymbol{K}\right\} \\ = \; & \mathrm{Tr}\left\{f\left[\boldsymbol{\chi}^0\boldsymbol{K}-\boldsymbol{\chi}^0\boldsymbol{K}^x\boldsymbol{\chi}^0\boldsymbol{K}(\boldsymbol{\chi}^0\boldsymbol{K})^{-1}\right]\boldsymbol{\chi}^0\boldsymbol{K}\right\} \\ = \; & \mathrm{Tr}\left\{f\left[\boldsymbol{\chi}^0\boldsymbol{N}\boldsymbol{N}^T-\boldsymbol{\chi}^0\boldsymbol{M}\boldsymbol{M}^T\boldsymbol{\chi}^0\boldsymbol{N}\boldsymbol{N}^T(\boldsymbol{\chi}^0\boldsymbol{N}\boldsymbol{N}^T)^{-1}\right]\boldsymbol{\chi}^0\boldsymbol{N}\boldsymbol{N}^T\right\} \\ = \; & \mathrm{Tr}\left\{\boldsymbol{N}^Tf\left[\boldsymbol{\chi}^0\boldsymbol{N}\boldsymbol{N}^T-\boldsymbol{\chi}^0\boldsymbol{M}\boldsymbol{M}^T\boldsymbol{\chi}^0\boldsymbol{N}\boldsymbol{N}^T(\boldsymbol{\chi}^0\boldsymbol{N}\boldsymbol{N}^T)^{-1}\right]\boldsymbol{\chi}^0\boldsymbol{N}\right\} \\ = \; & \mathrm{Tr}\left\{f\left[\boldsymbol{N}^T\boldsymbol{\chi}^0\boldsymbol{N}-\boldsymbol{N}^T\boldsymbol{\chi}^0\boldsymbol{M}\boldsymbol{M}^T\boldsymbol{\chi}^0\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{\chi}^0\boldsymbol{N})^{-1}\right]\boldsymbol{N}^T\boldsymbol{\chi}^0\boldsymbol{N}\right\}, \qquad (3.4) \end{aligned}$$

where we have defined the function of a matrix $f(\boldsymbol{A}) = \ln(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{A}^{-1}$ and used the cyclic property of the trace operator; additionally, the result in Eq. 3.4 exploits properties of functions of a matrix that are described in more detail in Ref. 99. The dependence of $\boldsymbol{\chi}^0$ on $i\omega$ is kept implicit in Eq. 3.4.

It is now convenient to introduce the definitions

$$\begin{aligned} C_{QR}(i\omega) &= (\boldsymbol{N}^T\boldsymbol{\chi}^0\boldsymbol{N})_{QR} \\ &= \langle\Phi_Q|V^{\frac{1}{2}}\hat{\chi}^0V^{\frac{1}{2}}|\Phi_R\rangle, \\ &= 4\sum_{vc}\langle\Phi_Q|V^{\frac{1}{2}}|\phi_v\phi_c\rangle\,\mathfrak{R}\frac{1}{i\omega+\epsilon_v-\epsilon_c}\,\langle\phi_v\phi_c|V^{\frac{1}{2}}|\Phi_R\rangle, \qquad (3.5) \end{aligned}$$

$$J^{vv'}_{QG}(i\omega) = 2\sum_{c}\langle\Phi_Q|V^{\frac{1}{2}}|\phi_v\phi_c\rangle\,\Re\frac{1}{i\omega+\epsilon_v-\epsilon_c}\langle\phi_{v'}\phi_c|V^{\frac{1}{2}}|\Phi_G\rangle, \tag{3.6}$$

$$L^{vv'}_{GR}(i\omega) = 2\sum_{c'}\langle\Phi_G|V^{\frac{1}{2}}|\phi_v\phi_{c'}\rangle\Re\frac{1}{i\omega+\epsilon_{v'}-\epsilon_{c'}}\langle\phi_{v'}\phi_{c'}|V^{\frac{1}{2}}|\Phi_R\rangle, \tag{3.7}$$

and

$$Z_{QR}(i\omega) = (\boldsymbol{N}^T\boldsymbol{\chi}^0\boldsymbol{M}\boldsymbol{M}^T\boldsymbol{\chi}^0\boldsymbol{N})_{QR} = 2\sum_{vv'}\sum_{G}J^{vv'}_{QG}(i\omega)L^{vv'}_{GR}(i\omega); \tag{3.8}$$

the compact notation in terms of the operator $\hat{\chi}^0$ on the second line of Eq 3.5 will be useful later. In Eqs. 3.5-3.8 the subscripts Q, R, and G are used to denote different vectors Φ belonging to the auxiliary basis set. By using the definitions in Eqs. 3.5-3.8 we can finally reformulate the eh-TDHF correlation energy expression in Eq. 2.26 to obtain

$$\begin{aligned} E_c^{\text{eh-TDHF}} &= \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\text{Tr}\big\{\ln\big[\boldsymbol{I}-\mathbf{C}(i\omega)+\mathbf{Z}(i\omega)\mathbf{C}^{-1}(i\omega)\big] \\ &\qquad \times\big[\mathbf{C}(i\omega)-\mathbf{Z}(i\omega)\mathbf{C}^{-1}(i\omega)\big]^{-1}\mathbf{C}(i\omega)+\mathbf{C}(i\omega)\big\}. \end{aligned} \tag{3.9}$$

This final expression for the correlation energy is more convenient than the original form Eq. 2.26 since it involves $N_{aux}\times N_{aux}$ matrices, where $N_{aux}\ll N_{eh}$.

If the exchange term in the eh-TDHF kernel is set to zero we have that $\mathbf{Z}(i\omega)=0$ and the correlation energy reduces to the RPA expression

$$E_c^{\text{RPA}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\text{Tr}\big\{\ln\left[\boldsymbol{I}-\mathbf{C}(i\omega)\right]+\mathbf{C}(i\omega)\big\}; \tag{3.10}$$

where $\boldsymbol{I}-\mathbf{C}$ is nothing but the symmetrized dielectric matrix $\tilde{\varepsilon}$ represented in the Φ basis set [124].

By Taylor expanding the logarithm in Eq. 3.9 up to second order we obtain the MP2 approximation:

$$E_c^{\text{MP2}} \quad = -\tfrac{1}{2}\int_0^{\infty}\tfrac{d\omega}{2\pi}\text{tr}\Big\{\boldsymbol{C}^2(i\omega)-\boldsymbol{Z}(i\omega)\Big\}. \tag{3.11}$$

This equation reformulates the MP2 correlation energy in a form that could be easily computed using an implementation derived from our eh-TDHF and AC-SOSEX code. However, RPA and beyond-RPA methods work as post-processing to a standard DFT calculation using local or semi-local functionals. While being numerically stable, the MP2 correlation energies computed through Eq. 3.11 using a DFT starting point are incorrect, since the Brillouin theorem is not anymore valid. Indeed the MP2 correlation energy is rather derived in a post-Hartree-Fock framework and is not supposed to use DFT orbitals and energy levels as starting point [141]. We proposed an implementation of Eq. 3.11 in a Hartree-Fock framework in Ref. [30]. However, the calculation of the HF exact exchange is highly expensive within the PW basis set and this also limits the applicability of our MP2 code to systems with few valence electrons. This MP2 implementation will not be further discussed on this manuscript, which will remain focused on RPA-type methods.

Also the AC-SOSEX correlation energy (Eq. 2.34) can be rewritten in terms of the compact auxiliary basis set:

$$\begin{aligned} E_c^{\text{AC-SOSEX}} = & \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\mathrm{Tr}\big\{\left[\ln\left(\boldsymbol{I}-\mathbf{C}(i\omega)\right)+\mathbf{C}(i\omega)\right] \\ & \times\left[\boldsymbol{I}-\mathbf{C}^{-1}(i\omega)\mathbf{Z}(i\omega)\mathbf{C}^{-1}(i\omega)\right]\big\}. \end{aligned} \tag{3.12}$$

Similarly to Eq. 3.9, also the AC-SOSEX reduces to MP2 by Taylor expanding the logarithm and to RPA by setting $\mathbf{Z}(i\omega)=0$.

To conclude this section it is useful to show explicitly the form of Eq. 3.5 when plane-waves are used as an auxiliary basis set. Indeed, this formulation is used in traditional implementations to compute RPA correlation energies [39, 92, 58]. By denoting the reciprocal-lattice vectors with $\mathbf{G}$ and $\mathbf{G}'$ we have

$$\begin{aligned} C_{\mathbf{G},\mathbf{G}'}(i\omega) &= 4\sum_{vc}\langle\mathbf{G}|V^{\frac{1}{2}}|\phi_v\phi_c\rangle\,\Re\frac{1}{i\omega+\epsilon_v-\epsilon_c}\,\langle\phi_v\phi_c|V^{\frac{1}{2}}|\mathbf{G}'\rangle \\ &= 4\sum_{vc}\frac{1}{\mathbf{G}}\langle\phi_v|e^{i\mathbf{G}\cdot\mathbf{r}}|\phi_c\rangle\,\Re\frac{1}{i\omega+\epsilon_v-\epsilon_c}\,\langle\phi_v|e^{-i\mathbf{G}'\cdot\mathbf{r}'}|\phi_c\rangle\frac{1}{\mathbf{G}'} \\ &= \frac{1}{\mathbf{G}}\chi^0_{\mathbf{G},\mathbf{G}'}\frac{1}{\mathbf{G}'} \end{aligned} \tag{3.13}$$

Since in traditional RPA implementations the response functions are represented using the PW basis set, practical calculations involve storage and diagonalization of matrices that become rapidly large by increasing the kinetic energy cut-off and/or the cell size [39, 92, 58, 137]. Additionally, the full diagonalization of the Hamiltonian is required to evaluate RPA and beyond-RPA correlation energies and also this task is influenced by the number of PWs. This last issue will be addressed in the next section.

3.2 Elimination of the empty states

Eqs. 3.5-3.8 can be evaluated directly by diagonalizing the ground-state KS Hamiltonian $\hat{H}$ to obtain conduction and valence states. However, the large dimension of the PW basis set leads to a very high number of conduction states and fully converged results can be obtained only through extrapolation techniques [92] or by diagonalizing the full Hamiltonian. Similarly to recent work on the calculation of RPA correlation energies [100, 85, 124], our numerical implementation takes instead advantage of DFPT techniques [3]. Indeed, we can notice that Eq. 3.5 and Eq. 3.6 can be rewritten in the form

$$\begin{aligned} C_{QR}(i\omega) &= 4\sum_{vc}\langle\Phi_Q|V^{\frac{1}{2}}|\phi_v\phi_c\rangle\,\Re\frac{1}{i\omega+\epsilon_v-\epsilon_c}\,\langle\phi_v\phi_c|V^{\frac{1}{2}}|\Phi_R\rangle \\ &= 4\sum_{vc}\langle\phi_v|\Phi_Q V^{\frac{1}{2}}|\phi_c\rangle\langle\phi_c|\,\Re\frac{1}{i\omega+\epsilon_v-\hat{H}}\,|\phi_c\rangle\langle\phi_c|V^{\frac{1}{2}}\Phi_R|\phi_v\rangle \\ &= 4\sum_{v}\langle\phi_v|\Phi_Q V^{\frac{1}{2}}\hat{P}\,\Re\frac{1}{i\omega+\epsilon_v-\hat{H}}\,\hat{P}V^{\frac{1}{2}}\Phi_R|\phi_v\rangle \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} J^{vv'}_{QG}(i\omega) &= 2\sum_{c}\langle\Phi_Q|V^{\frac{1}{2}}|\phi_v\phi_c\rangle\,\Re\frac{1}{i\omega+\epsilon_v-\epsilon_c}\,\langle\phi_{v'}\phi_c|V^{\frac{1}{2}}|\Phi_G\rangle \\ &= 2\sum_{c}\langle\phi_v|\Phi_Q V^{\frac{1}{2}}|\phi_c\rangle\langle\phi_c|\,\Re\frac{1}{i\omega+\epsilon_v-\hat{H}}\,|\phi_c\rangle\langle\phi_c|V^{\frac{1}{2}}\Phi_G|\phi_{v'}\rangle \\ &= 2\langle\phi_v|\Phi_Q V^{\frac{1}{2}}\hat{P}\,\Re\frac{1}{i\omega+\epsilon_v-\hat{H}}\,\hat{P}V^{\frac{1}{2}}\Phi_G|\phi_{v'}\rangle, \end{aligned} \tag{3.15}$$

respectively; also Eq. 3.7 can be written in an analogous way. In Eqs. 3.14-3.15 $\hat{P} = \sum_c |\phi_c\rangle\langle\phi_c|$ is the projector onto the conduction state subspace. By exploiting the completeness of the basis set, the projector onto the conduction state subspace can be expressed as $\hat{P} = \hat{I} - \sum_v |\phi_v\rangle\langle\phi_v|$, where in this case the sum is limited to the occupied states only [3]. This approach does not require approximations and automatically includes all the conduction states described by the PW basis set. Once the auxiliary basis set elements Φ_Q and Φ_R are fixed the matrix elements involved in the last line of Eq. 3.14 and Eq. 3.15 can be efficiently computed by using the Lanczos algorithm for response functions [126, 146, 102]. This approach will be discussed in detail in the next section.

3.3 Lanczos algorithm

Eqs. 3.14-3.15 allow us to compute RPA and beyond-RPA correlation energies without referring explicitly to unoccupied states. However, this formulation requires the computation of the resolvent of the Hamiltonian $\hat{H}$ for many different values of $i\omega$. Specifically, the mathematical problem that needs to be solved in Eqs. 3.14-3.15 can be formulated in the following general way:

$$g(i\omega) = \langle u|(i\omega - \hat{A})^{-1}|v\rangle, \tag{3.16}$$

where $\hat{A}$ is a Hermitian operator. For the specific case of Eqs. 3.14-3.15 we have $\hat{A} = \hat{H} - \epsilon_v$. Depending on the number of occupied states and the size of the auxiliary basis set, the matrix element of the resolvent in Eq. 3.16 has to be computed for several different vectors u and v (see also Secs. 3.4 and 3.5).

The calculation of $g(i\omega)$ can be obtained by using standard iterative techniques to solve Hermitian linear systems (e.g., the conjugate gradient algorithm) [131]. In this case the linear system $(i\omega - \hat{A})^{-1}|z\rangle = v\rangle$ is solved and $g(i\omega)$ is computed as $\langle u|z\rangle$. The disadvantage of this approach is that a different linear system has to be solved for each different value of $i\omega$.

The Lanczos algorithm is known to be an efficient method to compute the resolvent of Hermitian operators by performing a single iterative recursion independent of ω [61].The standard Lanczos algorithm is limited to the case u=v; however, for our purpose it is necessary to compute also the off-diagonal elements of $g(i\omega)$. For this task we use the Lanczos algorithm proposed in Ref. 126 for time-dependent density functional theory and later applied to GW calculations [146, 102]. This algorithm generates iteratively a series of vectors $\{q_1, q_2, \cdots\}$ by using the following procedure:

$$\begin{aligned} q_0 &= 0, \\ q_1 &= v/\sqrt{\langle v|v\rangle}, \\ \beta_{n+1}|q_{n+1}\rangle &= \hat{A}|q_n\rangle - \alpha_n|q_n\rangle - \beta_n|q_{n-1}\rangle \end{aligned} \tag{3.17}$$

where β_{n+1} is determined in order to impose the normalization condition $\langle q_{n+1}|q_{n+1}\rangle = 1$ and $\alpha_n = \langle q_n|\hat{A}|q_n\rangle$. In the orthonormal basis set of the vectors $\{q_1, q_2, \cdots\}$ the matrix $\hat{A}$ has tridiagonal form:

$$T^j = \begin{pmatrix} \alpha_1 & \beta_2 & 0 & \cdots & 0 \\ \beta_2 & \alpha_2 & \beta_3 & 0 & \vdots \\ 0 & \beta_3 & \alpha_3 & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \beta_j \\ 0 & \cdots & 0 & \beta_j & \alpha_j \end{pmatrix} \tag{3.18}$$

where α and β are the coefficients of the Lanczos recursion (Eq. 3.17) and j is the maximum number of Lanczos iterations performed. The value of $g(i\omega)$ can then be approximated as

$$g(i\omega) \approx \langle \zeta^j|(i\omega - T^j)^{-1}|e_1^j\rangle. \tag{3.19}$$

where ζ^{jT} is a j-dimensional vector defined as $(\langle u|q_1\rangle, \langle u|q_2\rangle, \cdots, \langle u|q_j\rangle)$ and e_1^j is the j-dimensional unit vector $(1, 0, \cdots, 0)$. In general, the dimension j (the number of Lanczos iterations) of the matrix T^j necessary to obtain an accurate approximation of $g(i\omega)$ is much smaller than the dimension of the full matrix $\hat{A}$.

There are two important advantages in using Eq. 3.19. First, the matrix T^j and the corresponding Lanczos iterative recursion do not depend on $i\omega$. Once T^j is generated, the value of $g(i\omega)$ can be computed for many values of $i\omega$ by simple linear algebra operations in a small j-dimensional space (Eq. 3.19). For all the numerical applications considered in this manuscript 30 Lanczos iterations have been found to be sufficient to achieve a satisfactory level of accuracy. Second, the Lanczos recursion (Eq. 3.17) depends only on the right vector $|v\rangle$ and different ζ^j vectors can be generated "on the fly" during the Lanczos chain in order to compute matrix elements of $g(i\omega)$ for different left vectors $\langle u|$. This feature is particularly convenient to evaluate Eqs. 3.14-3.15, where the calculation of several off-diagonal elements $(u \neq v)$ is required.

One of the main advantages of the present method is the efficient treatment of the dynamical effects in the response function. Once the tridiagonal matrix T^l is computed through the Lanczos algorithm the response functions can be efficiently evaluated for several different values of ω (see Eq. 3.19). In this work the integrals over ω in Eqs. 3.9, 3.10, and 3.12 are evaluated by a 20-point Gauss-Legendre quadrature in the interval 0 to 200 Ry. The accuracy of this integration scheme can be estimated by comparison with the simple trapezoidal rule. An example has been presented in Ref. [124]: By considering 100 kinetic-PDEP basis vectors for the benzene dimer the difference between the values of ΔE_c obtained with the trapezoidal rule (uniform grid with a 0.0001 Ry step) and the Gauss-Legendre quadrature is of the order of 10^{-6} kcal/mol. The results are also well converged with respect to the integration range: By integrating between 0 and 100 Ry the change in the value of ΔE_c is of the order of 10^{-5} kcal/mol.

3.4 Construction of a compact auxiliary basis set by diagonalizing an approximate response function

Despite the several advantages of the Lanczos algorithm, in order to make this procedure efficient it is necessary to consider a basis set $\{\Phi_I\}$ as small as possible. Indeed, even if the Lanczos procedure requires only a few iterations, it is still necessary to perform a number of Lanczos chains which is proportional to the size of the basis set $\{\Phi_I\}$. In order to discuss the construction of the optimal basis we redefine Eq. 3.5 as:

$$\begin{aligned} C_{QR} &= 4\sum_{vc} \langle \Phi_Q | V^{\frac{1}{2}} | \phi_v \phi_c \rangle \, \Re \frac{1}{\epsilon_v - \epsilon_c} \, \langle \phi_v \phi_c | V^{\frac{1}{2}} | \Phi_R \rangle \\ &= 4\sum_{vc} \langle \bar{\Phi}_Q | \phi_v \phi_c \rangle \, \Re \frac{1}{\epsilon_v - \epsilon_c} \, \langle \phi_v \phi_c | \bar{\Phi}_R \rangle \qquad (3.20) \\ &= \int \bar{\Phi}_Q(\mathbf{r}) \chi^0(\mathbf{r}, \mathbf{r}') \bar{\Phi}_R(\mathbf{r}') d\mathbf{r} d\mathbf{r}', \qquad (3.21) \end{aligned}$$

where we focus on the static case ($\omega = 0$). It is possible to switch between the two alternative forms of the auxiliary basis set Φ and $\bar{\Phi}$ simply by multiplying by $V^{\frac{1}{2}}$ or its inverse (which are diagonal in reciprocal space). In Eq. 3.21 we used the static polarizability in the real-space representation

$$\chi^0(\mathbf{r}, \mathbf{r}') = 4\sum_{cv} \frac{\phi_v(\mathbf{r})\phi_c(\mathbf{r})\phi_c(\mathbf{r}')\phi_v(\mathbf{r}')}{\epsilon_v - \epsilon_c}; \qquad (3.22)$$

this definition will be useful below. From Eq. 3.20 it is clear that an optimally small basis set $\bar{\Phi}$ needs to efficiently represent the $\phi_c\phi_v$ products. While ϕ_c and ϕ_v are orthonormal orbitals, the products $\phi_c\phi_v$ might have a strong linear dependence. This observation suggests the possibility of representing χ^0 on a relatively small basis set (at least much smaller than the number of products $\phi_c\phi_v$). The resolution of the identity techniques used to speed up localized basis set RPA calculations are based on this idea [35, 117]. However, in a PW implementation the number of products $\phi_c\phi_v$ could reach several millions and the corresponding subspace would still require a large basis set to be accurately represented. In a PW framework an optimally small basis set for χ^0 (or $\tilde{\epsilon}$) can be generated by "eliminating" the linear dependence of the products $\phi_c\phi_v$ keeping into account at the same time the weighting factor $1/(\epsilon_v - \epsilon_c)$ (see Eq. 3.22). For example, for a molecule the importance of the $\phi_{LUMO}\phi_{HOMO}$ term in the sum in Eq. 3.22 is enhanced by the factor $1/(\epsilon_{HOMO} - \epsilon_{LUMO})$; similarly, the contribution of the $\phi_{(LUMO+N)}\phi_{HOMO}$ term (where N is a large integer) to the sum in Eq. 3.22 might be much smaller because $1/(\epsilon_{HOMO} - \epsilon_{(LUMO+N)})$ goes to 0 for $N \to +\infty$. Accordingly, the optimal basis set should well represent the $\phi_{LUMO}\phi_{HOMO}$ term rather than $\phi_{(LUMO+N)}\phi_{HOMO}$.

The idea of keeping into account the linear dependence of the $\phi_c\phi_v$ products together with the weight $1/(\epsilon_v - \epsilon_c)$ is implicit in projective dielectric eigenpotential (PDEP) method [150, 100, 85, 112]. Within this approach the response functions χ^0 or, equivalently, $\tilde{\varepsilon}$ are iteratively diagonalized. At each iteration χ^0

is applied to a set of trial potentials $\{\bar{\Phi}_I^{trial}\}$ by using Eqs. 3.21 and 3.14:

$$\int \chi^0(\mathbf{r},\mathbf{r}')\bar{\Phi}_I^{trial}(\mathbf{r}')d\mathbf{r}' \\ = 4\sum_v \phi_v(\mathbf{r})\langle\mathbf{r}|(\epsilon_v - \hat{H})^{-1}\hat{P}|\bar{\Phi}_I^{trial}\phi_v\rangle \tag{3.23}$$

until the iterative algorithm (e.g. conjugate gradient) reaches convergence; Eq. 3.23 can be easily generalized to the dynamical case. The diagonalization of the static $\tilde{\varepsilon}$ has been used to build the optimal basis set used in the Lanczos-based GW implementation of Refs. 102, 109. In the PDEP method only a limited number of dominant eigenvalues and eigenpotentials (eigenvectors) are necessary to represent χ^0 or $\tilde{\varepsilon}$. Indeed, the eigenpotentials corresponding to the largest (in absolute value) eigenvalues have a strong superimposition with the products $\phi_c\phi_v$ through the term $\hat{P}|\bar{\Phi}_I^{trial}\phi_v\rangle$ in Eq. 3.23, while at the same time $(\epsilon_v - \hat{H})^{-1}$ keeps into account the weight of the different transitions. This is done without explicitly computing the empty states ϕ_c of the system. This qualitative argument suggests why the PDEP technique provides a compact representation for χ^0 and $\tilde{\epsilon}$. The PDEP method has been used to compute RPA correlation energies by diagonalizing $\tilde{\epsilon}$ for each value of $i\omega$ included in the numerical integration of Eq. 3.10 [85, 100]. The main drawback of the PDEP procedure is that, each time χ_0 is applied to a vector through Eq. 3.23, the $(\epsilon_v - \hat{H})^{-1}$ operator has to be evaluated by solving a linear system.

In this manuscript we discuss a new and more efficient method to construct a basis set to compactly represent dielectric matrices. The main idea consists in noticing that for sufficiently high energies ϵ_c is mostly determined by the kinetic energy contribution ϵ_c^{kin}. For this reason we approximate the weighting factor in Eq. 3.22 as $1/(\epsilon_v - \epsilon_c) \approx 1/(\epsilon_v - \epsilon_c^{kin})$. This approximation describes well high energy excitations but is certainly not valid for the low energy ϵ_c's. However, since the main purpose of this procedure is to build a basis set that "discard" the high energy contributions to Eq. 3.22, this approximation is acceptable and works well in practice (see also the numerical results of Chapter 4). Since we want to obtain a compact representation for χ^0 without computing explicitly empty states, similarly to the PDEP procedure, we consider the iterative diagonalization of a χ^0_{kin} operator containing only the kinetic energy contribution. This operator can be applied to a trial potential in the following way:

$$\int \chi^0_{kin}(\mathbf{r},\mathbf{r}')\bar{\Phi}_I^{trial}(\mathbf{r}')d\mathbf{r}' \\ = 4\sum_v \phi_v(\mathbf{r})\langle\mathbf{r}|\hat{P}(\epsilon_v + \nabla^2/2)^{-1}\hat{P}|\bar{\Phi}_I^{trial}\phi_v\rangle, \tag{3.24}$$

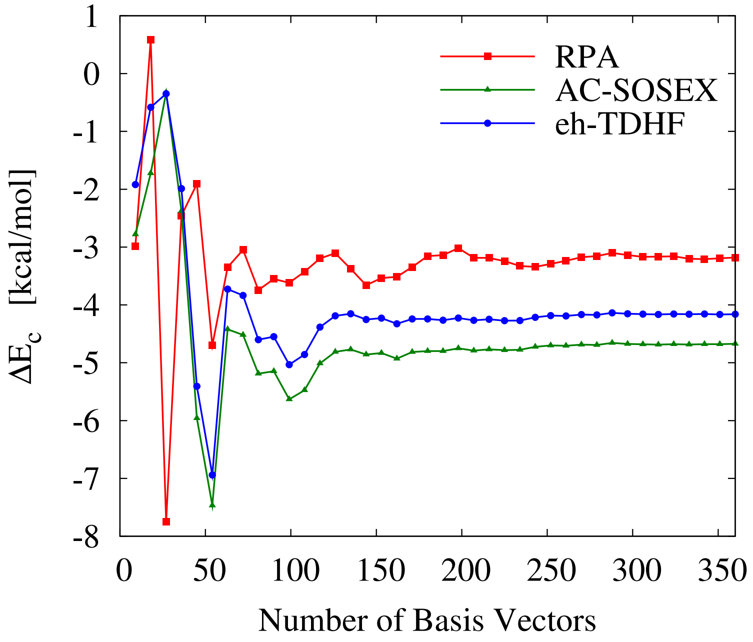


Figure 3.1: Convergence with respect to the auxiliary basis set size of the correlation energy contribution to the energy of the C_2H_2+H_2O → CH_3CHO reaction. Specifically, $\Delta E_c = E_c^{CH_3CHO} - E_c^{C_2H_2} - E_c^{H_2O}$.

where the kinetic energy operator $-\nabla^2/2$ is diagonal in a plane-wave representation (it is proportional to G^2 in reciprocal space). This is one of the main advantages of Eq. 3.24: Since $(\epsilon_v + \nabla^2/2)$ is diagonal its inversion is trivial and the implementation to solve Eq. 3.24 is simple and efficient. We will call this procedure kinetic-PDEP. Eq. 3.24 gives an accurate description of high energy excitations while the low energy excitations are approximated in a rather poor way. For example, the term $(\epsilon_v + \nabla^2/2)$ does not contain any information on the exact value of the gap of the system. However, it is important to stress that the only purpose of the kinetic-PDEP procedure is to build a compact basis whose size can be systematically increased to achieve convergence. The details of the real electronic energy levels (as well as the dynamical contributions) are included by the Lanczos algorithm.

In Eq. 3.24 χ^0_{kin} is considered exclusively in the static limit ($i\omega = 0$); this approximation has been proven reliable to built an effective basis set also for the dynamical case ($i\omega \neq 0$) [102, 109, 50]. However, the structure of the matrix elements C_{QR} in Eq. 3.14 is different from the structure of $J^{vv'}_{QG}$ (or, similarly, $L^{vv'}_{GR}$) in Eq. 3.15: $J^{vv'}_{QG}$ does not involve a summation over the occupied states. From a rigorous point of view the basis set obtained from the kinetic-PDEP (or PDEP) method is designed to compactly represent C_{QR} in Eq. 3.5, which is the main ingredient for RPA calculations. As a consequence, it is not straightforward to conclude that the kinetic-PDEP basis set would also efficiently represent the matrix elements in Eq. 3.6 and Eq. 3.7. However, as discussed qualitatively in Ref. 124, the eigenvectors obtained by diagonalizing the operator in Eq. 3.24 well represent the occupied-virtual orbital products by eliminating at the same time their linear dependence and the excitations too high in energy, that weakly contribute to the response functions. Since Eq. 3.2 and Eq. 3.3 also involve occupied-virtual state products and excitation energies at the denominator, it is reasonable to suppose that the kinetic-PDEP basis set would also well represent these matrix elements. From numerical calculations we observed that this is indeed the case. In Fig. 3.1, we show the convergence with respect to the auxiliary basis set of the correlation energy contribution to the energy of the C_2H_2+H_2O $\rightarrow$ CH_3CHO reaction (this reaction is part of a test set that will be considered in detail in Sec. 4.3). Specifically, $\Delta E_c = E_c^{CH_3CHO} - E_c^{C_2H_2} - E_c^{H_2O}$. Fig. 3.1 allows us to observe the rapid convergence of ΔE_c as a function of the size of the auxiliary basis set. For example, by increasing N_{aux} from 270 to 360 the values ΔE_c evaluated at the AC-SOSEX and eh-TDHF are all contained within a 0.04 kcal/mol interval for AC-SOSEX and eh-TDHF and within a 0.1 kcal/mol interval for RPA. In general, we observed that the eh-TDHF and AC-SOSEX correlation energies have a convergence rate with respect to N_{aux} which is similar to the RPA case and sometimes even slightly better. This allows us to use a more compact basis set to solve Eq. 2.26, whose direct solution would be very challenging by expressing the response functions on the eh basis set or the plane-wave basis set. For example, in the case shown in Fig. 3.1 the use of the full plane-wave basis set for the potential to directly represent the response functions in Eq. 2.26 or Eq. 2.34 would involve matrices with a 850000×850000 size; alternatively, the use of Eq. 3.9 and Eq. 3.12 with the kinetic-PDEP auxiliary basis set leads to converged results with as little as 360 basis elements corresponding to 360×360 matrices.

In order to converge energy differences as done in Fig. 3.1 it is necessary to use a number of basis vectors proportional to the number of electrons for each component. For example, if 360 basis vectors are used to compute the correlation energy of CH_3CHO, for C_2H_2 and H_2O we should use 200 and 160 vectors, respectively. This procedure will be described in detail in Chapter 5.

3.5 Gram-Schmidt orthogonalization

Even though the use of the compact auxiliary basis greatly reduces the dimensionality of dielectric matrices, the calculation requires the evaluation $(N_{aux})^2$ matrix elements for $\boldsymbol{C}$ and $(N_{aux}N_v)^2$ matrix elements for $\boldsymbol{J}$ and $\boldsymbol{L}$. For the following discussion, we introduce the compact notation $|U^v_G\rangle = \hat{P}V^{\frac{1}{2}}\Phi_G|\phi_v\rangle$ and a similar definition for $\langle U^v_G|$. As seen in Eq. 3.10 the calculation of the RPA correlation energy requires the evaluation of matrix elements of the type:

$$C_{QR}(i\omega) = 4\sum_v \langle U^v_Q|\Re\left(i\omega + \epsilon_v - \hat{H}\right)^{-1}|U^v_R\rangle; \tag{3.25}$$

which is just Eq. 3.14 written in a more compact notation. The AC-SOSEX and eh-TDHF require the evaluation of additional terms such as:

$$J^{vv'}_{QG}(i\omega) = 2\langle U^v_Q|\Re\left(i\omega + \epsilon_v - \hat{H}\right)^{-1}|U^{v'}_G\rangle \tag{3.26}$$

The matrix elements in Eqs. 3.25-3.26 can be efficiently computed by using the Lanczos algorithm. Since the subscripts v and v' are in general different in Eq. 3.26, the numerical complexity of beyond-RPA methods is higher than that of RPA, which requires only the evaluation of Eq. 3.25. This is a major bottleneck for the AC-SOSEX and eh-TDHF methods whose numerical complexity scales as $N_{PW} \times N^2_{aux} \times N^2_v$, where N_{PW} is the number of plane-wave, N_{aux} the number of auxiliary basis vectors, and N_v the number of valence states (compared to $N_{PW} \times N^2_{aux} \times N_v$ for RPA). Intuitively, this can be understood by considering that the Lanczos algorithm computes the solution to the $N_v \times N_{aux}$ linear systems $(i\omega + \epsilon_v - \hat{H})^{-1}|U^{v'}_G\rangle$; each solution, represented on N_{PW} plane-waves, is projected over the $N_v \times N_{aux}$ left vectors $\langle U^v_Q|$. Additionally, in order to store all the left vectors the memory requirements for beyond-RPA calculations scale as $N_{PW} \times N_v \times N_{aux}$ (in RPA it is sufficient to store only one value of v at a time); this problem could be overcome by recomputing these matrix elements on the fly or by storing them on disk but this comes at a big loss in efficiency. Because of these computational and memory issues, the applicability of the AC-SOSEX and eh-TDHF methods would be limited to molecules with a maximum of twenty to thirty electrons. For example, the largest dimers in the S22 test set considered in Sec. 4.1.2 (≈ 100 electrons) would be completely out of reach.

In order to overcome these difficulties in applying the beyond-RPA methods to large systems, we use a procedure inspired by previous work of Umari and Baroni in the context of GW calculations [146]. Specifically, we reduce the large number of right and left vectors in Eq. 3.26 by applying a Gram-Schmidt orthogonalization procedure to eliminate vectors that are linearly dependent. To do so, we define a new set of orthonormal vectors as,

$$\begin{aligned}
|O_1\rangle &= |U_1^1\rangle / N_1 \\
|O_2\rangle &= \left[|U_1^2\rangle - \langle O_1 | U_1^2\rangle \, |O_1\rangle \right] / N_2 \\
&\cdots \\
|O_{N_{aux}\times N_v}\rangle &= \left[|U_{N_{aux}}^{N_v}\rangle - \sum_{i=1}^{N_{aux}\times N_v} \langle O_i | U_{N_{aux}}^{N_v}\rangle \, |O_i\rangle \right] / N_{(N_{aux}\times N_v)},
\end{aligned} \tag{3.27}$$

where the factors N_n at the denominator are used to normalize each vector of the basis $|O\rangle$. Up to this point this procedure does not allow for any computational saving but rather introduces additional expensive numerical computations. However, if for a certain $|O_n\rangle$ we have $N_n \approx 0$, this vector can be fully represented by the previous ones $(|O_1\rangle, \cdots, |O_{n-1}\rangle)$ and does not need to be included in the orthonormalized basis set. By setting the threshold for the normalizing factors to a reasonable value,this idea can be used to generate a basis set $(|O_1\rangle, \cdots, |O_{N_{ortho}}\rangle)$ where $N_{ortho} \ll N_{aux} \times N_v$. The threshold on the normalizing factors N_n has to be chosen with a compromise between computational time and accuracy. In this work we chose a variable threshold, that will be referred to as "default threshold" (DT): 10^{-5} for the first 20% of the vectors $|U\rangle$, 10^{-4} between 20% and 50%, and 10^{-3} for all the remaining vectors. In this way the first vectors, which give a more important contribution to the final result, are more accurately represented. While this specific choice is to a certain extent arbitrary, we will further discuss this point at the end of this section.

The gain in practical calculations can be quantitatively understood by looking at Fig. 3.2, where data for all the dimers and monomers considered in Secs. 4.1.1 and 4.1.2 are presented (92 monomers and 46 dimers). In this figure N_{aux} is fixed to forty times the number of KS valence states N_v, which is sufficient to reach convergence for almost all systems (see discussion in Sec. 4.1 and Chapter 5). While by definition $N_{aux} \times N_v$ grows quadratically with the number of electrons, N_{ortho} has an approximately linear growth. This leads to a significant saving, with the memory necessary to store the vectors reduced to $N_{PW} \times N_{ortho}$ and the numerical complexity reduced to $N_{PW} \times N_{ortho}^2$.

This linear growth is to a certain extent related to the threshold choice. Indeed, in the case of a sufficiently small threshold N_{ortho} would be comparable to $N_{aux} \times N_v$. With the DT this approximately linear behavior has been deduced empirically considering 138 molecular systems. Since this is a rather extended ensemble, we believe that this effect is robust and generalizable to all the molecules. However, in order to be meaningful, this observation of a linear behavior has to be associated with a high level of accuracy when only N_{ortho} vectors are included in the calculations rather than the full $N_{aux} \times N_v$ vector set. This point is further discussed at the end of this section and in Sec. 4.1.1.

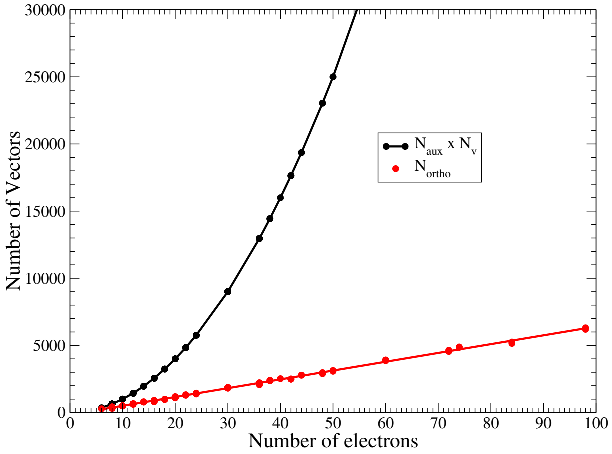


Figure 3.2: This figure shows how the number of vectors involved in the computation of dielectric matrices can be decreased by a Gram-Schmidt orthogonalization (GSO) procedure. The original method requires $N_{aux} \times N_v$ vectors, which grows quadratically with the number of electrons. The GSO version requires N_{ortho} matrix elements, which grows linearly with the number of electrons. See text for further details.

As it will be shown in Sec. 4.1.2, this procedure allows for calculations for systems with up to 98 electrons and about 400k plane-waves. In order to compute the correlation energy in Eqs. 3.9 and 3.12 the Lanczos algorithm is used to evaluate the matrix elements $\langle O_i|\Re\left(i\omega+\epsilon_v-\hat{H}\right)^{-1}|O_j\rangle$ and the matrix $\boldsymbol{J}$ in Eq. 3.26 is then reconstructed as

$$J_{QG}^{vv'} = 2\sum_{i,j=1}^{N_{ortho}} \langle U_Q^v|O_i\rangle\langle O_i|\Re\left(i\omega+\epsilon_v-\hat{H}\right)^{-1}|O_j\rangle\langle O_j|U_G^{v'}\rangle. \tag{3.28}$$

Similar reconstructions can be obtained for $\boldsymbol{C}$ and $\boldsymbol{L}$. The evaluation of Eq. 3.28 can be performed with a $N_{ortho} \times N_{aux}^2 \times N_v^2$ scaling. Accordingly, this implementation still scales with the fifth power of the system size. However, since N_{ortho} is much smaller than N_{PW} (up to a factor 60 for the largest system considered in this work), computational time is significantly reduced and calculations of much larger systems become feasible. In the following we will use “GSO” to denote the new optimized version of our code and “original” to denote the code that does not use the orthogonalization.

To conclude this section we further discuss the specific choice of the threshold we made. Table 3.1 shows the effect of this parameter on the accuracy of the

Table 3.1: Binding energy of the formaldehyde dimer in the A24 test set as computed for different approximations (RPA, eh-TDHF, and AC-SOSEX) with different thresholds for the Gram-Schmidt orthogonalization (GSO). DT denotes the default threshold described in the text. In the last line of the table we report the total time taken to complete the calculation on an Intel Xeon mchine with 14 cores.

	Original code	GSO DT	DT$\times 10$	DT $/10$
RPA	-2.739	-2.674	-2.312	-2.728
eh-TDHF	-3.280	-3.255	-3.283	-3.263
AC-SOSEX	-3.571	-3.572	-3.589	-3.558
Time taken	28h 15m	6h 19m	6h 00m	9h 44m

binding energy of the formaldehyde dimer, the largest system in the A24 test set (see Sec. 4.1.1). Specifically, we considered the default threshold (DT), a larger threshold (DT$\times 10$), and a smaller threshold (DT$/10$) and compared the results with the original version of the code; the last line on the table reports the time taken to obtain the results on a node of our cluster (Intel Xeon CPU E5-2680 v4 @ 2.40GHz, 14 cores). Since the RPA, eh-TDHF, and AC-SOSEX methods require the evaluation of the same matrices, the three correlation energies are all provided by a single calculation and, accordingly, a single time is provided in Table 3.1. As expected, the GSO implementation significantly speed up calculations and the time decreases by increasing the threshold in the GSO procedure. The difference in the time taken by the DT and DT$\times 10$ calculations is rather small; this can be understood by considering that a large threshold significantly decreases N_{ortho} but at the same time the cost of rebuilding the response function (Eq. 3.28) increases. Discarding some small fluctuations in part related to the auxiliary basis set convergence, the original, the GSO DT, and the GSO DT/10 implementations all give results quantitatively in good agreement. Differently, the GSO DT$\times 10$ calculation leads to an error of about 0.4 kcal/mol for the RPA result. By considering these results the choice of the DT provides an excellent compromise between accuracy and computer time. In Sec. 4.1.1 we will systematically discuss the influence of the Gram-Schmidt procedure with DT on the accuracy of the binding energies of the full A24 test set.

Chapter 4

Numerical applications to molecules and solids

This chapter is based on original work previously published in Refs. [28, 31, 99, 74].

In this chapter we present a series of different applications of the methods introduced in Chapter 2 using the implementation discussed in Chapter 3. Specifically, we will consider weakly bound molecular dimers, molecular reaction energies involving the breaking and formation of covalent bonds, and bulk solids (for the latter the discussion will be limited to the RPA, since an implementation of eh-TDHF and AC-SOSEX including the first Brillouin zone integration is currently not yet available). The accuracy of RPA and beyond-RPA methods will be determined by comparison with high-accuracy reference values obtained from coupled cluster theory (molecules) and experiment (solids). It will be shown that the inclusion of exchange effects beyond the RPA significantly improves the results.

Ideally, ACFDT correlation energies should be computed self-consistently [103], as typically done for DFT approaches. In practice, these types of calculations are cumbersome and here we will rely exclusively on a fixed starting point obtained from an affordable DFT approximation. Specifically, all the results for molecules presented in this manuscript are based on response functions built from GGA orbitals and energy levels [106]. Instead, GGA-based RPA results for solids will be compared with correlation energies using a LDA starting point.

4.1 Applications to weakly bound molecular dimers

4.1.1 A24 Test Set

We present here the results for the A24 test set proposed by Řezáč and Hobza [121], which contains interaction energies for 24 molecular dimers. The results are compared with the CCSD(T) and CCSDT(Q) values from the same reference. Our calculations do not include core correlation and relativistic effects, and the reference values have been accordingly chosen. All dimers were placed in a $30 \times 30 \times 30\, a_0^3$ box, and the plane-wave cut-off was set to 80 Ry to achieve sufficient convergence. This cut-off corresponds to about 160k plane-waves. Norm-conserving FHI pseudopotentials were used [40]. For the RPA and beyond-RPA calculations, the number of auxiliary basis vectors was set to twenty times the number of electrons. Since we have non-spin polarized calculations, this corresponds to $N_{aux} = 40 \times N_v$. The number of elements included in the auxiliary basis is orders of magnitude smaller than the number of plane-waves.

We start the discussion by validating the Gram-Schmidt orthogonalization (GSO) algorithm introduced in Sec. 3.5. Since the dimers in the A24 test set are relatively small, results with and without GSO can be compared to understand the implications of this procedure on the numerical accuracy. For the S22 test set, discussed in the next subsection, the GSO algorithm is indispensable to obtain the correlation energies for the largest dimers. Figure 4.1 shows the differences between the interaction energies calculated with the original and the GSO versions. As can be observed, the differences are small, below 0.085 kcal/mol for all the molecules. As indicated in Table 4.1, the mean absolute error (MAE) of the GSO results with respect to the "exact" results is 0.047 kcal/mol for the RPA, 0.033 kcal/mol for the eh-TDHF, 0.029 kcal/mol for the AC-SOSEX approximations. Therefore, the GSO algorithm can be applied to calculate correlation energies at the RPA and beyond-RPA levels of theory for different types of systems, e.g., weakly interacting systems, which require stricter convergence criteria.

Table 4.1: Mean absolute errors (MAE) and maximum absolute deviations (|MAX|) in kcal/mol for the A24 test set introduced by applying the Gram-Schmidt orthogonalization procedure described in the text.

System	RPA	eh-TDHF	AC-SOSEX
MAE	0.047	0.033	0.029
|MAX|	0.084	0.075	0.078

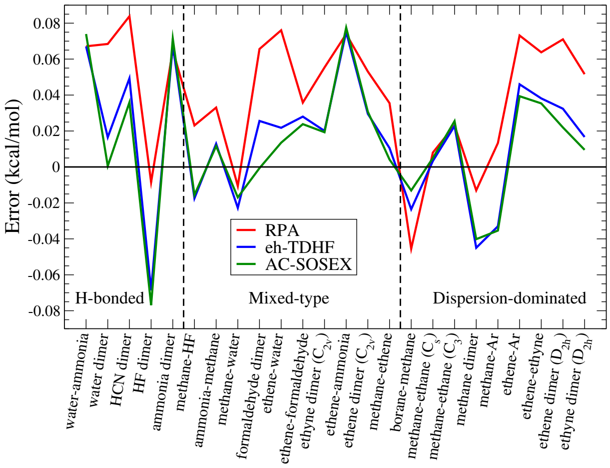


Figure 4.1: Differences in kcal/mol for the A24 test set introduced by applying the Gram-Schmidt orthogonalization procedure described in the text.

From now we will discuss exclusively results for the A24 and S22 (in the next subsection) test sets obtained using the GSO version of the code. To compare our results with CCSD(T) and CCSDT(Q) values, we used slightly more strict convergence parameters for a few systems. Specifically, N_{aux} was increased to sixty times the number of valence states for the borane-methane, ethene-Ar, methane-HF and methane-water dimers to ensure a convergence of the RPA interaction energy to within 0.09 kcal/mol.

The fully converged results are shown in Table 4.2. The maximum absolute deviation ($|\mathrm{MAX}|$) of the RPA and beyond-RPA methods from the CCSDT(Q) reference values is found for the formaldehyde dimer, with almost 1.8 kcal/mol for the RPA, 1.2 kcal/mol for the eh-TDHF and 0.9 kcal/mol for the AC-SOSEX approximations. It could also noticed that in the comparison between the CCSD(T) and CCSDT(Q), the maximum deviation was also found for the formaldehyde dimer. In general, the inclusion of approximate exchange effects in the beyond-RPA methods improve results over the RPA. This is most visible for the hydrogen-bonded systems, where the mean absolute errors for the eh-TDHF and AC-SOSEX are 0.356 kcal/mol and 0.339 kcal/mol, respectively, as compared to the value of 0.605 kcal/mol for the RPA, and for the mixed-type systems, where the mean absolute errors for the eh-TDHF and AC-SOSEX are 0.224 kcal/mol and 0.179 kcal/mol, respectively, as compared to the value of 0.442 kcal/mol for the RPA. In the case of dispersion-dominated systems, the improvement is less clear, with the mean absolute error for the eh-TDHF and AC-SOSEX are 0.242 kcal/mol and 0.314 kcal/mol respectively compared to 0.264 kcal/mol for the RPA. While these differences are small compared to numerical accuracy, for these systems only the eh-TDHF method slightly improves over the RPA. The deviations with respect to CCSDT(Q) are plotted in Fig. 4.2. Positive mean errors for the RPA and beyond-RPA methods indicate a general tendency of all these methods to consistently underestimate the absolute binding energies of the various systems. We observe that the beyond-RPA methods improve over the MP2 values reported in the literature [19], where MAEs of 0.389 kcal/mol and 0.364 kcal/mol were found for the counter-poise corrected and uncorrected case, respectively.

We also observe a good agreement of our values with earlier works. In the case of RPA, our values have a mean absolute error of 0.409 kcal/mol compared to 0.338 kcal/mol obtained by Bates and Furche [6] and 0.4 kcal/mol obtained by Grundei and Burow [54] within a localized basis set implementation. For the AC-SOSEX calculations, our results have a mean absolute error of 0.263 kcal/mol compared to 0.208 kcal/mol obtained by Bates and Furche [6].

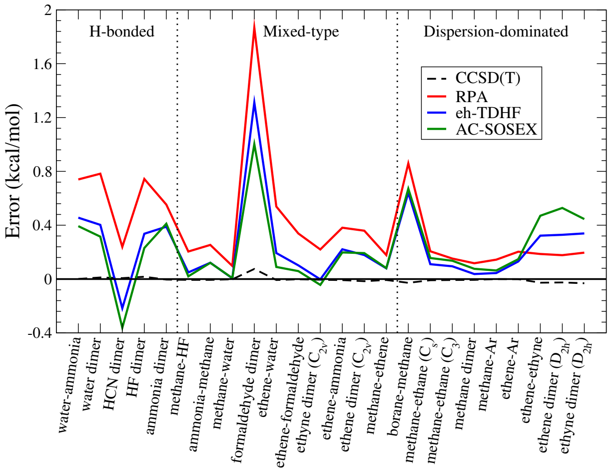


Figure 4.2: Differences in kcal/mol for the A24 test set for the RPA, eh-TDHF, and AC-SOSEX methods computed with respect to CCSDT(Q) reference values. For the sake of completeness the errors of CCSD(T) [121] are also included.

4.1.2 S22 Test Set

We apply here the more efficient algorithm based on the GSO to the S22 test set proposed by Jurečka *et al.*, which consists of interaction energies for 22 molecular dimers [73]. This test set has been very valuable for benchmarking new and existing methods for non-covalent complexes, as it contains 22 dimers of various types (H-bonded, dispersion dominated and mixed), and system sizes (from

Table 4.2: Interaction energies (in kcal/mol) for the A24 test set as obtained with the RPA, eh-TDHF, and AC-SOSEX methods compared with CCSDT(Q) and CCSD(T) reference values. The mean absolute errors (MAE), mean errors (ME) and the maximum deviations (|MAX|) are provided with respect to CCSDT(Q) values for the three types of molecular dimers and for the total test set.

	System		CCSDT(Q) [121]	CCSD(T) [121]	RPA	eh-TDHF	AC-SOSEX
			Hydrogen-bonded systems				
1	water…ammonia	C_s	−6.492	−6.493	−5.753	−6.037	−6.100
2	water dimer	C_s	−4.994	−5.006	−4.223	−4.603	−4.691
3	HCN dimer	C_s	−4.738	−4.745	−4.506	−4.962	−5.107
4	HF dimer	C_s	−4.564	−4.581	−3.837	−4.244	−4.350
5	ammonia dimer	C_{2h}	−3.141	−3.137	−2.584	−2.748	−2.726
	MAE			0.008	0.605	0.356	0.339
	ME			−0.007	0.605	0.267	0.191
	|MAX|			0.017	0.771	0.455	0.415
			Mixed-type systems				
6	methane…HF	C_{3v}	−1.660	−1.654	−1.371	−1.626	−1.651
7	ammonia…methane	C_{3v}	−0.771	−0.765	−0.511	−0.644	−0.645
8	methane…water	C_s	−0.665	−0.663	−0.510	−0.664	−0.660
9	formaldehyde dimer	C_s	−4.479	−4.554	−2.674	−3.255	−3.572
10	ethene…water	C_s	−2.564	−2.557	−2.018	−2.364	−2.467
11	ethene…formaldehyde	C_s	−1.623	−1.621	−1.282	−1.519	−1.562
12	ethyne dimer	C_{2v}	−1.529	−1.524	−1.304	−1.526	−1.568
13	ethene…ammonia	C_s	−1.382	−1.374	−0.993	−1.153	−1.176
14	ethene dimer	C_{2v}	−1.106	−1.090	−0.731	−0.912	−0.898
15	methane…ethene	C_s	−0.509	−0.502	−0.324	−0.422	−0.421
	MAE			0.013	0.442	0.224	0.179
	ME			−0.002	0.442	0.224	0.171
	|MAX|			0.075	1.792	1.237	0.927
			Dispersion-dominated systems				
16	borane…methane	C_s	−1.513	−1.485	−0.636	−0.883	−0.850
17	methane…ethane	C_s	−0.836	−0.827	−0.621	−0.717	−0.671
18	methane…ethane	C_3	−0.614	−0.607	−0.455	−0.512	−0.471
19	methane dimer	D_{3d}	−0.539	−0.533	−0.416	−0.496	−0.457
20	methane…Ar	C_{3v}	−0.405	−0.405	−0.263	−0.362	−0.343
21	ethene…Ar	C_{2v}	−0.365	−0.364	−0.165	−0.226	−0.211
22	ethene…ethyne	C_{2v}	+0.794	+0.821	+1.007	+1.143	+1.291
23	ethene dimer	D_{2h}	+0.909	+0.934	+1.112	+1.263	+1.462
24	ethyne dimer	D_{2h}	+1.084	+1.115	+1.311	+1.454	+1.561
	MAE			0.015	0.264	0.242	0.314
	ME			0.015	0.264	0.242	0.314
	|MAX|			0.031	0.887	0.668	0.699
			Total				
	MAE			0.013	0.409	0.259	0.263
	ME			0.004	0.409	0.240	0.229
	|MAX|			0.075	1.792	1.237	0.927

water dimer to adenine-thymine complexes), which has led to a rise of its popularity [68, 123, 94, 97, 132, 90, 48, 11, 91, 51, 142, 64, 34]. It has even been used to parametrize and adjust certain methods so that the relative errors with respect to the S22 values are minimized [135, 27, 113]. The kinetic energy cut-off was set to 80 Ry to achieve well-converged PBE orbitals. As with the A24 test set, the number of auxiliary basis vectors was set to forty times the number of valence states to ensure convergence, except for the formic acid ($N_{aux} = 60 \times N_v$), formamide ($N_{aux} = 60 \times N_v$), phenol ($N_{aux} = 32 \times N_v$), uracil ($N_{aux} = 32 \times N_v$) and adenine-thymine dimers ($N_{aux} = 32 \times N_v$). This ensured that the variations of the RPA and beyond-RPA interaction energies were within 0.1 kcal/mol. The results are tabulated in Table 4.3, and the errors with respect to the CCSD(T) values of Takatani *et al.* [142] are shown graphically in Fig. 4.3.

As was observed with the A24 test set, the inclusion of exchange effects generally improve the RPA interaction energy values, with the mean absolute errors for the eh-TDHF and AC-SOSEX methods around 0.43 kcal/mol and 0.42 kcal/mol respectively, compared to 0.81 kcal/mol for the RPA. Specifically, the improvement is more clear in the hydrogen-bonded systems, where the mean absolute errors for the beyond-RPA methods are 0.48 kcal/mol for the eh-TDHF and 0.36 kcal/mol for the AC-SOSEX approximations, compared to the value of 1.26 kcal/mol for the RPA, and in the mixed-type dimers, where the mean absolute errors for eh-TDHF and AC-SOSEX methods is 0.20 kcal/mol and 0.21 kcal/mol respectively, compared to 0.46 kcal/mol for the RPA. In case of dispersion-dominated systems, as was seen with the A24 test set, the improvement is more limited, where the mean absolute errors for the eh-TDHF and AC-SOSEX methods are 0.60 kcal/mol and 0.65 kcal/mol respectively compared to the value of 0.72 kcal/mol for the RPA. A similar behavior was previously noticed for the RPA+SOSEX method within localized basis sets implementations [119]. Similar to the A24 test set, we observe that all methods have positive mean errors, indicating a general tendency of these methods to consistently underestimate the absolute binding energies when compared with CCSD(T) interaction energy values. We observe that the beyond-RPA methods significantly improve over the MP2 values reported by Takatani *et al.* [142], whose MAE is 0.88 kcal/mol.

Overall, we obtain a good agreement of our RPA results with earlier works, with a mean error of 0.81 kcal/mol compared to 0.9 kcal/mol obtained by Ren *et al.* [117, 119] and 0.79 kcal/mol obtained by Eshuis and Furche [34]. The AC-SOSEX values have a mean absolute percentage error of 9.5% in close agreement with the value of 10.5% reported for a localized basis set implementation [119]. We highlight that our implementation was done within the plane-wave basis set, which can treat molecules and solids on the same footing, and does not use extrapolation techniques to obtain complete basis set-quality results (see Chapter 5).

Table 4.3: Interaction energies (in kcal/mol) for the S22 test set as obtained with the RPA, eh-TDHF, and AC-SOSEX methods compared with CCSD(T) reference values [142]. The mean absolute errors (MAE), mean errors (ME) and the maximum deviations (|MAX|) are provided with respect to CCSD(T) values [142] for the three types of molecular dimers and for the total test set.

	System		CCSD(T) [142] Takatani *et al.*	CCSD(T) [73] Jurečka *et al.*	RPA	eh-TDHF	AC-SOSEX
	Hydrogen-bonded systems						
1	ammonia dimer	C_{2h}	−3.17	−3.17	−2.54	−2.72	−2.71
2	water dimer	C_s	−5.02	−5.02	−4.20	−4.58	−4.65
3	formic acid dimer	C_{2h}	−18.80	−18.61	−17.05	−18.24	−18.60
4	formamide dimer	C_{2h}	−16.12	−15.96	−14.75	−15.57	−15.71
5	uracil dimer	C_{2h}	−20.69	−20.65	−18.78	−20.32	−21.02
6	2-pyridoxine…2-aminopyridine	C_1	−17.00	−16.71	−15.78	−16.46	−16.61
7	adenine…thymine (WC)	C_1	−16.74	−16.37	−15.63	−16.30	−16.37
	MAE			0.15	1.26	0.48	0.36
	ME			0.15	1.26	0.48	0.27
	|MAX|			0.37	1.91	0.56	0.46
	Dispersion-dominated systems						
8	methane dimer	D_{3d}	−0.53	−0.53	−0.47	−0.52	−0.48
9	ethene dimer	D_{2d}	−1.50	−1.51	−1.15	−1.33	−1.31
10	benzene…methane	C_3	−1.45	−1.50	−1.18	−1.32	−1.28
11	benzene dimer	C_{2h}	−2.62	−2.73	−1.92	−1.86	−1.70
12	pyrazine dimer	C_s	−4.20	−4.42	−3.40	−3.39	−3.24
13	uracil dimer	C_2	−9.74	−10.12	−8.61	−9.11	−9.26
14	indole…benzene	C_1	−4.59	−5.22	−3.41	−3.35	−3.19
15	adenine…thymine (stack)	C_1	−11.66	−12.23	−10.36	−10.59	−10.62
	MAE			0.25	0.72	0.60	0.65
	ME			−0.25	0.72	0.60	0.65
	|MAX|			0.63	1.30	1.24	1.40
	Mixed-type systems						
16	ethene…ethine	C_{2v}	−1.51	−1.53	−1.29	−1.50	−1.57
17	benzene…water	C_s	−3.29	−3.28	−2.89	−3.10	−3.10
18	benzene…ammonia	C_s	−2.32	−2.35	−1.91	−2.07	−2.06
19	benzene…hydrogen cyanide	C_s	−4.55	−4.46	−4.03	−4.49	−4.61
20	benzene dimer	C_{2v}	−2.71	−2.74	−2.31	−2.45	−2.42
21	indole…benzene (T-shape)	C_1	−5.62	−5.73	−5.18	−5.38	−5.37
22	phenol dimer	C_1	−7.09	−7.05	−6.31	−6.76	−6.79
	MAE			0.05	0.46	0.20	0.21
	ME			−0.01	0.46	0.20	0.17
	|MAX|			0.11	0.78	0.33	0.33
	Total						
	MAE			0.15	0.81	0.43	0.42
	ME			−0.04	0.81	0.43	0.38
	|MAX|			0.63	1.91	1.24	1.40

4.2 Binding curves

Up to this point we have only considered weakly bound dimers in their equilibrium geometry. In this subsection we consider the application of RPA, AC-

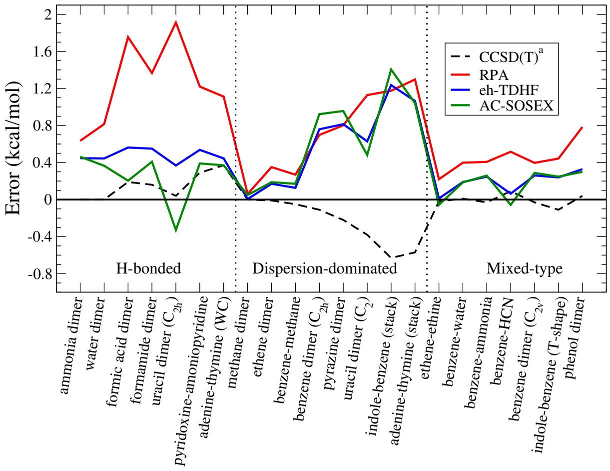


Figure 4.3: Differences in kcal/mol for the S22 test set for the RPA, eh-TDHF, and AC-SOSEX methods computed with respect to CCSD(T) reference values reported by Takatani *et al.* [142]. For the sake of completeness we included the deviations of the CCSD(T) values previously reported by Jurečka *et al.* [73]

SOSEX, and eh-TDHF to compute the full energy curve of three weakly bound dimers from the S22 test set [73]: The ethene dimer, which is categorized as a dispersion-dominated system, the water dimer, which is hydrogen bonded, and the ethene-ethyne dimer, which belongs to the mixed-type complexes. This allows us to estimate the accuracy of the different approximations also for equilibrium positions. In PW calculations a kinetic-energy cut-off of 80 Ry was used and supercell sizes were chosen to guarantee a distance of at least a 10 Å between periodically repeated images. The computed curves for the three dimers are shown in Figs. 4.4-4.6. In the case of the ethene and ethene-ethyne dimers CCSD(T) curves from Ref. 65 are shown as a reference; in the case of the water dimer we included only the CCSD(T) value at equilibrium. A summary of the equilibrium binding energies is also given in Table 4.4. While the RPA reproduces already at least 80% of the binding energy, the methods with exchange systematically improve this result. For example, eh-TDHF gives 92% of the binding energy of the ethene dimer, 99% for the ethene-ethyne dimer, and 93% for the water dimer. A similar result is obtained for the AC-SOSEX method, with a slight overestimation of the binding energy of the ethene-ethyne dimer. The RPA and beyond-RPA methods also perform well in reproducing the minimum of the binding curve, which is exactly reproduced within the 0.1 Å accuracy of the grid used to discretize the intermolecular distance. This is not the case for PBE, that for the ethene dimer predicts a too large binding distance. With exception of the water dimer, whose interaction is dominated by polar forces, PBE also significantly underestimates

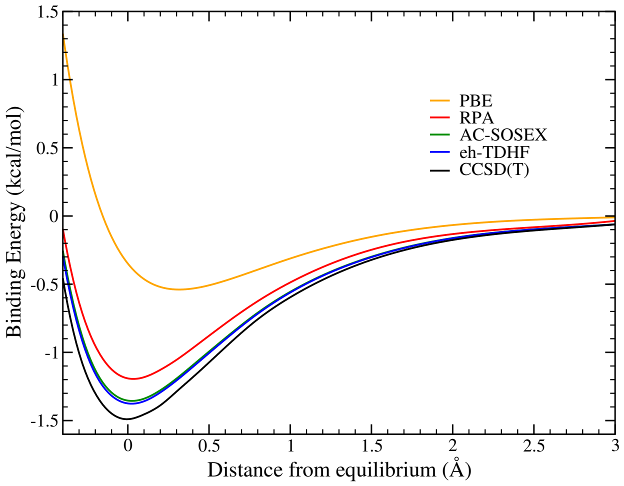


Figure 4.4: Binding curve of the ethene dimer.

binding energies.

In Table 4.4 we also compare the results obtained with the present PW implementation and the Gaussian basis set implementation that we introduced in Ref. 99. This is also useful to understand the effect of pseudopotentials in PW results. The PBE and HF binding energies were computed by using the aug-cc-pCV5Z basis set; the RPA and beyond-RPA correlation energies were extrapolated to the complete basis set (CBS) limit by using aug-cc-pCVQZ and aug-cc-pCV5Z basis sets. In this case the two different methods lead to close results with differences of the order of about 0.1 kcal/mol at most. This is not surprising, since weak interactions only marginally involve core regions and the effect of pseudopotentials is expected to be weak. While comparing the two different implementations of the RPA, AC-SOSEX, and eh-TDHF methods it is also important to keep into account that the PW results are converged within 0.03-0.04 kcal/mol with respect to the auxiliary basis set and Gaussian basis set results might be affected by errors due to the CBS extrapolation and basis set superimposition. Accordingly, the agreement between the two different implementations can be considered fully satisfactory in the case of the weakly bound dimers.

4.3 Application to reaction energies involving covalent bonds

In this section we present the results for the reaction energy test set proposed by Hesselmann [65]. These reactions involve the breaking and formation of covalent

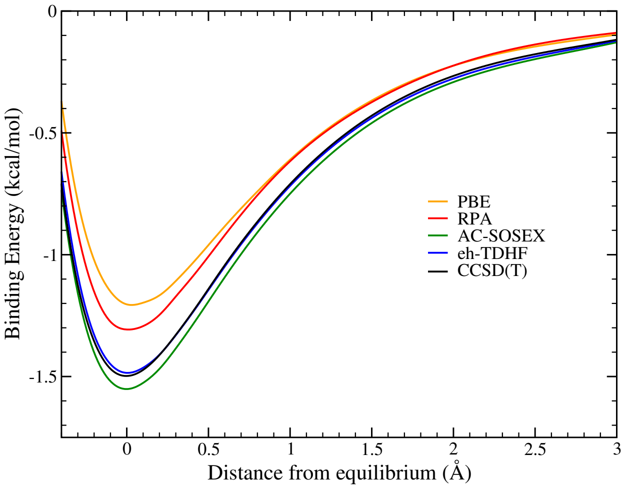


Figure 4.5: Binding curve of the ethene-ethyne dimer.

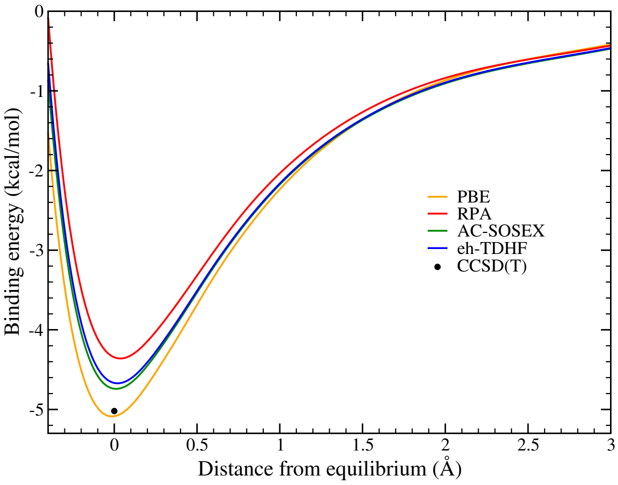


Figure 4.6: Binding curve of the water dimer.

Table 4.4: Binding energies in kcal/mol for three weakly bound dimers obtained by different approximations: PBE, HF@PBE, RPA, AC-SOSEX, and eh-TDHF. Reference CCSD(T) results are provided in the last column. Results have been obtained by using plane-wave (PW) and Gaussian basis sets (see text for discussion and comparison).

Dimer	PBE PW	HF@PBE PW	RPA PW	AC-SOSEX PW	eh-TDHF PW	PBE Gaussian	HF@PBE Gaussian	RPA Gaussian	AC-SOSEX Gaussian	eh-TDHF Gaussian	CCSD(T)
ethene dimer	-0.35	1.12	-1.19	-1.35	-1.37	-0.34	1.15	-1.11	-1.27	-1.26	-1.50
ethene-ethyne dimer	-1.20	-0.32	-1.31	-1.55	-1.49	-1.18	-0.27	-1.24	-1.47	-1.40	-1.51
water dimer	-5.08	-2.73	-4.35	-4.74	-4.67	-4.95	-2.64	-4.36	-4.70	-4.63	-5.02

bonds. Results obtained within the plane-wave framework will also be compared to values obtained using a previous localized basis set implementation that we developed [99].

In order to decrease the spurious interaction between periodically repeated images, each molecule from the reaction test set has been placed in a $26 \times 26 \times 26$ a_0^3 supercell. Norm-conserving pseudopotentials were taken from the library of Schlipf and Gygi [134], which has been generated within the optimized norm-conserving Vanderbilt pseudopotential scheme recently introduced by Hamann [57]. A kinetic-energy cut-off of 80 Ry was used to expand the orbitals.

In the columns 2-6 of Table 4.5 we show the results obtained within the PW implementation. The last three lines of this table contain the mean absolute error (MAE), the mean error (ME), and the absolute value of the largest deviation (|MAX|) in kcal/mol computed with respect to reference localized orbital CCSD(T) results [65]; these results are also pictorially represented in Fig. 4.7. Similarly to the results obtained with the Gaussian basis set of Ref. 99, the RPA significantly improves over PBE (the MAE is decreased from 4.49 kcal/mol to 3.27 kcal/mol) and this result is improved further by including the exchange contribution; in particular, the eh-TDHF is the most accurate approach with a MAE of only 1.19 kcal/mol. This is particularly promising in the context of plane-wave implementations, where access to the most accurate quantum chemical approaches are numerically challenging [15] and applications rely mostly on semi-local DFT.

We now discuss the comparison of the results obtained within the present PW implementation and the localized basis set implementation in Molpro [149, 148, 99]. For a reliable comparison with the plane-wave basis set, we used localized basis sets larger than in Ref. 99. Specifically, here the PBE and HF from a PBE starting point (HF@PBE) total energies have been computed by using the aug-cc-pCV6Z basis set. Since HF@PBE results are used to compute RPA and beyond-RPA total energies by adding the correlation term, their inclusion in the discussion is useful to better understand the origin of differences between PW and Gaussian results. The RPA and beyond-RPA correlation energies were extrapolated to the complete basis set (CBS) limit by the usual $1/X^3$ formula [79] considering aug-cc-pCVXZ basis sets with X going from Q to 6. In the columns 7-11 of Table 4.5 and in Fig. 4.8 we present the reaction energy values for the localized basis set implementation. The general trend of the errors is similar to the PW case. However, a sizeable difference in the MAE can be seen in the RPA case (which decreases from 3.27 kcal/mol for PWs to 2.42 kcal/mol for the localized basis sets) and differences are present also in the ME and |MAX|. This issue has to be ascribed to the use of norm-conserving pseudopotentials in the plane-wave basis set. Indeed, Hesselmann's test set involves strong covalent interactions and short bond lengths whose energy might be affected by the core regions of pseudopotentials. As shown in Sec. 4.1, in the case of weakly bound systems the interaction energy is mostly

Table 4.5: Reaction energies in kcal/mol for 16 chemical reactions obtained by different approximations: PBE, HF@PBE, RPA, AC-SOSEX, and eh-TDHF. The mean absolute error (MAE), mean error (ME), and maximum absolute deviation (|MAX|) are computed with respect to the CCSD(T) reference values [65] in the last column. Results have been obtained by using plane-wave (PW) and Gaussian basis sets (see text for discussion and comparison).

Reaction	PBE	HF@PBE	RPA	AC-SOSEX	eh-TDHF	PBE	HF@PBE	RPA	AC-SOSEX	eh-TDHF	CCSD(T)
Basis set	PW	PW	PW	PW	PW	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
C_2H_2+H_2 → C_2H_4	-52.08	-50.49	-46.96	-51.40	-50.55	-52.14	-50.97	-47.63	-51.78	-51.38	-49.44
C_2H_4+H_2 → C_2H_6	-40.53	-39.12	-37.20	-42.19	-39.90	-40.59	-39.60	-37.65	-42.49	-40.36	-39.47
C_2H_6+H_2 → $2CH_4$	-18.71	-22.18	-17.39	-19.13	-18.62	-18.61	-21.99	-17.23	-18.88	-18.32	-18.18
CO+H_2 → HCHO	-11.97	2.57	-1.17	-3.66	-3.41	-12.36	1.55	-2.63	-4.72	-4.60	-5.47
HCHO+H_2 → CH_3OH	-29.52	-29.53	-25.61	-31.96	-29.66	-29.60	-30.20	-26.47	-32.53	-30.74	-29.70
H_2O_2+H_2 → $2H_2O$	-82.37	-97.51	-81.05	-92.07	-88.31	-81.82	-96.95	-82.10	-92.43	-89.37	-87.63
C_2H_2+H_2O → CH_3CHO	-44.05	-34.91	-38.09	-39.58	-39.07	-44.19	-35.39	-38.37	-39.90	-39.51	-38.28
C_2H_4+H_2O → C_2H_5OH	-15.37	-9.38	-13.08	-15.45	-13.94	-15.57	-10.06	-13.50	-15.92	-14.56	-14.12
CH_3CHO+H_2 → C_2H_5OH	-23.39	-24.96	-21.95	-27.27	-25.41	-23.52	-25.64	-22.76	-27.80	-26.42	-25.28
CO+NH_3 → $HCONH_2$	-20.89	3.96	-4.75	-8.18	-7.48	-20.92	2.98	-6.55	-9.71	-9.28	-10.26
CO+H_2O → CO_2+H_2	-16.76	7.64	-4.61	-2.92	-3.16	-17.55	7.06	-5.57	-4.08	-4.08	-6.18
HNCO+NH_3 → NH_2CONH_2	-18.95	-19.45	-16.68	-22.33	-20.70	-18.40	-19.87	-17.37	-22.87	-22.00	-20.70
CO+CH_3OH → $HCOOCH_3$	-21.84	3.65	-8.68	-11.21	-10.71	-22.31	2.75	-10.03	-12.45	-12.10	-13.59
HCOOH+NH_3 → $HCONH_2$+H_2O	-2.28	-1.83	-0.52	-1.63	-1.39	-1.75	-1.69	-0.76	-1.72	-1.56	-1.16
CO+H_2O_2 → CO_2+H_2O	-99.13	-89.87	-85.66	-94.99	-91.47	-99.37	-89.90	-87.67	-96.51	-93.45	-93.81
H_2CCO+HCHO → C_2H_4O+CO	5.21	-9.21	-6.29	-6.45	-5.70	5.04	-9.22	-5.30	-5.83	-5.43	-3.83
MAE	4.49	5.53	3.27	2.02	1.19	4.64	5.19	2.42	1.97	1.10	
ME	-2.22	2.91	2.96	-0.83	0.48	-2.28	2.49	2.22	-1.41	-0.38	
|MAX|	10.63	17.24	8.15	4.44	3.02	11.37	16.34	6.14	4.80	2.10	

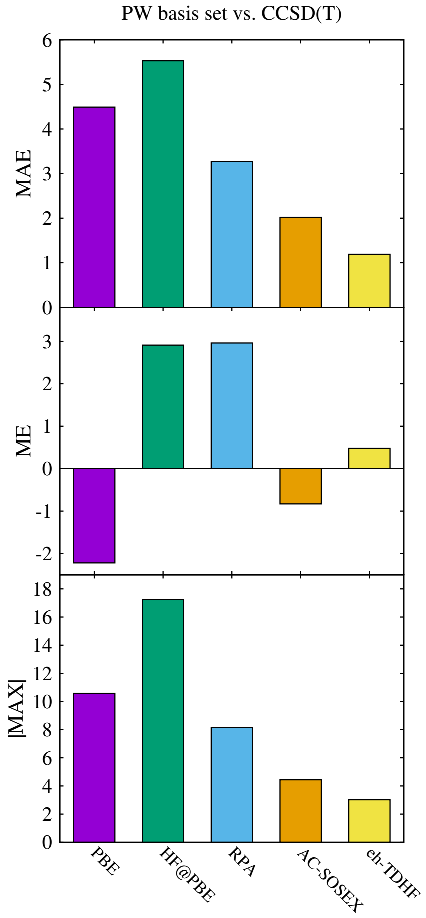


Figure 4.7: Mean absolute error (MAE), mean error (ME), and maximum absolute deviation (|MAX|) for a test set of 16 reaction energies [65]. Errors have been computed for PBE, HF@PBE, RPA, AC-SOSEX, and eh-TDHF results obtained within the pane-wave (PW) basis set implementation with respect to CCSD(T) reference values.

determined by long range forces far from the core region and pseudopotentials weakly affect the results.

To better understand the PW-Gaussian basis set comparison in Fig. 4.9 we present the MAE, ME, and |MAX| of the PW basis set results with respect to the corresponding localized basis set results. The first important observation is that errors increase going from PBE, to HF@PBE, and to ACFDT methods. This is not surprising, since the pseudopotentials have been optimized to provide accurate PBE results [134]. Another important point is that the MEs for HF@PBE and for ACFDT methods are positive and close in value to the corresponding MAE values; this indicates that the error induced by the pseudopotentials is rather systematic and tends to underestimate the absolute value of reaction energies. This explains why only in the RPA case the PW MAE with respect to CCSD(T) sizeably increases with respect to Gaussian basis set results: Since the RPA approach already presents a systematic underestimation of the absolute value of reaction energies, the error induced by the pseudopotential accumulates over this intrinsic error of the RPA. All the other approaches (PBE, HF@PBE, AC-SOSEX, and eh-TDHF) have more erratic behaviors: The significant difference between the MAE and ME in Figs. 4.7-4.8 implies that these methods sometimes overestimate and sometimes underestimate reaction energies. Accordingly, the effect of the pseudopotential is mostly compensated when computing the MAE. While the effect of pseudopotentials in the PW implementation is relatively large with respect to the level of accuracy of the AC-SOSEX and eh-TDHF methods, by considering MAE and |MAX| in Fig. 4.7 it is possible to clearly appreciate the improvement of these approaches with respect to the RPA.

The pseudopotential approximation affects HF@PBE and ACFDT results mainly in two ways: (1) The lack of valence-core contributions; (2) The use of pseudo-orbitals to compute the HF exchange and the response functions. However, it is important to mention that also the accuracy of correlation energies computed within a Gaussian basis set might be affected to a certain extent by the CBS extrapolation.

In principle, it might be possible to improve the comparison of PW basis set results with Gaussian basis set results by using the projector augmented-wave method [13]. However, within this framework the formalism is significantly more involved and approximations are typically introduced to compute response functions.

4.4 Applications to solids

In this section we present applications of the RPA within the implementation of Chapter 3 to compute the structural properties of covalent solids (C in the dia-

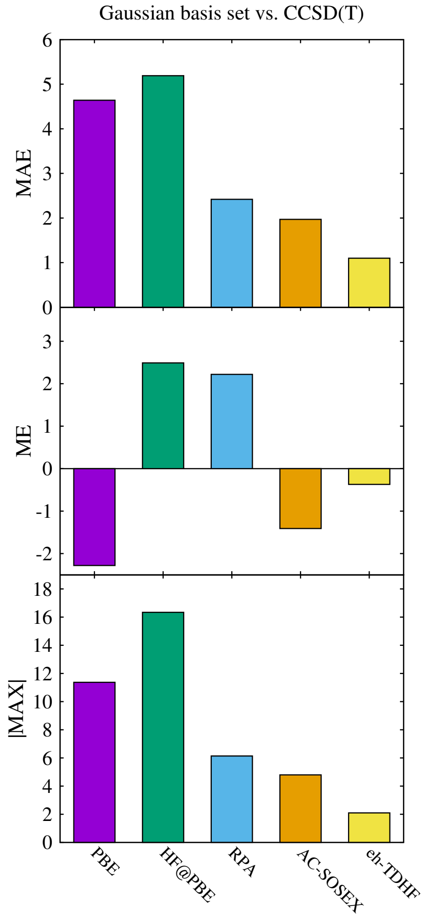


Figure 4.8: Mean absolute error (MAE), mean error (ME), and maximum absolute deviation (|MAX|) for a test set of 16 reaction energies [65]. Errors have been computed for PBE, HF@PBE, RPA, AC-SOSEX, and eh-TDHF results obtained within the Gaussian basis set implementation with respect to CCSD(T) reference values.

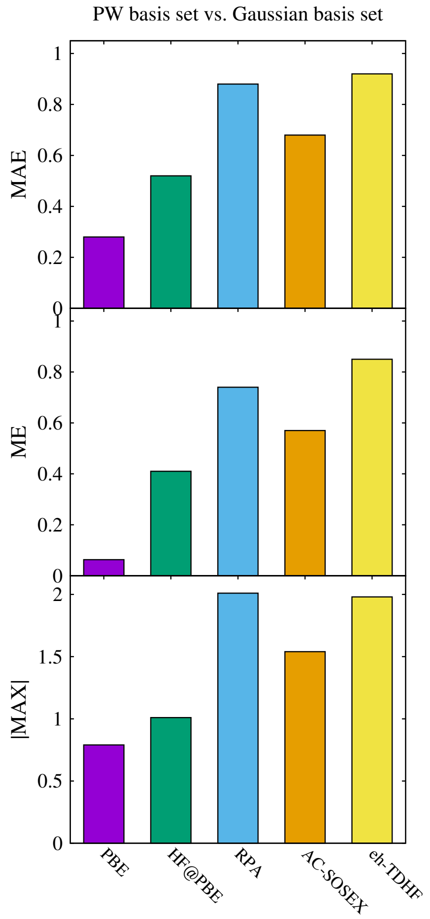


Figure 4.9: Mean absolute error (MAE), mean error (ME), and maximum absolute deviation (|MAX|) for a test set of 16 reaction energies [65]. Errors have been computed for PBE, HF@PBE, RPA, AC-SOSEX, and eh-TDHF results obtained within the plane-wave (PW) basis set implementation with respect to values obtained within the Gaussian basis set implementation of the same methods.

mond structure, silicon, and silicon carbide) and of weakly bonded solids (Ne, Ar, and Kr). While this set of bulk materials is not fully comprehensive, it provides a test-bed to prove accuracy and efficiency of the methodology discussed in this work. The discussion will be limited to the RPA, since an implementation of eh-TDHF and AC-SOSEX for solids is currently not yet available. The calculations were performed with norm-conserving pseudopotentials from the Quantum Espresso website [47, 46]. Both LDA [108] and GGA [106] have been considered as starting points for the RPA calculations (these approximations will be denoted as RPA@LDA and RPA@GGA). The convergence with respect to the kinetic energy cut-off and the k-point and q-point grids have been tested carefully[1]. All the total energy curves were obtained by interpolating the total energy computed for at least seven different lattice parameters (up to eleven for Ne) with the Murnaghan equation of state (EOS) [98]. For all the materials studied it was found that $N_{aux} = 250$ is sufficient to obtain accurate results, with the exception of solid Ne, for which the size of the auxiliary basis set was increased up to 480. The use of the auxiliary basis set was instrumental to reach convergence without extrapolation techniques (see also Chapter 5). For example, the dimensionality reduction is particularly important for noble gas solids, with N_{PW}/N_{aux} up to 150 for neon and up to 120 for argon and krypton.

In Tables 4.6 and 4.7 we show the results for lattice constants and bulk moduli, respectively, for the six bulk materials considered in this work. For comparison purposes we also provide experimental (see Refs. 139, 129 and references therein) and other results based on the RPA in the literature [58, 60, 84].

In general, compared to experimental values, the RPA preserves the good level of accuracy of the LDA and GGA for covalently bonded systems [60]. In the case of silicon the RPA@LDA lattice constant is slightly less accurate than the LDA lattice constant (the 0.6% underestimation by LDA increases up to 1.5% for RPA@LDA). Previous RPA and RPA+ calculations based on pseudopotentials had already shown a similar trend of RPA@LDA to underestimate the lattice constant of bulk silicon [100, 43, 84]; in particular, as shown in Table 4.6, the RPA result is in excellent agreement with our value [84]. In Ref. 60 a RPA@LDA lattice constant was obtained in perfect agreement with experiments; the discrepancy with respect to our results might be due to the details of the implementation,

---

[1] Converged parameters: RPA@LDA for C: 70 Ry (kinetic energy cut-off for wavefunctions), $6 \times 6 \times 6$ (k-point grid), $4 \times 4 \times 4$ (q-point grid shifted from the origin); RPA@GGA for C: 70 Ry, $6 \times 6 \times 6$, $4 \times 4 \times 4$; RPA@LDA for Si: 40 Ry, $6 \times 6 \times 6$, and $4 \times 4 \times 4$; RPA@GGA for Si: 40 Ry, $6 \times 6 \times 6$, and $4 \times 4 \times 4$; RPA@LDA for SiC: 70 Ry, $6 \times 6 \times 6$, and $4 \times 4 \times 4$; RPA@GGA for SiC: 70 Ry, $6 \times 6 \times 6$, and $4 \times 4 \times 4$; RPA@LDA for Ne: 160 Ry, $4 \times 4 \times 4$, and $4 \times 4 \times 4$; RPA@GGA for Ne: 160 Ry, $4 \times 4 \times 4$, and $4 \times 4 \times 4$; RPA@LDA for Ar: 60 Ry, $4 \times 4 \times 4$, and $4 \times 4 \times 4$; RPA@GGA for Ar: 80 Ry, $4 \times 4 \times 4$, and $4 \times 4 \times 4$; RPA@LDA for Kr: 70 Ry, $4 \times 4 \times 4$, and $4 \times 4 \times 4$; RPA@GGA for Kr: 80 Ry, $4 \times 4 \times 4$, and $4 \times 4 \times 4$.

Table 4.6: Lattice parameter as computed within different approximations (LDA, GGA, RPA@LDA, and RPA@GGA) for six different materials. Experimental values and previous RPA results in the literature are also provided for reference.

Method	C	Si	SiC	Ne	Ar	Kr
LDA	6.67	10.20	8.21	7.27	9.32	10.09
RPA@LDA	6.68	10.11	8.17	8.97	10.23	10.83
GGA	6.73	10.33	8.26	8.69	11.22	12.12
RPA@GGA	6.73	10.21	8.21	8.18	10.00	10.72
RPA@LDA Literature	6.75[a]	10.27[a], 10.12[b]	/	8.88[c]	10.20[c]	10.96[c]
RPA@GGA Literature	6.75[a]	10.26[a]	8.25[a]	8.50[c]	10.02[c]	10.77[c]
Experiments	6.74[d]	10.26[d]	8.24[d]	8.22[e]	9.88[e]	10.60[e]

[a] Ref. 60.

[b] Ref. 84.

[c] Ref. 58.

[d] Experimental lattice constants extrapolated at 0 K. See Ref. 139 and references therein.

[e] Experimental lattice constants extrapolated at 0 K. See Ref. 129 and references therein.

since in this previous calculation a code based on the projector augmented wave (PAW) method was used and the RPA correlation energy was obtained by extrapolation techniques. When GGA is used as a starting point for the RPA calculation (RPA@GGA), the lattice constant of silicon is improved with respect to both LDA and GGA. As shown in Table 4.7, the RPA systematically improves the bulk modulus of covalent solids with respect to traditional (semi-)local functionals.

The case of noble gas solids is certainly more interesting. In this case, the interaction is dominated by weak dispersion forces and the RPA strongly improves the results of LDA and GGA [58]. In general, as shown in Table 4.6, the LDA importantly underestimates experimental values of the lattice constant (LDA relative error with respect to experiments: -11.6% for Ne, -5.7% for Ar, and -4.8% for Kr) while the GGA overestimates them (GGA relative error: +5.6% for Ne, +13.6% for Ar, and +14.3 % for Kr). Since the RPA accurately describes weak dispersion forces, the results for lattice constants are significantly improved both using LDA and GGA as a starting point (RPA@LDA relative error: +9.1% for Ne, +3.5% for Ar, and +2.2% for Kr; RPA@GGA relative error: -0.5% for Ne, +1.2% for Ar, and +1.1% for Kr). As shown in Table 4.7, the RPA strongly improves also the bulk moduli of noble gas solids. For example, LDA strongly overestimates the bulk moduli (LDA relative error: +657% for Ne, +182% for Ar, and +70% for Kr) while overall RPA@GGA gives results much closer to experiments (RPA@GGA relative error: +35% for Ne, +17% for Ar, and -17% for Kr). In general, while LDA and GGA lead to very different structural properties, this behavior is less

Table 4.7: Bulk modulus as computed within different approximations (LDA, GGA, RPA@LDA, and RPA@GGA) for six different materials. Experimental values and previous RPA results in the literature are also provided for reference.

Method	C	Si	SiC	Ne	Ar	Kr
LDA	449.7	92.10	216.1	8.25	6.71	6.15
RPA@LDA	441.1	102.1	222.4	0.26	1.58	2.24
GGA	416.2	85.70	204.4	1.17	0.70	0.61
RPA@GGA	417.0	97.30	215.0	1.47	2.79	2.98
RPA@LDA Literature	439[a]	98[a], 103[b]	/	/	/	/
RPA@GGA Literature	441	99	223	/	/	/
Experiments	443.0[c]	99.2[c]	225.0[c]	1.09[d]	2.38[d]	3.61[d]

[a] Ref. 60.

[b] Ref. 84.

[c] See Ref. 139 and references therein.

[d] See Ref. 129 and references therein.

pronounced for the RPA, that gives much closer results to experiment starting from LDA or GGA states.

For comparison purposes, in Tables 4.6 and 4.7 we included results from previous RPA calculations available in the literature [58, 60, 84]. The overall agreement of our results with previous work is good. However, a straightforward comparison is not possible since the majority of previous results were obtained with an implementation based on PAW and extrapolation techniques [58, 60]. Furthermore, bulk moduli of noble gas solids have not been computed before at the RPA level of theory.

By considering the examples shown in this section we can conclude, in agreement with previous work [58, 60], that the RPA preserves the accuracy of (semi-)local functionals for covalently bonded solids, such as C, Si, and SiC, and strongly improves the structural properties of weakly bonded noble gas solids. In particular, the RPA approximated starting from GGA orbitals has been shown to have an impressive accuracy for the structural properties of the systems considered in this work.

Chapter 5

Auxiliary basis set convergence

This chapter is based on original work previously published in Ref. [29].

In this chapter we discuss in detail a point that was only briefly addressed in Chapter 3, the technique used to achieve basis set convergence of correlation energies. Differently from semi-local or hybrid functionals, the ACFDT approaches are characterized by a slow convergence with respect to this parameter. This behavior is analogous to that of traditional quantum chemical methods. Indeed, convergence of the correlation energy is known to be slow in correlated methods such as Møller-Plesset perturbation theory or coupled cluster theory [136, 62, 56, 36, 121, 122, 95, 4]. In order to overcome this issue, techniques to extrapolate the results to the complete basis set (CBS) limit are routinely used [79, 56, 122].

In Chapter 4 we did not rely on any complete basis set extrapolation technique for the the application of RPA or beyond-RPA methods, but we rather converged energy differences. This is a peculiar feature of our implementation that takes advantage of a compact auxiliary basis set which is designed to rapidly converge the eigenvalues of the response functions [150, 124]. In this chapter we give a detailed description of the methodology that we have been using so far, and compare it to a more traditional approach, the complete basis set extrapolation.

5.1 Convergence of energy differences vs. complete basis set extrapolation

To understand how correlation energy differences are converged, a simple observation is necessary, illustrated by a simple example. Let us suppose that we are interested in computing the interaction energy of the argon dimer at the RPA level of theory [101, 86]. We first consider the two atoms at a very large distance so that

$$E^{int} = E^{dimer} - 2E^{atom} \approx 0, \tag{5.1}$$

where E^{int} is the interaction energy, E^{dimer} is the total dimer energy, and E^{atom} is the total atom energy. Eq. 5.1 is a consequence of the size consistency of the RPA [49]. Clearly, each single contribution to the energy should be exactly compensated for, including the correlation energy: $E_c^{int} = E_c^{dimer} - 2E_c^{atom} \approx 0$. By using a single eigenvalue of the polarizability for each frequency $\lambda_1(i\omega)$ (the largest in absolute value), the RPA correlation energy (Eq. 3.10) of each single argon atom is

$$E_c^{atom} = \frac{1}{2}\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi}\left[\ln\left\{1 - \lambda_1(i\omega)\right\} + \lambda_1(i\omega)\right]. \tag{5.2}$$

In order to obtain $E^{int} \approx 0$ it is now clear that the correlation energy of the dimer should be computed including two eigenvalues per frequency in the correlation energy expression:

$$E_c^{dimer} = \frac{1}{2}\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi}\left[\ln\left\{1 - \lambda_1(i\omega)\right\} + \lambda_1(i\omega) + \ln\left\{1 - \lambda_2(i\omega)\right\} + \lambda_2(i\omega)\right]. \tag{5.3}$$

Indeed, since the interatomic distance is large, we have $\lambda_2(i\omega) \approx \lambda_1(i\omega)$. This error cancellation would have not been possible if we would have naively used the same number of eigenvalues for the atoms and the dimer. The same idea can be also applied to small interatomic distances; there should be a compensation between the eigenvalue contributions of the dimer and the monomers to induce error cancellation. Generally, at small or intermediate distances it is necessary to significantly increase the number of eigenvalues N_{eig} included in the evaluation of the correlation energy to reach convergence, while maintaining the condition $N_{eig}^{dimer} = 2N_{eig}^{monomer}$. Henceforth, we refer to this approach as the size-consistent energy difference approach.

This idea can also be generalized to systems, whose components are different (e.g. a dimer of two different molecules), or which involve different types of interactions. When the PDEP technique was first introduced it was already understood that the number of eigenvalues necessary to reach a certain level of accuracy in the description of the response function is approximately proportional to the number of electrons [150]. Accordingly, the number of eigenvalues (namely of auxiliary basis vectors) used to describe the response function should be chosen to be proportional to the number of electrons. For example, if in a certain molecular dimer, the monomer 1 has double the number of electrons than the monomer 2, in the correlation energy equations we should always use $N_{aux}^{monomer1} = 2N_{aux}^{monomer2}$ and $N_{aux}^{dimer} = N_{aux}^{monomer1} + N_{aux}^{monomer2} = 2N_{aux}^{monomer2} + N_{aux}^{monomer2}$. As shown by the numerical results in Sec. 5.2 this approach works well in practice.

All this discussion has been presented for the RPA. The eh-TDHF and AC-SOSEX correlation energies have a similar structure to the RPA (see Eqs. 3.9 and 3.12), but contain additional exchange terms. As shown in Eqs. 3.6-3.8, the additional exchange contributions have a structure similar to the response function in Eq. 3.5, the main difference being related to the $v \neq v'$ terms. It was already observed that the PDEP basis set, specifically designed for the polarizability (Eq. 3.5), represents well also the exchange terms (Eqs. 3.6-3.8) [28, 31]. Because of these affinities between the RPA and the beyond-RPA methods discussed in this work, the same convergence techniques can be used in practical calculations for beyond-RPA methods as well.

By using the size-consistent energy difference technique, the interaction energy converges quite rapidly with respect to the auxiliary basis set size. As an example, we present in Fig. 5.1 the interaction energy of the formaldehyde dimer computed at the RPA and beyond-RPA levels of theory. The convergence with respect to N_{aux} is quite rapid, with the beyond-RPA energies seeming to converge faster than RPA. Since the measurable physical observables are energy differences, convergence of the interaction energy with respect to the auxiliary basis set size is sufficient for practical applications. Empirically we have observed that $N_{aux} = 40 \times N_v$ is sufficient to achieve convergence for most molecular systems, where N_v is the number of Kohn-Sham states (half the number of valence electrons) in the system. This is a very small number compared to the number of plane-waves ($N_{PW} = 163000$ for the formaldehyde dimer), and convergence of the interaction energy can be achieved without resorting to extrapolation techniques. However, if we consider the correlation energies of the formaldehyde dimer and of the monomer separately, convergence is far from achieved. For example, by increasing the basis set from $35 \times N_v$ to $70 \times N_v$, the correlation energy decreases by about 10 kcal/mol for the dimer and by about 5 kcal/mol for the monomers; this leads to the error cancellation shown in Fig. 5.1.

It is important to mention that the size-consistent energy difference technique can in principle be used with different types of auxiliary basis sets. For example, plane-waves are traditionally used to represent dielectric matrices. For a given size of the dielectric matrix, the accuracy of the eigenvalues in the PW representation is lower than for the kinetic-PDEP auxiliary basis sets and, accordingly, also the energy differences less accurate. This implies that the use of PWs or other basis set, while compatible with the energy difference method, might lead to slower convergence.

The approach described in this section to converge energy differences has been previously used for homogeneous [101, 86, 124] and heterogeneous systems [28,

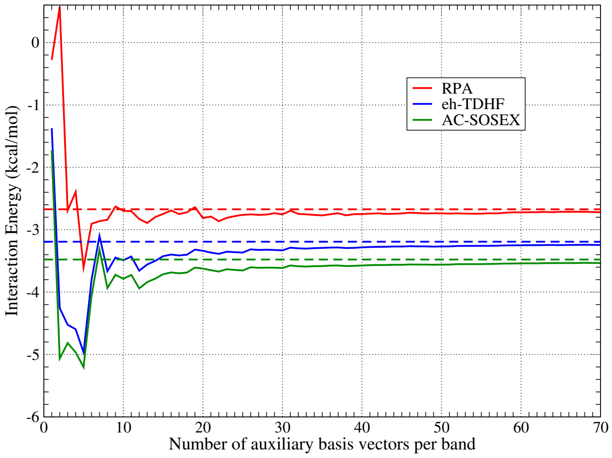


Figure 5.1: Interaction energy (kcal/mol) of the formaldehyde dimer at the RPA (red), eh-TDHF (blue) and AC-SOSEX (green) levels of theory as a function of N_{aux}/N_v. Dashed lines indicate the CBS interaction energy values computed with Eq. 5.4.

30, 31]. Here we will systematically study its accuracy for weakly bound systems and reaction energies involving the breaking and formation of covalent bonds (see Sec. 5.2).

The complete basis set extrapolation scheme for the RPA and beyond-RPA correlation energies used here is,

$$E_c^{N_{aux}} = A + \frac{B}{N_{aux}}, \tag{5.4}$$

where $A = E_c^{\infty}$ is the CBS correlation energy and B is a system-dependent parameter. As shown in Fig. 5.2 for the formaldehyde dimer, the function in Eq. 5.4 provides a reasonable description of the behavior of the correlation energy as a function of N_{aux}. An analogous expression with an additional C/N_{aux}^2 term was previously used to compute the interaction energy of molecular crystals [85]. The higher order contribution can be discarded for sufficiently high values of N_{aux} (this is the case for the results of Sec. 5.2). In Ref. 85, the PDEP method was used to exactly diagonalize the dielectric matrix for each different value of the frequency. By comparison our approach is approximated, as the basis set generated in the static case is used in the dynamical case as well.

Beyond working well in practice, Eq. 5.4 is similar to the function routinely used in the quantum chemistry community to extrapolate the correlation energy of approximations based on Møller-Plesset perturbation theory and coupled cluster theory. In these approaches the correlation energy is extrapolated to the infinite

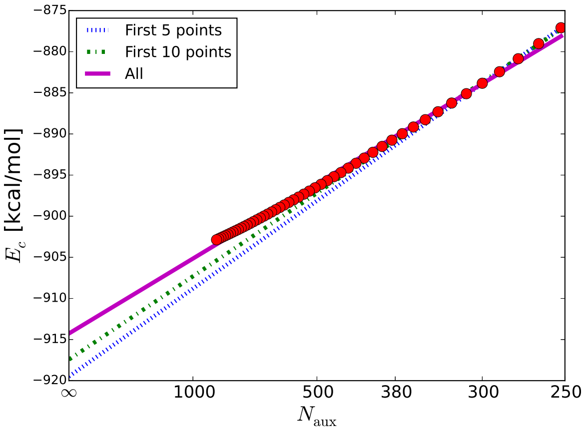


Figure 5.2: Correlation energy (red dots) of the formaldehyde dimer at the RPA level of theory as a function of $1/N_{aux}$. The points correspond to the range from $N_{aux} = 21 \times N_v$ to $N_{aux} = 70 \times N_v$. The continuous line denotes a linear fit of the data. The dotted and dot-dash lines show partial fits, demonstrating how rapidly the approach converges.

basis set limit as [136, 62, 56, 36],

$$E_c^X = E_c^\infty + \frac{B}{X^3}, \tag{5.5}$$

with X denoting the cardinal number of the basis set, E_c^∞ being its "true" correlation energy evaluated at the complete basis set limit, and B being a system-, and method-dependent constant. The number of the basis set functions is typically

$$N_X = \frac{1}{3}(X+1)\left(X+\frac{3}{2}\right)(X+2), \tag{5.6}$$

namely $N_X \propto X^3$, and for sufficiently large values of N_X we get to leading order

$$E_c^X = E_c^\infty + \frac{B'}{N_X}, \tag{5.7}$$

which has a form similar to Eq. 5.4. As with B above, B' depends on the system and method.

Alternative extrapolation techniques used in the quantum chemistry community include expressions with variable exponents such as [121, 122],

$$E_c^X = E_c^\infty + \frac{B}{X^C}, \tag{5.8}$$

where, typically, small deviations are observed from $C = 3$ (Eq. 5.5).

In standard implementations using plane-wave basis sets, the RPA correlation energy is often extrapolated to CBS as [58],

$$E_c^{|\boldsymbol{G}|_{cut}} = E_c^{\infty} + \frac{B}{|\boldsymbol{G}_{cut}|^3} = E_c^{\infty} + \frac{B'}{N_{PW}}, \tag{5.9}$$

where $|\boldsymbol{G}_{cut}|$ is the maximum reciprocal lattice vector used to represent the response function and N_{PW} is the number of plane-waves. This is different with respect to Eq. 5.7, where N_X is the basis set used to describe the full electronic structure and not just the dielectric matrices (which are not even used in typical coupled-cluster or Møller-Plesset perturbation theory formulations). Our implementation is also based on a plane-wave basis set but the response functions are represented on a system-dependent optimized basis set. Because of this difference, the range of values of N_{aux} used to apply the extrapolation function in Eq. 5.4 might be different from the range of values of N_{PW} used for the extrapolation function in Eq. 5.9. Since the basic numerical framework is the same, the two implementations should give indistinguishable results in the CBS limit.

To complete the literature review, we mention that Björkman *et al.* [12] and Gulans [55] used a different approach to converge RPA results for layered materials like hexagonal boron nitride and molybdenum di-sulphide. In these works it was observed that the correlation energy differences of layered materials converge as $|\boldsymbol{G}_{cut}|^{-5}$, and this observation was used to obtain CBS results. By using this approach it was possible to obtain accurate CBS interaction energies with much lower kinetic energy cut-offs ($\sim$ 100-150 eV). However, it is not clear if this methodology is valid in general for different classes of materials, or is a specific property of spatially infinite planar solids. Though attractive, such behavior of correlation energy differences was not observed in our calculations.

Both approaches based on the CBS extrapolation and on the size-consistent energy differences present advantages and disadvantages. In the case of the CBS extrapolation, the choice of the model is to a certain extent arbitrary (for example, Eq. 5.5 or Eq. 5.8); additionally, the final result might be affected by small fluctuations in the correlation energies used for the fitting. On the plus side this approach can be applied to any kind of system. By contrast, the alternative technique proposed here is not biased by the model chosen, but is based on reasonable assumptions about size-consistency. However, this approach might suffer from slow convergence when energy differences are computed between systems that have a very different dielectric eigenvalue structure.

Let us consider an extreme example: the interaction between a small molecule ($monomer1$), with only one occupied state, and a larger molecule ($monomer2$), with nine occupied states. Specifically, we suppose that $\lambda_{400}^{monomer2}(i\omega) > \lambda_{1}^{monomer1}(i\omega)$, where $\lambda_i(i\omega)$ indicates the i^{th} dielectric eigenvalue for each system. If we compute the RPA correlation energy for the corresponding dimer when the two molecules are separated by a large distance, by using just 400 eigenvalues ($40 \times N_v$), we would have,

$$\begin{aligned} E_c^{dimer} &= \frac{1}{2}\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \sum_i^{400} \left[\ln\left\{1 - \lambda_i^{dimer}(i\omega)\right\} + \lambda_i^{dimer}(i\omega)\right] \\ &= \frac{1}{2}\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \sum_i^{400} \left[\ln\left\{1 - \lambda_i^{monomer1}(i\omega)\right\} + \lambda_i^{monomer1}(i\omega)\right] ; \end{aligned} \tag{5.10}$$

it is clear that in this case, the energy of the dimer is fully dominated by the eigenvalues of the larger system. In other words, there are no eigenvalues of $monomer2$ present in the eigenvalue listing of the dimer, and no error cancellation can be obtained for $N_{aux} = 40 \times N_v$. Accordingly, convergence can be achieved only for very large values of N_{aux}. This slow convergence could also appear in cases where the correlation energy difference is computed between two systems and a corresponding complex containing the same atoms rearranged in a significantly different structure with a significantly different dielectric eigenvalues. In this context the reaction energy test set discussed in Sec. 5.2.C could represent a problematic example, as it contains reactants and products with different molecular structures and possibly significantly different dielectric spectrum; still the size-consistent energy difference method provides accurate estimates of the reaction energies. In general, the size-consistent energy difference approach allows for a systematic study of convergence as a function of N_{aux}, and problematic cases can be identified and potentially dealt with, a significant advantage over the the CBS technique.

In the following section, we present the RPA and beyond-RPA interaction energy results for the A24 test set [121], for the methane-formaldehyde complex (MFC), and for reaction energies as computed for different auxiliary basis set sizes (N_{aux}) with and without the extrapolation technique of Eq. 5.4. The A24 and reaction energy test sets were already considered in Chapter 4; here the discussion will be limited exclusively to the convergence properties with respect to the auxiliary basis set.

5.2 Numerical examples

5.2.1 A24 Test Set

To verify the energy differences, and compare the values with those obtained with the CBS extrapolation technique, we start by considering the A24 test set proposed by Řezáč *et al.*, which provides interaction energies for 24 molecular dimers of three types – hydrogen-bonded, dispersion-dominated and mixed-type [121]. A comparison of the RPA and beyond-RPA methods with the reference CCSDT(Q) values has been discussed in Chapter 4, and is not repeated here. Here we focus on estimating the inaccuracies due to the use of an incomplete auxiliary basis set; as the molecules included in the A24 are relatively small, a systematic study is possible by increasing N_{aux} to large values. For establishing a reference, the RPA and beyond-RPA calculations were first performed using an auxiliary basis set size of $70 \times N_v$ (in Chapter 4 this number was limited to $40 \times N_v$ for most systems) and then the absolute values of the correlation energies were extrapolated to CBS using Eq. 5.4 with 30 points from $41 \times N_v$ to $70 \times N_v$. Although small numerical fluctuations might still be present, this procedure, by including a large number of points in the fitting and a large N_{aux}, provides reasonably accurate reference values to study the convergence of RPA and beyond-RPA methods.

The choice of the $41N_v - 70N_v$ CBS value as a reference will also be used for the other numerical examples considered in the following subsections. As we will consider below values from size-consistent energy differences and CBS extrapolations from different intervals of N_{aux}, it is important to keep in mind that our specific choice of the reference introduces a bias in favor of CBS results, which in general will be characterized by a more regular and predictable convergence.

In the following we study the level of convergence of CBS values obtained by extrapolating 20 points in three different ranges: $N_{aux}/N_v =$ 11-30, 21-40, and 31-50. We compare these results with size-consistent energy differences computed at $N_{aux}/N_v =$ 30, 40, and 50 by using the procedure discussed in Sec. 5.1. For convenience, we denote with $E_A^{CBS(n,m)}$ the CBS interaction energy computed by extrapolating 20 correlation energies between $N_{aux} = n \times N_v$ and $m \times N_v$ at the A level of theory ($A =$ RPA, eh-TDHF, or AC-SOSEX); more specifically, Eq. 5.4 was used to obtain CBS correlation energies for each system (dimer and monomer) and these values were then used to compute correlation energies. Similarly, $E_A^{N_{aux}=nN_v}$ corresponds to the non-extrapolated (namely based on size-consistent energy differences) interaction energy at the A level of theory obtained with $N_{aux} = n \times N_v$.

The top panel of Fig. 5.3 shows the mean absolute errors (MAE) with respect to $E_{\text{RPA}}^{CBS(41,70)}$ for the RPA interaction energies obtained from the CBS for three different intervals, $E_{\text{RPA}}^{CBS(11,30)}$, $E_{\text{RPA}}^{CBS(21,40)}$, and $E_{\text{RPA}}^{CBS(31,50)}$, and obtained

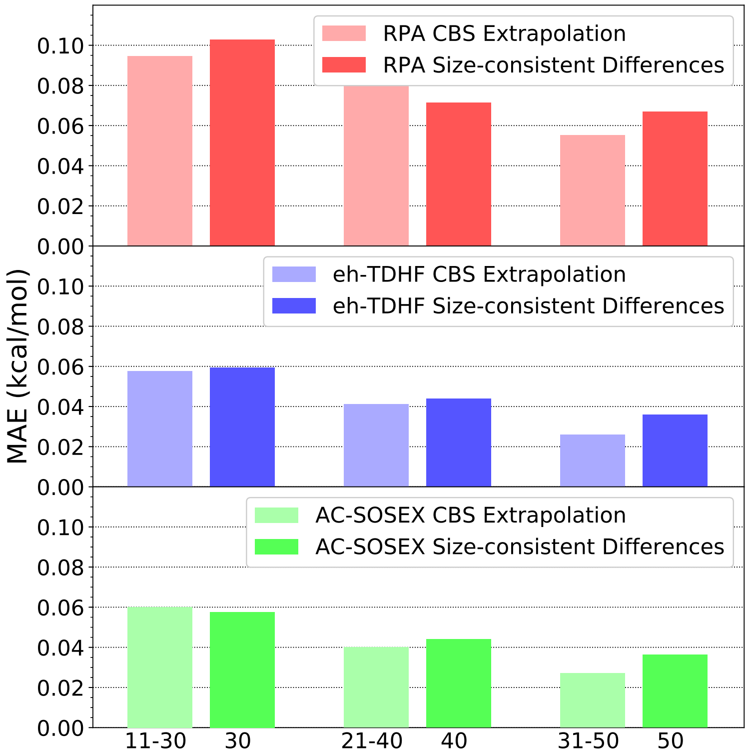


Figure 5.3: Mean absolute errors (MAEs) of RPA (red), eh-TDHF (blue) and AC-SOSEX (green) for the A24 test set interaction energies using CBS extrapolations for $E^{CBS(11,30)}$, $E^{CBS(21,40)}$, and $E^{CBS(31,50)}$, and size-consistent energy differences for $E^{N_{aux}=30}$, $E^{N_{aux}=40}$, and $E^{N_{aux}=50}$. The $E^{CBS(41,70)}$ interaction energy was used as a reference (see text).

from energy differences for three different values of N_{aux}, $E_{\text{RPA}}^{N_{aux}=30}$,$E_{\text{RPA}}^{N_{aux}=40}$, and $E_{\text{RPA}}^{N_{aux}=50}$. In the middle and lower panels of Fig. 5.3, analogous results are shown for the eh-TDHF and AC-SOSEX methods respectively. From the three panels, we observe that the MAEs are rather small in all cases. However, considering the small energy scales involved in the A24 test set (the average interaction energy at the CCSDT(Q) level of theory is about 2 kcal/mol [121]), a high level of convergence is also particularly important. By considering Fig. 5.3 we can notice that: (1) RPA presents larger MAEs than the beyond-RPA methods, hinting that its convergence is slower and/or less stable; this empirically observed behavior might be related to a reduced cusp problem in beyond-RPA approximations, although further investigation will be necessary in the future for a better understanding; (2) if the maximum value for N_{aux} is the same, results based on the CBS extrapolation or on the energy differences have similar accuracy; (3) MAEs systematically decrease by increasing the maximum value of N_{aux}; (4) the value of $N_{aux} = 40 \times N_v$ used previously in Chapter 4 to compute interaction energies provides a good level of accuracy, especially for the beyond-RPA methods.

According to this discussion for the A24 test set, the CBS results and the size-consistent energy differences have similar convergence rates and the two approaches are equivalent for most practical calculations. However, the size-consistent energy differences have the advantage of not being biased by the choice of a particular function to extrapolate to CBS.

5.2.2 Methane-Formaldehyde Complex

To further study the convergence rate of the RPA and beyond-RPA methods with respect to the auxiliary basis set size, we computed six points on the potential energy surface (PES) of the methane-formaldehyde complex (MFC). For these points accurate coupled-cluster benchmark results have been recently published [96]. The numerical results for the MFC complex are also based on FHI norm-conserving pseudopotentials. This system allows us to verify the accuracy of the RPA and beyond-RPA methods and the reliability of the convergence techniques discussed in Sec. 5.1 also for the case of out-of-equilibrium geometries. As in the case of the A24 test set, we use $E^{CBS(41,70)}$ energies as the reference and we test the convergence of the CBS extrapolated values for three intervals, $E^{CBS(11,30)}$, $E^{CBS(21,40)}$, and $E^{CBS(31,50)}$, and the energy differences at three values of N_{aux}, $E_{\text{RPA}}^{N_{aux}=30N_v}$, $E_{\text{RPA}}^{N_{aux}=40N_v}$, and $E_{\text{RPA}}^{N_{aux}=50N_v}$. In Fig. 5.4 we present the corresponding MAEs for the RPA, the eh-TDHF, and the AC-SOSEX. For the MFC, the MAEs decrease by increasing the value of N_{aux}, as was observed with the A24 test set. In contrast though, the convergence of energy differences significantly outperform CBS extrapolations for the beyond-RPA methods, whereas the converse is true for the RPA. To a certain extent this behavior can be interpreted by considering that size-consistent energy differences benefit from error cancellations, whereas the CBS extrapolations are sensitive to small fluctuations in the data. If for example we consider $E_{RPA}^{CBS(41,60)}$, which includes a subset of the points used for the reference $E_{RPA}^{CBS(41,70)}$, the MAE actually increases to 0.07 kcal/mol, showing that small fluctuations influence the CBS extrapolation. On the other hand, the MAE on the energy differences continues to decrease with an increase in N_{aux}, reaching a value of 0.12 kcal/mol for $E_{RPA}^{N_{aux}=60N_v}$. Thus, the convergence of the RPA is slower and more unstable, similar to that observed for the A24 test set. For the beyond-RPA methods, convergence is instead more systematic, with both approaches leading to similar results for sufficiently high values of N_{aux}. The method based on size-consistent energy differences is thus confirmed as a reliable and unbiased approach for the MFC as well, in particular, for the eh-TDHF and AC-SOSEX methods.

Fig. 5.5 shows the CBS results for the MFC PES compared to CCSD(T) reference values [96]. The RPA and beyond-RPA methods all capture the PES correctly, with the correct ordering in energy of the configurations. With respect to the coupled-cluster reference the mean absolute error for the RPA is 0.085 kcal/mol. This value reduces to 0.061 and 0.066 kcal/mol for the eh-TDHF and AC-SOSEX respectively, highlighting the accurate nature of these methods. However, the convergence of RPA results is more problematic and, by considering the small energy scales involved in this PES, the eh-TDHF and AC-SOSEX results are certainly more reliable.

5.2.3 Reaction energies

After considering weakly bound systems, in this section we discuss results for the reaction energy test set by Hesselmann [65], which involves the breaking and formation of covalent bonds. While results for this test set were previously presented in Chapter 4, here we extend the auxiliary basis set up to $70 \times N_v$ and compare extrapolated results with energy differences. Studying this test set is particularly interesting, since going from the reactants (e.g. HCOOH and NH_3) to the products (e.g. $HCONH_2$ and H_2O) involves the breaking and formation of covalent bonds. Accordingly, reactants and products could have a rather different structure of the dielectric eigenspectrum and this could represent a challenge for the size-consistent energy difference method.

In Fig. 5.6 we present the MAEs obtained at the RPA, the eh-TDHF, and the AC-SOSEX levels of theory. As for the weakly bound systems, the MAEs are computed with respect to the $E^{CBS(41,70)}$ extrapolated value. Not surprisingly for this test set the deviation between CBS and size-consistent energy differences is more sizeable than for the weakly-interacting systems. A few observations are

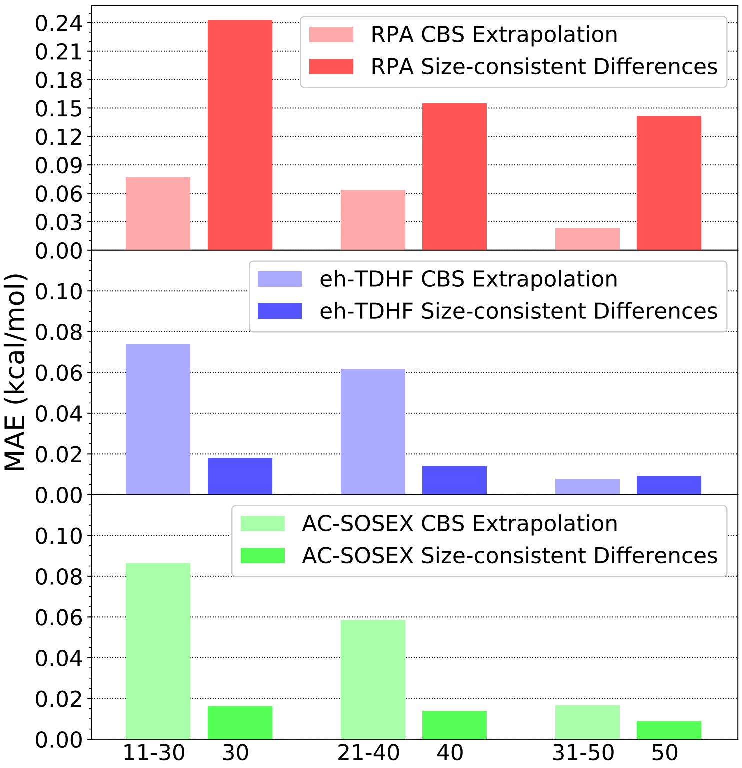


Figure 5.4: Mean absolute errors (MAEs) of RPA (red), eh-TDHF (blue) and AC-SOSEX (green) for the methane-formaldehyde complex interaction energies using CBS extrapolations for $E^{CBS(11,30)}$, $E^{CBS(21,40)}$, and $E^{CBS(31,50)}$, and size-consistent energy differences for $E^{N_{aux}=30}$, $E^{N_{aux}=40}$, and $E^{N_{aux}=50}$. The $E^{CBS(41,70)}$ interaction energy was used as a reference (see text).

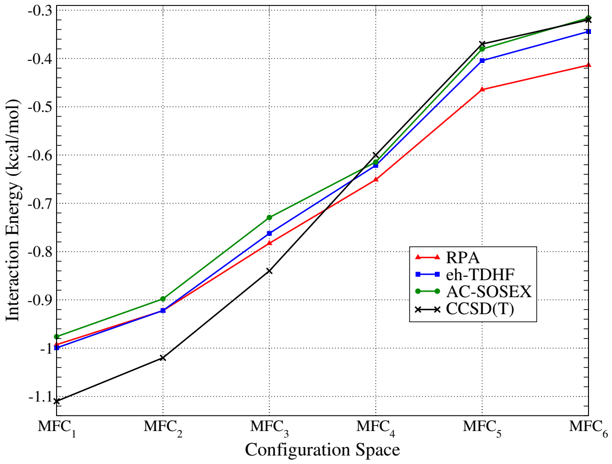


Figure 5.5: Potential energy surface points (kcal/mol) of the methane-formaldehyde complex at the extrapolated RPA (red), eh-TDHF (blue) and AC-SOSEX (green) levels of theory compared with the CCSD(T) values (black).

important for a full understanding of the results in Fig. 5.6: (1) As the average reaction energy at the CCSDT(Q) level of theory is about 29 kcal/mol, the energy scales involved in this test set are usually larger than for weakly interacting systems and, accordingly, higher errors can be tolerated; (2) As mentioned before, the use of an $E^{CBS(41,70)}$ reference is to some extent arbitrary and creates a bias in favor of the CBS results; by using $E^{N_{aux}=70}$ as a reference instead, the behavior shown in Fig. 5.6 would be swapped, with the size-consistent energy difference method showing a faster convergence; (3) The MAE of $E^{N_{aux}=70}$ with respect to $E^{CBS(41,70)}$ (or vice versa) is within 0.09 kcal/mol for all the approximations and this observation point to the fact that the two methods studied in this work converge to the same values.

According to this discussion, the size-consistent energy difference method works well also beyond the expected range of applicability. The convergence of correlation energies for applications involving the breaking and formation of covalent bonds is more problematic but can achieve a good level of reliability.

5.3 Conclusions

We analyzed the basis set convergence of the RPA, eh-TDHF, and AC-SOSEX methods. Specifically, two techniques were used to converge correlation energies: one based on a complete basis set extrapolation, and the other that converges energy differences by accounting for the size-consistency property. These two methods were tested by considering the A24 test set and six points on the poten-

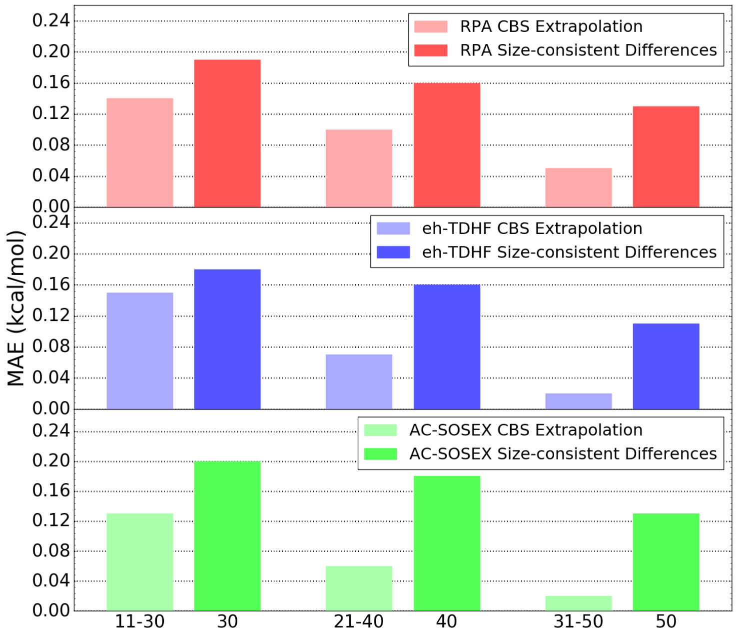


Figure 5.6: Mean absolute errors (MAEs) of RPA (red), eh-TDHF (blue) and AC-SOSEX (green) for the reaction energy test set of Hesselmann [65] using CBS extrapolations for $E^{CBS(11,30)}$, $E^{CBS(21,40)}$, and $E^{CBS(31,50)}$, and size-consistent energy differences for $E^{N_{aux}=30}$, $E^{N_{aux}=40}$, and $E^{N_{aux}=50}$. The $E^{CBS(41,70)}$ interaction energy was used as a reference (see text).

tial energy surface of the methane-formaldehyde complex. It was found that both the CBS extrapolation and size-consistent energy differences converge to similar results at similar speeds as a function of the auxiliary basis set size. It was also shown that the RPA is in general characterized by a slower convergence than the beyond-RPA methods. At least for the weakly bound systems considered in this manuscript, the two approaches have been found to be equivalent for practical purposes. If the breaking and formation of covalent bonds is involved, convergence is less trivial but satisfactory levels of accuracy can be reached.

The size-consistent energy difference technique represents a new and alternative scheme to conventional basis set extrapolation for converging results of correlated methods within a dielectric matrix formulation. While possibly less general than the CBS extrapolation, this approach has the advantage of not relying on any assumption on the model used for the fitting.

Chapter 6

Applications to realistic systems by machine learning thermodynamic perturbation theory

This chapter is based on original work previously published in Refs. [125, 20].

Up to this point we have discussed new approximations for the electronic correlation that hold the promise for a higher level of accuracy than traditional DFT approximations but involve a significantly high computational cost. Additionally, the implementation of analytical gradients, necessary for geometry optimizations or molecular dynamics, is challenging for these methods and further increases the computational requirements. Accordingly, RPA molecular dynamics are by far too expensive to be routinely used in applications that require higher levels of accuracy. In this chapter we will consider examples of computations of adsorption enthalpies in zeolites. Directly performing a RPA *ab initio* molecular dynamics (AIMD) for these systems is estimated to require several tens of millions of CPU hours. In order to compute finite-temperature properties at the RPA level of theory in a simple and numerically efficient way, here we discuss the MLPT method, that couples machine learning (ML) techniques with thermodynamic perturbation theory (TPT). The numerical applications of this methodology presented below show that a high level of accuracy can be achieved by training a ML model with as few as 10 RPA energies. The MLPT approach represents an important step towards the possibility of applying the RPA and other high-accuracy/high-cost *ab initio* methodologies to reach a new level of reliability in finite-temperature materials simulations.

6.1 Methodological approach: Machine learning thermodynamic perturbation theory

Since the work of Behler and Parrinello (BP) in 2007 [10] the use of ML algorithms in MD simulations is growing in popularity [24, 23, 152, 9]. In the BP approach and its variants, a certain number of configurations of a system is generated by performing a relatively short MD simulation or by properly sampling the configuration space. A ML learning model, usually a neural network (NN), is trained on the energies and/or forces computed for these configurations. A full molecular dynamics (MD) simulation is then performed by using the previously trained neural network. At each step of the MD simulation, instead of directly solving the Schrödinger equation, forces and energies are inexpensively "predicted" by the ML model and used to generate the configuration in the next step. In this way the timescale accessible by a typical MD run is significantly increased while keeping to a large extent the accuracy of an *ab initio* simulation. However, the training of a ML learning algorithm for a MD simulation typically requires thousands or even tens of thousands of configurations and, accordingly, the BP approach is almost exclusively trained on DFT energies and/or forces.

Similarly to the BP approach, here we use a ML learning regression model to predict the energy surface from a certain number of training examples. Differently from the BP approach, this model is not used to perform a full MD simulation but it is rather coupled with thermodynamic perturbation theory, whose brute force application also requires an impractically high number of RPA calculations. The corresponding MLPT calculation is executed in the following steps: (1) an AIMD simulation is performed using a numerically affordable DFT functional; (2) a small number N_{train} of configurations generated by the initial MD is selected and RPA calculations are performed on them; (3) a ML model is trained on the selected configurations to predict the RPA energies for all the remaining configurations; (4) based on TPT, the contributions from the configurations generated at step 1 are reweighted using the predicted energies and RPA ensemble averages are computed. For the specific objective of this work, the present MLPT approach has the following advantages with respect to BP MD: (1) The ML regression algorithm is exclusively trained on energies; this is useful since most condensed phase implementations of correlated methods do not support the calculation of analytical gradients. (2) The implementation is easier, since the BP method requires also the derivative of the ML model with respect to atomic positions in order to compute forces. (3) The accuracy of the ML predictions can be enhanced in a straightforward way by using a Δ-machine learning approach [115], which trains the model on energy differences instead of total energies (see Sec. 6.1). In the two following subsections we discuss the technical details of thermodynamic perturbation theory and machine learning that are at the base of the MLPT method.

6.1.1 Thermodynamic perturbation theory

Starting from a molecular dynamics driven by a numerically inexpensive approximation, our goal is to use thermodynamic perturbation theory (TPT) to obtain finite-temperature observables at a high level of theory. This methodology is general but for the specific numerical applications of this manuscript the van der Waals corrected PBE+D2 functional [53, 106] will be used as a (cheap) starting point to compute RPA ensemble averages.

A general review of TPT in the context of classical MD can be found e.g. in Refs. 22, 114. Here we will discuss the main ideas necessary to understand the MLPT method. Molecular dynamics is used to evaluate the finite-temperature expectation values of a certain observable O:

$$\langle O \rangle_H = \frac{\int O(\mathbf{q},\mathbf{p})\mathrm{exp}\left\{-\beta H(\mathbf{q},\mathbf{p})\right\} d\mathbf{q}\, d\mathbf{p}}{\int \mathrm{exp}\left\{-\beta H(\mathbf{q},\mathbf{p})\right\} d\mathbf{q}\, d\mathbf{p}}, \tag{6.1}$$

where $\langle \ldots \rangle_H$ is used to compactly denote the average over the canonical ensemble corresponding to the Hamiltonian H, $\beta = 1/k_B T$, and $\mathbf{q}$ and $\mathbf{p}$ denote nuclear positions and momenta, respectively. In Eq. 6.1 the nuclear motion is driven by the classical Hamiltonian

$$H(\mathbf{q},\mathbf{p}) = T(\mathbf{p}) + V(\mathbf{q}), \tag{6.2}$$

where $T(\mathbf{p})$ is the nuclear kinetic energy and $V(\mathbf{q})$ is the potential acting on the nuclei. Here we will focus on AIMD, where the potential $V(\mathbf{q})$ is determined by the solution of the electronic Schrödinger equation within different approximations, usually based on DFT:

$$V(\mathbf{q}) = \langle \Psi(\mathbf{q}) | \hat{H}_{DFT} | \Psi(\mathbf{q}) \rangle + U_{II}(\mathbf{q}), \tag{6.3}$$

where U_{II} is the repulsive nuclear energy and $\hat{H}_{DFT}$ is the electronic DFT Hamiltonian (not to be confused with the classical Hamiltonian H). As explained above, our main goal is to use thermodynamic perturbation theory to compute observables within a sophisticated quantum chemical approximation (e.g. the RPA) starting from a molecular dynamics driven by a computationally "cheap" DFT approximation (e.g. PBE+D2). Below the Hamiltonian H and the corresponding potential V will be associated with the less expensive quantum-mechanical approximation while primed quantities (H' and V') will be associated with the more expensive approach. By using the identity

$$H'(\mathbf{q},\mathbf{p}) = H(\mathbf{q},\mathbf{p}) + V'(\mathbf{q}) - V(\mathbf{q}) = H(\mathbf{q},\mathbf{p}) + \Delta V(\mathbf{q}), \tag{6.4}$$

the average of the observable $O(\mathbf{q},\mathbf{p})$ over the ensemble corresponding to H' can be written as

$$\begin{aligned}\langle O\rangle_{H'} &= \frac{\int O(\mathbf{q},\mathbf{p})\text{exp}\left\{-\beta H'(\mathbf{q},\mathbf{p})\right\}d\mathbf{q}\,d\mathbf{p}}{\int \text{exp}\left\{-\beta H'(\mathbf{q},\mathbf{p})\right\}d\mathbf{q}\,d\mathbf{p}} \\ &= \frac{\int O(\mathbf{q},\mathbf{p})\text{exp}\left\{-\beta \Delta V(\mathbf{q})\right\}\text{exp}\left\{-\beta H(\mathbf{q},\mathbf{p})\right\}d\mathbf{q}\,d\mathbf{p}}{\int \text{exp}\left\{-\beta \Delta V(\mathbf{q})\right\}\text{exp}\left\{-\beta H(\mathbf{q},\mathbf{p})\right\}d\mathbf{q}\,d\mathbf{p}} \\ &= \frac{\langle O\text{exp}\left\{-\beta\Delta V\right\}\rangle_H}{\langle \text{exp}\left\{-\beta\Delta V\right\}\rangle_H}, \end{aligned} \tag{6.5}$$

which provides the main result of TPT used in this discussion. This equation has important theoretical and practical implications, showing that the ensemble average $\langle O\rangle_{H'}$ can be redefined in terms of an ensemble average corresponding to H; this is achieved through the reweighting factor exp $\{-\beta\Delta V\}$. From a numerical point of view Eq. 6.5 implies that the statistics sampled by a MD driven by H can be used to obtain finite-temperature expectation values corresponding to H'. This is a crucial point since directly performing a MD based on H' would be extremely costly if feasible at all.

From a practical point of view, in order for the perturbative procedure in Eq. 6.5 to be effective, it is necessary that a MD driven by H would sample a configurational space that is similar to that sampled by an hypothetical H' MD. If the corresponding high probability ensembles are disjoint or only partially superimposed, the perturbative approach will have low accuracy. A detailed discussion of this issue can be found in Refs. 22, 114. Following Ref. 125, we suppose that PBE+D2 represents a reliable starting point for RPA.

In this manuscript the TPT formalism will be applied to the specific problem of computing adsorption energies in zeolites at the RPA level of theory. The adsorption enthalpy can be computed as

$$\Delta_{\text{ads}}H(T) = \langle V'_{S+A}\rangle_{H'} - [\langle V'_S\rangle_{H'} + \langle V'_A\rangle_{H'}] - k_B T, \tag{6.6}$$

where S denotes the clean substrate (a zeolite in this case) and A the adsorbate; the values of $\langle V'\rangle_{H'}$ can be obtained from Eq. 6.5 for the specific case $O = V'$. In order to evaluate numerically the ensemble averages in practical numerical calculations, a MD based on H is first performed and then the ensemble average is computed as

$$\langle V'\rangle_{H'} \approx \frac{\sum_{i=1}^{N_{tot}} V'_i\text{exp}\left\{-\beta(V'_i - V_i)\right\}}{\sum_{i=1}^{N_{tot}}\text{exp}\left\{-\beta(V'_i - V_i)\right\}} = \frac{\sum_{i=1}^{N_{tot}} V'_i\text{exp}\left\{-\beta\Delta V_i\right\}}{\sum_{i=1}^{N_{tot}}\text{exp}\left\{-\beta\Delta V_i\right\}}, \tag{6.7}$$

where the index i runs over all the N_{tot} configurations generated by the production MD run. Since in the MD simulations considered below N_{tot} = 190000, Eq. 6.7 highlights the limited practical applicability of TPT alone. Indeed, even by considering a smaller subset of uncorrelated configurations, several thousands of calculations of V' (namely of RPA total energies) are expected to be necessary to reach convergence. This issue can be significantly alleviated by using a ML regression algorithm. Within this approach only few V_i''s are explicitly computed and used to train a ML model; all the remaining values up to N_{tot} are then predicted at a negligible computational cost.

6.1.2 Machine Learning approach

In order to significantly decrease the number of expensive calculations (i.e. based on the RPA) required to apply thermodynamic perturbation theory, we propose the use of a ML regression model. To this purpose we chose the smooth overlap of atomic positions (SOAP) [5, 26] as descriptor for the geometric configurations and the kernel ridge regression (KRR) [130] as ML algorithm to predict total energies. This choice, that allowed us to reach a satisfactory level of accuracy, was partially driven by a recent comparative study that reported a good performance of SOAP in predicting adsorption energies of hydrogen on nanoclusters [72]. Several other descriptors for periodic materials have also been proposed in the literature, such as, for example, the many-body tensor representation (MBTR) [71] or atom-centered symmetry functions (ACSF) [8]. While a systematic comparison of different ML approaches is beyond the purpose of this discussion, other descriptors and/or ML algorithms might also perform well for the problem we are considering.

Below we summarize the main ideas of the SOAP and KRR approaches; a more detailed presentation can be found in Refs. 5, 130, 26. The numerical implementation of the ML models considered here is based on the DScribe libraries [69][1].

For a structure A of a given material let us define an environment $\mathcal{X}_i^A$ that includes the atoms surrounding a specific atom i within a certain radius. The density of the atoms ρ in the environment $\mathcal{X}_i^A$ is defined as a sum of Gaussians

$$\rho_{\mathcal{X}_i^A}(\mathbf{r}) = \sum_{j \in \mathcal{X}_i^A} \exp\left\{-\frac{(\mathbf{r}-\mathbf{x}_j)^2}{2\sigma^2}\right\}, \tag{6.8}$$

where the Gaussians of variance σ^2 are centered at the positions $\mathbf{x}_j$ of all the atoms in the environment $\mathcal{X}_i^A$ (including the central one). For systems containing different atomic species, as those considered here, a different environment has to be defined for each species. The SOAP kernel is then defined as an overlap integral of the densities of different environments, possibly from different structures A and

---

[1] The DScribe libraries are available at https://github.com/SINGROUP/dscribe

B:

$$\tilde{k}(\mathcal{X}_i^A, \mathcal{X}_j^B) = \int d\hat{R} \left(\int \rho_{\mathcal{X}_i^A}(\mathbf{r}) \rho_{\mathcal{X}_j^B}(\hat{R}\mathbf{r}) d\mathbf{r} \right)^2 , \tag{6.9}$$

where the first integral is performed over the three-dimensional rotations, as required in order to have a rotationally invariant representation. For practical purposes a normalised version of the kernel in Eq. 6.9 is typically used:

$$k(\mathcal{X}_i^A, \mathcal{X}_j^B) = \frac{\tilde{k}(\mathcal{X}_i^A, \mathcal{X}_j^B)}{\sqrt{\tilde{k}(\mathcal{X}_i^A, \mathcal{X}_i^A) \tilde{k}(\mathcal{X}_j^B, \mathcal{X}_j^B)}}. \tag{6.10}$$

This normalized kernel provides a measure of similarity since the maximum value (one) is attained when the two environments $\mathcal{X}_i^A$ and $\mathcal{X}_j^B$ are identical (independently of rotation and translations of the environment). Because of the specific form of the density of atoms, expressed as an expansion of Gaussians, the integral over the rotations in Eq. 6.9 can be carried out analytically. This is achieved by using a basis of spherical harmonics and the corresponding formalism is discussed in details in Refs. 5, 26.

The kernels introduced in Eqs. 6.9-6.10 can be used to compare environments surrounding specific atoms and, in this respect, describe only a local similarity. For our specific purpose it is necessary to determine the similarity among different structures (all with the same chemical composition) generated by a MD simulation. This global information is included in the full covariance matrix $\mathbf{C}(A, B)$, whose elements

$$C_{ij}(A, B) = k(\mathcal{X}_i^A, \mathcal{X}_j^B) \tag{6.11}$$

contain all the possible pairings of environments belonging to structure A and to structure B. Based on the matrix in Eq. 6.11 different global kernels can be defined. For our purpose the regularized entropy match (REMatch) kernel will be used, that is typically coupled with SOAP descriptors [5, 26]. The REMatch kernel is defined as

$$K^\gamma(A, B) = \mathrm{Tr}\mathbf{P}^\gamma \mathbf{C}(A, B), \tag{6.12}$$

where $\mathbf{P}$ is a doubly stochastic matrix that satisfies the condition

$$\mathbf{P}^\gamma = \arg\min_{\mathbf{P}} \sum_{i,j} P_{ij}(1 - C_{ij} + \gamma \ln P_{ij}), \tag{6.13}$$

and γ is a hyperparameter that controls the entropic contribution.

The REMatch kernel in Eq. 6.12 can then be coupled with the kernel ridge regression ML algorithm. Since our purpose is to predict RPA total energies for the configurations generated by a PBE+D2 MD, a certain number of configurations N$_{train}$ is selected to train the ML model (the specific selection criterion used will be discussed later). From these N$_{train}$ configurations the REMatch kernel $\mathbf{K}^{\gamma}_{train}$ is built and used to train the ML model:

$$\mathbf{w} = (\mathbf{K}^{\gamma}_{train} + \lambda \mathbf{1})^{-1} \mathbf{y}_{train} \tag{6.14}$$

where $\mathbf{y}_{train}$ in this case contains the RPA energies corresponding to the configurations in the training set and λ is a regularization hyperparameter used to prevent overfitting. The weights $\mathbf{w}$ determined by Eq. 6.14 are then used to predict the RPA energies for all the remaining N$_{predict} = N_{tot} - N_{train}$ configurations in the MD:

$$\mathbf{y}_{predict} = \mathbf{K}^{\gamma}_{predict} \mathbf{w}, \tag{6.15}$$

where $\mathbf{y}_{predict}$ contains the predicted energies and $\mathbf{K}^{\gamma}_{predict}$ is the N$_{predict}$×N$_{train}$ REMatch kernel.

As RPA calculations are highly expensive it is crucial to develop a ML model that can be trained with a number of configurations which is as small as possible. Indeed, N$_{train}$ explicit RPA calculations are required while the remaining N$_{predict}$ are inexpensively predicted by Eq. 6.15. To this purpose it is useful to use the energy difference $\Delta V_i = V_i' - V_i$ as a variable for training rather than the absolute energy V_i'. Indeed, the energy difference ΔV_i has a much smoother behavior than V_i' and the error on the prediction can be lowered by one order of magnitude with respect to a ML model that uses total energies. This idea is based on the Δ-machine learning approach [115]. For example, for CH_4 adsorbed in protonated chabazite (see Sec. 6.2), ΔV_i (and, correspondingly, V_i') can be predicted with a root mean square error of about 0.4 kcal/mol by using only N$_{train} = 10$ and with a systematic decrease of the error as a function of N$_{train}$.

6.1.3 Statistical error evaluation

The accuracy of a ML model is characterized by a statistical error that can be possibly decreased by increasing the quantity of data used for training. In our specific case the energies V_i' or, equivalently, the energy differences ΔV_i are predicted with a statistical error that has implications on the estimate of the internal energy (Eq. 6.7) and enthalpy (Eq. 6.6).

As a starting point for our discussion it is necessary to determine the error involved in the prediction of ΔV_i and V_i'. To this purpose the root mean square error (RMSE) of the predicted energies with respect to the exact RPA results is computed for a small test set. This set includes configurations that are reasonably independent and uncorrelated with respect to configurations used for training the ML model (see Sec. 6.2 for additional details). Since the prediction is not biased (mean error is always well below 0.1 kcal/mol), the RMSE can be interpreted as the standard deviation of the distribution of the residuals in the regression model and will be denoted by s. The RMSE in the prediction of the RPA energies of individual configurations leads to both a systematic and random error in the internal energy expressed by Eq. 6.7.

To discuss the systematic error we will split the ensemble energy in Eq. 6.7 into two contributions

$$\langle V' \rangle_{H'} = \frac{\sum_{i=1}^{N_{tot}} V_i \exp\{-\beta \Delta V_i\}}{\sum_{i=1}^{N_{tot}} \exp\{-\beta \Delta V_i\}} + \frac{\sum_{i=1}^{N_{tot}} \Delta V_i \exp\{-\beta \Delta V_i\}}{\sum_{i=1}^{N_{tot}} \exp\{-\beta \Delta V_i\}} = f_1 + f_2 \quad (6.16)$$

and we will focus on the last term f_2 that depends only on ΔV_i. Because of the statistical error associated with the machine learning model, we suppose that the exact ΔV_i is replaced by values E normally distributed with variance s^2:

$$P(E) = K \exp\left\{-\frac{(E - \Delta V_i)^2}{2s^2}\right\}, \quad (6.17)$$

where K is a normalizing factor. The estimate of f_2 that keeps into account this distribution can be written in the following way:

$$\begin{aligned}
\tilde{f}_2 &= \frac{\sum_i \int dE\, E \exp\{-\beta E\} \times \exp\left\{-\frac{(E-\Delta V_i)^2}{2s^2}\right\}}{\sum_i \int dE \exp\{-\beta E\} \times \exp\left\{-\frac{(E-\Delta V_i)^2}{2s^2}\right\}} \\
&= \frac{\sum_i \int dE\, E \exp\{-\beta \Delta V_i\} \times \exp\left\{-\frac{(E-\Delta V_i-\beta s^2)^2}{2s^2}\right\}}{\sum_i \int dE \exp\{-\beta \Delta V_i\} \times \exp\left\{-\frac{(E-\Delta V_i-\beta s^2)^2}{2s^2}\right\}} \\
&= \frac{\sum_i \int dE' (E' + \beta s^2) \exp\{-\beta \Delta V_i\} \times \exp\left\{-\frac{(E'-\Delta V_i)^2}{2s^2}\right\}}{\sum_i \int dE' \exp\{-\beta \Delta V_i\} \times \exp\left\{-\frac{(E'-\Delta V_i)^2}{2s^2}\right\}} \\
&= f_2 + \beta s^2, \qquad (6.18)
\end{aligned}$$

where the sum over i is implicitly intended between 1 and N$_{tot}$. Here f_2 is the reweighted average obtained using the exact energy differences ΔV_i while $\tilde{f}_2$ is based on the values of ΔV_i predicted by a ML algorithm that introduces a random statistical error. Accordingly, when ML and TPT are coupled, a systematic error of βs^2 in the prediction of ensemble averages has to be expected. In the assumption that the residuals are normally distributed and that the variance s^2 is known, this bias can be easily evaluated using Eq. 6.18. The results that will be discussed below have all been corrected for this systematic error. The value of s and, consequently, of βs^2 decreases by increasing the number of training examples N$_{train}$. For example, if we consider CH_4 adsorbed in protonated chabazite (see Sec. 6.2 and in particular Table 6.1) the βs^2 correction is 0.33 kcal/mol for N$_{train}$=10 and

0.07 kcal/mol for N$_{train}$=200. By the same logic, it is easy to see that $\tilde{f}_1 = f_1$ and hence the first term on the right hand side of Eq. 6.16 is unaffected by the systematic error arising from s.

We now consider the evaluation of the random error in the ensemble averages associated with the ML model. To this purpose it is important to notice that Eq. 6.7 is a non linear function of V_i' and, accordingly, an error propagation formula based on first derivatives will be used here. We have that the standard error δ_{ML} in the ensemble average is

$$\delta_{ML} = \sqrt{\sum_{i=1}^{N_{tot}} \left(\frac{\partial \langle V' \rangle_{H'}}{\partial V_i'} \right)^2 s_{V_i'}^2} = s \sqrt{\sum_{i=1}^{N_{tot}} \left(\frac{\partial \langle V' \rangle_{H'}}{\partial V_i'} \right)^2} \tag{6.19}$$

where $s_{V_i'} = s$ is the standard deviation of the V_i''s distribution and the derivative is explicitly given by

$$\frac{\partial \langle V' \rangle_{H'}}{\partial V_i'} = w_i \left(1 - \beta V_i' + \beta \langle V' \rangle_{H'}\right), \tag{6.20}$$

where we defined the weight

$$w_i = \frac{\exp\left\{-\beta(V_i' - V_i)\right\}}{\sum_{i=1}^{N_{tot}} \exp\left\{-\beta(V_i' - V_i)\right\}}. \tag{6.21}$$

In all calculations discussed further, δ_{ML} is typically much smaller than s.

Finally, we discuss the source of random error that is associated with the finite sample size. This error is not related to the ML approach but is typical of MD simulations, which span a finite time interval. As the term $\langle V' \rangle_{H'}$ in Eq. 6.7 can be considered as a weighted average of the form

$$\langle V' \rangle_{H'} = \sum_{i=1}^{N_{tot}} V_i' w_i \tag{6.22}$$

with the weight w_i defined in Eq. 6.21, the corresponding statistical error can be estimated using the standard formula [44]

$$\delta_s = \sqrt{\frac{N_s}{(N_s - 1)} \sum_{i=1}^{N_s} w_i^2 \left(V_i' - \langle V' \rangle_{H'}\right)^2}, \tag{6.23}$$

where $N_s \ll N_{tot}$ is the number of uncorrelated samples. In practice, the latter can be identified e.g. by the block method of Flyvbjerg and Petersen. [37, 38]. The value of δ_s depends on the length and the quality of the MD trajectory and it can, in principle, be made arbitrarily small by performing a sufficiently long simulation.

In our numerical examples discussed below, the error bars are evaluated as

$$\delta = \sqrt{\delta_{ML}^2 + \delta_s^2} \tag{6.24}$$

where δ_{ML} is defined in Eq. 6.19 and δ_s in Eq. 6.23. The values of δ_s are in all cases discussed in Sec. 6.2 significantly larger than those of δ_{ML} and hence the former term represents a dominant contribution to δ.

6.2 RPA enthalpies of adsorption in zeolites

In order to establish the accuracy and efficiency of the MLPT methodology we consider the calculation of molecular enthalpies of adsorption in zeolites, which are porous materials used in many applications [138] such as depollution [21, 17, 67], separation of chemicals [110, 88, 76], and catalysis [52, 18, 83]. Specifically, we consider four applications involving CH_4 and CO_2 in protonated chabazite (denoted as HCHAB) and siliceous chabazite (CHAB). The four corresponding adsorbed systems will be indicated as HCHAB+CH_4, HCHAB+CO_2, CHAB+CH_4, and CHAB+CO_2. The HCHAB+CH_4 example will be discussed in more detail to describe the numerical properties of the MLPT approach. The corresponding periodic model is shown in Fig. 6.1.

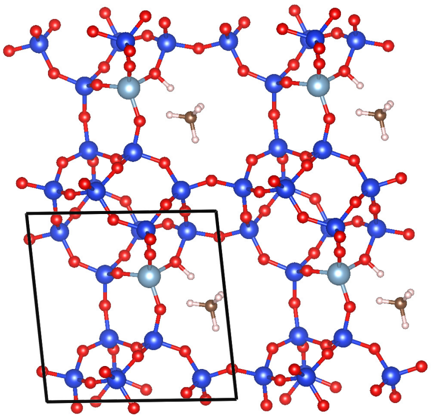


Figure 6.1: Periodic model used for CH_4 adsorbed in protonated chabazite. The unit cell used in our calculations is shown by solid lines.

Following Ref. 125 we used MD trajectories based on the vdW corrected PBE+D2 functional [53, 106] as a starting point for the MLPT method. MD sim-

Table 6.1: RMSE (kcal/mol) in the prediction of RPA energies of the configurations of protonated chabazite (HCHAB), CH_4, and adsorbed system (HCHAB+CH_4) as a function of the number of training structures N_{train}. The RMSE is estimated on a test set of 20 configurations (see text).

N_{train}	HCHAB+CH_4	HCHAB	CH_4
10	0.45	0.34	0.03
25	0.36	0.30	0.01
50	0.29	0.26	0.01
100	0.24	0.23	0.01
200	0.21	0.20	0.01

ulations were performed in the NVT ensemble. Temperature was set to 300 K and the length of trajectories was fixed at 100 ps with an integration step of 0.5 fs. In a production period succeeding the equilibration length of 5 ps, a total of 190,000 configurations were generated.

The first step of the MLPT is the selection of a proper training set with N_{train} elements. Since configurations along a MD trajectory are time correlated, we select configurations that are as distanced as possible in time. For example, N_{train}=10 indicates that out of the 190,000 configurations one configuration every 19,000 is selected to obtain a total of 10 training structures. Five training sets are considered here with $N_{train} = 10, 25, 50, 100,$ and 200. Additionally, a test set of 20 configurations is selected to estimate the RMSE in the prediction of the machine learning model. Those configurations are chosen randomly but with the constraint of being at a distance of at least 350 time steps from the geometries contained in the largest training set (N_{train}=200). In this way the test and training sets are reasonably independent. For the HCHAB+CH_4 system the RMSE error (also denoted with s) in the prediction of the RPA energies of configurations is shown in Table 6.1. The RMSE clearly decreases by increasing the number of training examples N_{train}.

Since configurations are selected using time separation as a criterion it is also important to understand to which extent they span the geometric space. This is particularly important for the adsorbed systems, where the molecule can move rather freely in the cage of the zeolite. If certain parts of the configurational space do not have representative structures included in the training set, the ML learning model might have limited predictive power in these regions. Fig. 6.2 helps visualizing the selected configurations for HCHAB+CH_4 using the t-Distributed Stochastic Neighbor Embedding algorithm (t-SNE) [87]. t-SNE is an unsupervised learning method that allows for a dimensionality reduction to visualize high-dimensional data: Points that are closer in the high-dimensional space have higher

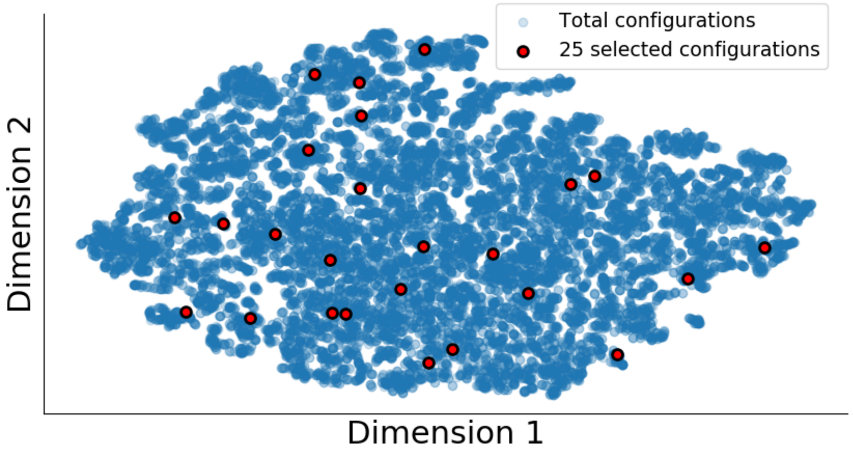
Table 6.2: RPA internal energies and enthalpy of adsorption (kcal/mol) of protonated chabazite (HCHAB), CH_4, and adsorbed system (HCHAB+CH_4) as estimated by MLPT for different values of N_{train}.

N_{train}	HCHAB+CH_4	HCHAB	CH_4	Δ_{ads} H
10	-21398.01 ± 0.19	-20213.82 ± 0.13	-1181.05 ± 0.02	-3.74 ± 0.23
25	-21398.13 ± 0.21	-20213.94 ± 0.13	-1181.07 ± 0.02	-3.72 ± 0.25
50	-21398.17 ± 0.21	-20213.99 ± 0.13	-1181.06 ± 0.02	-3.72 ± 0.25
100	-21398.13 ± 0.21	-20213.93 ± 0.12	-1181.06 ± 0.02	-3.74 ± 0.24
200	-21398.15 ± 0.20	-20213.95 ± 0.12	-1181.06 ± 0.02	-3.74 ± 0.23

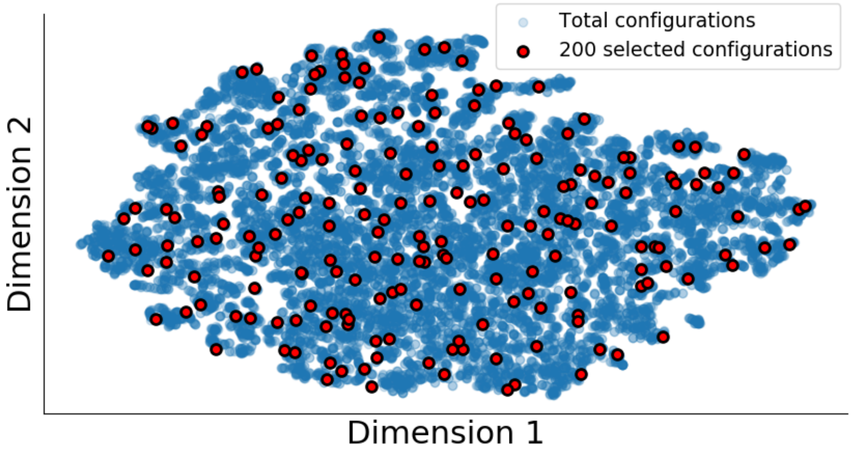
probability to be closer in the reduced two-dimensional space used for the plot. In order to generate Fig. 6.2, t-SNE is used to visualize the SOAP descriptors generated by considering only the carbon environment of CH_4 in HCHAB+CH_4. In this way the analysis is focused on the diffusion of the molecule in the zeolite rather than on the full system, that contains many more environments and degrees of freedom. Fig. 6.2(b) shows the configurations in the $N_{train} = 200$ training set with respect to the total number of configurations. The sampling is uniform and all the regions of the configuration space are fairly well represented by the training set. For $N_{train} = 25$ (Fig. 6.2(a)) the sampling is less satisfactory but, as shown below, the model can still be predictive. This is likely to be related to the use of the Δ-ML approach. Indeed, the model is trained on energy differences ΔV and a large contribution to the RPA total energy is already provided by the PBE+D2 total energy, whose exact value is available from the initial MD trajectory.

Detailed MLPT numerical results for CH_4 adsorbed in protonated chabazite are presented in Table 6.2, where values for the RPA ensemble averages of each system (CH_4, protonated chabazite, and adsorbed system) and enthalpies of adsorption (in kcal/mol) are reported. Results are rather stable with respect to the N_{train} variation and all the values of internal energies are well within a 0.2 kcal/mol interval. As few as 10 configurations can already provide highly accurate estimates of the RPA enthalphy of adsorption. As explained in Sec. 6.1, these values have been corrected for the systematic error βs^2 (see Eq. 6.18). This correction ranges from 0.33 kcal/mol (N_{train}=10) to 0.07 kcal/mol ($N_{train} = 200$). The table reports also error bars $\pm\delta$ associated with random errors Eq. 6.24. The values of δ are rather insensitive to the change of N_{train}; this is mostly due to the fact that the finite sampling component δ_s, which is independent of N_{train}, provides a larger contribution than δ_{ML}.

In order to highlight the importance of the ML approach in providing converged values for RPA internal energies and enthalpies, in Table 6.3 we present the results that would be obtained by applying thermodynamic perturbation the-



(a)



(b)

Figure 6.2: For HCHAB+CH_4 the dimensionality reduction t-SNE algorithm is used to visualize the configurations selected to train the machine learning model (in red) with respect to the total set of configurations (in blue). (a) $N_{train} = 25$; (b) $N_{train} = 200$.

Table 6.3: RPA internal energies and enthalpy of adsorption (kcal/mol) of protonated chabazite (HCHAB), CH_4, and adsorbed system (HCHAB+CH_4) as estimated by a brute force application of thermodynamic perturbation theory for a small number of configurations N_{conf}. Differently from Table 6.2, the ML approach is not applied in this case, causing thus a dramatic deterioration of convergence.

N_{conf}	HCHAB+CH_4	HCHAB	CH_4	Δ_{ads} H
10	-21401.99	-20213.75	-1180.97	-7.88
25	-21398.81	-20214.22	-1181.27	-3.91
50	-21399.19	-20214.20	-1181.24	-4.34
100	-21398.06	-20214.56	-1181.32	-2.78
200	-21398.41	-20214.56	-1181.16	-3.28

ory only on the small set of training configurations. Namely, in this case no ML model is used and, for example, $N_{conf} = 100$ means that TPT includes only the 100 configurations computed explicitly. Since configurations along a MD trajectory are time correlated, it is reasonable to expect convergence with a relatively small subset of configurations, as assumed in Ref. 33. However, as shown in Table 6.3 this is not the case for the sets considered in this work. For example, even the largest set $N_{conf} = 200$ still predicts results for the enthalpy of adsorption that deviate by about 0.5 kcal/mol with respect to the most converged MLPT values in Table 6.2. Importantly, the enthalpies $\Delta_{\text{ads}}H$ in Table 6.3 strongly oscillate and no convergence can be inferred from these results.

Finally, in Table 6.4 we summarize the RPA enthalpies of adsorption (in kcal/mol) as functions of N_{train} for the all four systems considered in this work. The RPA, which is often considered a reference method for periodic condensed matter applications, provides systematically accurate estimates within chemical accuracy in comparison to experiments. The tendency to slightly underestimate is a well known behavior of the RPA [120]. Enthalpies obtained from the PBE+D2 MD are instead less consistent, with large deviations for the adsorption in HCHAB and more reliable results for CHAB. Importantly, RPA results with $N_{train} = 10$ can be obtained with a CPU time comparable to or even shorter than the underlying PBE+D2 MD but can provide a significantly higher level of accuracy and more reliable results for different types of systems.

Table 6.4: PBE+D2, RPA, and experimental enthalpies of adsorption (kcal/mol) of CH_4 and CO_2 adsorbed in protonated (HCHAB) and siliceous (CHAB) chabazite. RPA results are provided as a function of N_{train}.

Δ_{ads} H	HCHAB+CH_4	HCHAB+CO_2	CHAB+CH_4	CHAB+CO_2
PBE+D2	-6.09	-9.72	-4.64	-5.14
RPA ($N_{train} = 10$)	-3.74 ± 0.23	-8.06 ± 0.44	-3.22 ± 0.19	-4.57 ± 0.25
RPA ($N_{train} = 25$)	-3.72 ± 0.25	-7.85 ± 0.37	-3.19 ± 0.19	-4.48 ± 0.22
RPA ($N_{train} = 50$)	-3.72 ± 0.25	-7.93 ± 0.38	-3.32 ± 0.19	-4.66 ± 0.21
RPA ($N_{train} = 100$)	-3.74 ± 0.24	-8.08 ± 0.34	-3.41 ± 0.19	-4.73 ± 0.21
RPA ($N_{train} = 200$)	-3.74 ± 0.23	-8.01 ± 0.32	-3.36 ± 0.19	-4.65 ± 0.21
Exp.	-4.06[a]	-8.41[b]	-4.09[c]	-5.02[c]

a Refs. 111

b Ref. 110

c Ref. 88

6.3 Conclusions

in this chapter we discussed the MLPT method that couples machine learning techniques with TPT to efficiently evaluate finite temperature properties using the RPA or other correlated quantum-chemical methods. By computing enthalpies of adsorption of small molecules in zeolites, it is shown that as few as 10 RPA energies used to train a ML model are sufficient to reach a satisfactory level of convergence. This work open the path to a broader use of high-accuracy/high-cost *ab initio* methodologies to reach a new level of reliability in materials simulations.

Chapter 7

Perspectives

In Chapter 6 we introduced a general purpose methodology, MLPT, to couple correlated methods with molecular dynamics at a reasonable computational cost. This approach allowed us to compute adsorption enthalpies of molecules in zeolites at the RPA of theory. Considering the high numerical cost of this approximation it would be inconceivable to directly perform a brute-force ab initio molecular dynamics. This recent work on MLPT will be at the base of several future developments and extensions. The ultimate goal will be to develop a highly accurate methodology to describe the properties of materials in realistic (finite-temperature) conditions. Specifically, the following research lines will be pursued:

- **Further development of machine learning perturbation theory (MLPT) -** The MLPT approach (see Chapter 6) is based on thermodynamic perturbation theory coupled to machine learning (ML) techniques. Within this approach a MD is performed using a numerically inexpensive functional and a small number of structures is chosen to sample the configurational space. The high-level correlated calculations are performed only on these geometries. Part of these high-level results are used to train a ML learning model and part of them are held out to test it (alternatively, a cross-validation approach could be used). The regression model is then used to predict the properties (e.g. the energy) of the full set of (hundreds of thousands) configurations and perturbation theory is applied to compute the finite temperature observables. This procedure can be improved in several aspects: (1) In order to build the training set algorithms could be developed to sample as uniformely as possible the configurational space (2) The performance of different types of descriptors and ML algorithms could be tested to make MLPT more efficient and/or accurate (e.g. the many-body tensor representation (MBTR) [71] or atom-centered symmetry functions (ACSF) [8].). (3)

The dependence of the perturbative approach on the starting point should be better understood and approaches to evaluate the corresponding error should be developed. (4) Using beyond-RPA methods or other more traditional quantum chemistry approaches to further improve the accuracy of thermodynamic calculations. (5) Properties other than then the energy could be computed.

The MLPT method is still at an early stage of development but has huge potential to significantly impact the field of materials simulations. To fully establish this approach it will be necessary to understand its range of applicability and numerical properties.

- **Machine learning molecular dynamics (MLMD) -** This is a methodology that has seen a growing popularity since the seminal work of Behler and Parrinello in 2007 [10]. In this approach a large number of configurations of a certain system is generated by performing a MD simulation or by properly sampling the configurational space. A ML learning model, usually a neural network (NN), is then trained on the energies and/or forces computed for these configurations. A full molecular dynamics is then performed by using this neural network. At each step of the MD simulation, instead of directly solving the Schrödinger equation, forces and energies are inexpensively "predicted" by the ML model and used to generate the next configuration. The accuracy of the model is strictly related to the quality of the sampling used to generate the training examples. One of the advantages of the MLMD approach is that a NN model can be trained by using only the energies and then forces can be obtained by deriving the NN with respects to atomic positions. This avoids the problem of the limited availability of analytic gradient implementations for correlated methods (e.g. RPA) for condensed matter applications. To couple the Behler-Parrinello with RPA-type methods it is crucial to decrease the number of configurations used to train the ML model. Indeed, a high level correlated calculation is required for every training example. Similarly to MLPT, this optimal sampling could be obtained from energetic or geometric arguments. For example, a large number of configurations could be generated at first; then the features corresponding to these geometric configurations could be clustered by using unsupervised learning algorithms; finally, a few representatives of each cluster could be randomly chosen to span the configurational space and to train the NN model. For certain systems it could also be possible to exploit the symmetry to "increase" the amount of information used to train the neural network model [23]. Differently from MLPT, the MLMD approach does not apply perturbation theory but rather performs a full molecular dynamics simulation. Accordingly, the MLMD methods could possibly reach a higher level of accuracy and should not be affected by the choice of the starting point as in perturbative approaches. From a practical point of view some libraries to perform MLMD calculations are already openly available [152] or, if necessary, we could develop our own codes by using Tensorflow.

- **Direct calculation of gradients -** As discussed before, directly preforming a molecular dynamics with correlated approximations (e.g. the RPA) requires impractical computational resources and complex algorithms to compute interatomic forces. An implementation of analytic forces has been recently proposed for the random phase approximation (RPA) within a periodic plane-wave basis set framework [116]; to the best of our knowledge this is the only one based on the PW approach. In the context of this project we will couple the ideas developed in Ref. 116 with the efficient algorithms described in Chapter 3 to obtain an efficient implementation to compute the interatomic forces for the RPA and possibly also for the beyond-RPA and MP2 methods. This part of the work will serve two purposes: (1) At least for the some small systems it will be possible to run a full molecular dynamics to compare and benchmark the MLPT and MLMD approaches; (2) The gradients will be used in the context of MLMD to train machine learning models that specifically require this information [23, 152].

These research directions represent different paths heading towards the same goal - the development of a general purpose methodology to compute finite temperature properties using correlated methods, such as the RPA. For the MLPT approach some proof-of-principle results are already available and show great promise. However, most of the ideas proposed here remains unexplored and it is not possible to establish which approach will give the best results in term of accuracy and numerical efficiency. Importantly, the different paths proposed above are not mutually exclusive. For example, the development of different techniques to sample the configurational space could improve both the MLPT and MLDM approaches. Similarly, an implementation to compute interatomic forces could be used to train the models used for MLMD. As the main outcome of our future work, we will establish which methodology is the most general, efficient, and accurate for finite temperature correlated simulations.

Beyond the methodologial development aspect, applications to realistic materials will be considered in the future with the purpose of testing the new methods, establishing them within the scientific community, and shedding new light on important experimental and technological problems. These will include:

- Study of the adsorption of molecules on surfaces or porous materials.

- Study of molecular crystals. These systems, which are important, for example, in pharmaceutical applications, are usually characterized by polymorphs with similar energetic stability. The methods described above will be used to rank the relative stability of different polymorphs.

- Study of few-layer materials and heterostructures of two-dimensional materials. The new methodologies will be used to study the structural parameters of these systems and the interlayer binding energy.

Appendix A

Curriculum Vitæ

General Information

Dario Rocca
Maître de Conférences/Associate Professor at *Université de Lorraine* (September 2012-Present)

Address: Laboratoire LPCT, UMR CNRS 7019
Faculté des Sciences et Technologies, Université de Lorraine, BP 70239,
Boulevard des Aiguillette, 54506 Vandoeuvre-lès-Nancy, France
Phone: +33 6 40 75 01 22
E-mail: dario.rocca@univ-lorraine.fr
Website: `https://dariorocca.wordpress.com/`
Github: `https://github.com/drocca`

Education

- **PhD in Physics** at the **International School for Advanced Studies (SISSA)** (Trieste, Italy), obtained *cum laude* in October 2007
  Dissertation Title: *Time-Dependent Density Functional Perturbation Theory: New algorithms with applications to molecular spectra*
  Supervisors: Prof. Stefano Baroni and Dr. Ralph Gebauer

- **MS in Physics** at the **International School for Advanced Studies (SISSA)** (Trieste, Italy), obtained in October 2004
  Supervisor: Prof. Sandro Sorella

- **BS in Physics** at **Università degli Studi di Pavia** (Italy), obtained *cum laude* in July 2003

Dissertation Title: *Quantum transport in localized basis mesoscopic systems* (in Italian)
Supervisor: Prof. Giuseppe Pastori Parravicini.

Postdoctoral Experience

- **Postdoctoral Associate**, Department of Chemistry, University of California, Davis, December 2007-August 2012.
  Supervisor: Prof. Giulia Galli

Languages

English (full professional proficiency), French (full professional proficiency), Italian (native)

Teaching Experience

Lecturer at the Université de Lorraine, Nancy, for the classes:

- Mécanique ondulatoire/Wave mechanics (academic year 2012/2013)
- Introduction à la physique/Introduction to physics (academic year 2012/2013)
- Mesures physiques 1/Physical measurements 1 (academic years 2012/2013, 2013/2014, and 2014/2015)
- Mécanique du point/Particle mechanics (academic year 2015/2016)
- Méthodes mathématiques et numériques en physique/Mathematical and numerical methods in physics (academics years 2013/2014, 2014/2015, 2015/2016, 2016/2017, and 2017/2018)
- Physique du solide avancée/Advanced solid state physics (academic year 2018/2019, 2019/2020)
- Mathématiques I/Mathematics I (academic year 2019/2020)

Guest lecturer at the University of California, Davis, for the classes:

- Upper division physical chemistry
- Theoretical and computational chemistry

Supervising Activities

- Contributed to the supervision of a PhD student at UC Davis: Yuan Ping (now Assistant Professor at UC Santa Cruz).
- Supervision of PhD students at the University of Lorraine: Julien Claudot (September 2015-July 2018) and Ali Abboud (September 2015-October 2018).
- Supervision of Master students at the University of Lorraine: Fawzi Kaoui (2014), Ali Abboud (2015), Mohamed Barhoumi (visiting student from the University of Monastir-2015), Mohamed Jizzini (2017), Bilal Chehaibou (2018 - 2019).
- Supervision of postdocs at the University of Lorraine: Anant Dixit (January 2016-December 2018), Won June Kim (2017-2019), Mauricio Chagas da Silva (from June 2018 - in collaboration with the University of Bremen), Minho Kim (2019-present).

Grants

As PI:

- ANR "Jeunes Chercheuses et Jeunes Chercheurs" 2015 (French National Research Agency Young Investigator Grant); Title: *Finding Accurate and Scalable Theories for Electronic Correlation (FASTcorrelation)*. Funding starting January 1st 2016; budget 180000 € (no overhead costs).
- CNRS PEPS "Physique théorique et ses interfaces" 2013; Title: *Techniques de théorie quantique des champs appliquées au problème de l'énergie solaire* (Quantum field theory techniques applied to the solar energy problem); awarded 4000 €.
- Project Institut Jean Barriol 2013; Title: *Techniques de théorie quantique des champs appliquées au problème de l'énergie solaire* (Quantum field theory techniques applied to the solar energy problem); awarded 3000 €.

Computer time:

- PRACE 2016; Title: *NEWcorrelation*; budget 4M core hours.
- GENCI 2015-2019; Title: *Modélisation de la structure électronique par calculs ab initio*; budget >1M core hours.

Participant in other two projects funded by the ANR (CARAPASS and LIQUI2D).

Invited Talks (presenter if first author)

1. **D. Rocca**, *Bethe-Salpeter equation without empty electronic states applied to charge-transfer excitations*, International Workshop on Computational Physics and Materials Science, January 7-9, 2010, Shanghai

2. **D. Rocca**, *Bethe-Salpeter equation without empty electronic states applied to charge-transfer excitations*, 15th International Workshop on Computational Physics and Materials Science: Total Energy and Force Methods, January 13-15, 2011, Trieste

3. **D. Rocca**, *Solution of the Bethe-Salpeter equation without empty electronic states*, ES11: the 23rd Annual Workshop on Electronic Structure Methods, June 6-9, 2011, Philadelphia

4. Z. Bai, **D. Rocca**, G. Galli, and G. Zimanyi, *Accelerating ab initio calculations of optical absorption spectra of nanostructures for solar energy applications*, ICIAM 2011, July 18-22, 2011, Vancouver

5. **D. Rocca**, *Optical absorption spectra: Overview of techniques to solve the Bethe-Salpeter equation*, LLNL-UC Davis Electronic Structure Workshop, October 20-21, 2011, Livermore

6. **D. Rocca** *Ab Initio Modeling and Optimization of Solar Energy Materials Using Newly Developed Many-Body Perturbation Theory Approaches*, CECAM Conference: "Energy from the Sun: Computational Chemists and Physicists Take up the Challenge", September 10-14, 2012, Cagliari

7. **D. Rocca** *Computational spectroscopy using many-body perturbation theory: Large scale calculations without virtual orbitals*, CECAM Workshop: “Calculation of Optical Properties of Nanostructures from First Principles”, February 19-22, 2013, Lausanne

8. **D. Rocca** *Computational spectroscopy using many-body perturbation theory: Large scale calculations without virtual orbitals*, APS March Meeting 2013, Baltimore

9. **D. Rocca** *Dielectric matrix formulation of correlation energies in the Random Phase Approximation: Inclusion of exchange effects*, Total Energy and Force Methods 2016, January 11-13, 2016, Luxembourg

10. **D. Rocca** *Ground-state correlation energy within the ACFDT: Theory and implementation*, Quantum Espresso Linear Response Workshop, January 18-21, 2016, Trieste

11. **D. Rocca** *Dielectric matrix formulation of correlation energies in the Random Phase Approximation: Inclusion of exchange effects*, CECAM workshop: "Density- and response density-based models for Intermolecular Interactions in Molecular Assemblies and in Solids", June 20-23, 2016, Nancy
12. **D. Rocca** *Dielectric matrix formulation of correlation energies in the Random Phase Approximation: Inclusion of exchange effects*, Workshop on "Theory and applications of RPA-and-beyond methods in physics and chemistry", May 2-4, 2017, Paris
13. **D. Rocca** *Accurate ground-state correlation energies within the RPA and beyond: Theory and applications to molecules and solids*, CECAM workhop: "Green's function methods: the next generation III", June 9-6, 2017, Toulouse
14. **D. Rocca** *Accurate ground-state correlation energies within the RPA and beyond: Theory and applications to molecules and solids*, EMN Meeting on Computation and Theory 2017, November 6-10 2017, Dubai
15. **D. Rocca** *Accurate ground-state correlation energies within the RPA and beyond: From theory to realistic materials*, Quantum simulations: From Chemistry to materials science, December 17-21 2018, Hong Kong
16. **D. Rocca** *Bridging molecular dynamics and correlated wave-function methods with machine-learning*, The 5th International Conference on Molecular Simulation (ICMS 2019), November 3-6 2019, Jeju (Korea)

Invited Seminars

1. **D. Rocca**, *Bethe-Salpeter equation without empty electronic states applied to charge-transfer excitations*, February 10, 2010, SISSA, Trieste
2. **D. Rocca**, *Bethe-Salpeter equation without empty electronic states applied to charge-transfer excitations*, March 1st, 2010, Lawrence Berkeley National Laboratory
3. **D. Rocca**, *New theoretical tools for the study of absorption spectra of semiconducting nanostructures*, July 14, 2011, CCI Solar Group Meeting
4. **D. Rocca**, *Nouvelles méthodes pour le calcul des états excités et de l'écrantage diélectrique: théorie et applications*, January 18, 2012, Laboratoire CRM^2, Université de Lorraine, Nancy

5. **D. Rocca**, *Nouvelles méthodes pour le calcul des états excités et de l'écrantage diélectrique: théorie et applications*, January 19, 2012, Laboratoire de Chimie Théorique, UPMC, Paris

6. **D. Rocca**, *Ab initio modeling and optimization of solar energy materials using newly developed many-body perturbation theory approaches*, October 11, 2012, Laboratoire CRM^2, Université de Lorraine, Nancy

7. **D. Rocca**, *New Methods and Algorithms for the Calculation of Electronic Excited States within Many-Body Perturbation Theory*, October 19, 2012, Department of Theoretical Physics, Budapest University of Technology and Economics, Budapest

8. **D. Rocca**, *Computational spectroscopy using many-body perturbation theory: Large scale calculations without virtual orbitals*, March 25, 2013, Center for Functional Nanomaterials, Brookhaven National Laboratory

9. **D. Rocca**, *Computational spectroscopy using many-body perturbation theory: Large scale calculations without virtual orbitals*, July 2, 2013, INSA, Rennes

10. **D. Rocca**, *Efficient calculation of random-phase approximation correlation energies using Lanczos chains and an optimal basis set*, Correlation Meeting, May 22, 2014, Ecole Polytechnique, Palaiseau

11. **D. Rocca**, *Efficient calculation of random-phase approximation correlation energies using Lanczos chains and an optimal basis set*, November 10, 2014, University of Erlangen

12. **D. Rocca**, *Efficient calculation of RPA and beyond-RPA correlation energies using Lanczos chains and an optimal basis set*, November 30, 2015, Griffith University, Brisbane

13. **D. Rocca**, *Efficient calculation of the dielectric screening: From electronic excitations to van der Waals interactions*, June 1st, 2016, Bogazici University, Istanbul

14. **D. Rocca**, *Efficient calculation of the dielectric screening: From electronic excitations to van der Waals interactions*, March 2nd, 2017, University of California, Santa Cruz

15. **D. Rocca**, *Accurate ground-state correlation energies within the RPA and beyond: Theory and applications to molecules and solids*, Argonne National Laboratory, August 14, 2017

Invited Lectures

1. **D. Rocca**, *Liouville-Lanczos calculation of the TDDFT spectrum of simple molecules*, Summer School on Materials Modeling from First Principles: Theory and Practice, ICMR, University of California at Santa Barbara, 2009

2. **D. Rocca**, *Liouville-Lanczos calculation of the TDDFT spectrum of simple molecules*, CECAMSpectra2010, Trieste, 2010

Contributed Talks (presenter if first author)

1. L. Huang, **D. Rocca**, S. Baroni, K. E. Gubbins, M. B. Nardelli, *Molecular Design for Organic Solar Cells*, 2008 AICHE Annual Meeting

2. L. Huang, **D. Rocca**, S. Baroni, K. E. Gubbins, M. B. Nardelli, *Rational design of photoactive materials for organic solar cells*, 2008 MRS Fall Meeting

3. **D. Rocca**, D. Lu and G. Galli, *Ab initio calculations of optical spectra by solving the Bethe-Salpeter equation without empty states*, 2009 APS March Meeting

4. Y. Ping, **D. Rocca** and G. Galli, *Ab-initio calculations of optical spectra of silicon nanowires*, 2009 APS March Meeting

5. A. Gali, M. Voros, **D. Rocca**, G. Zimanyi, G. Galli, *Theoretical investigation of the high energy excitations in silicon nanocrystals*, 2009 APS March Meeting

6. M. Voros, A. Gali, **D. Rocca**, G. Zimanyi, and G. Galli, *Effect of Surface Structure and Cluster-cluster Interactions on the Excitation Properties of Silicon Nanocrystals*, 2009 MRS Spring Meeting

7. **D. Rocca**, D. Lu and G. Galli, *Bethe-Salpeter equation without empty electronic states applied to charge-transfer excitations*, 2010 APS March Meeting

8. Y. Ping, **D. Rocca**, D. Lu, and G. Galli, *Ab-initio calculations of absorption spectra of Si nanostructures using iterative techniques to solve the Bethe-Salpeter Equation*, 2010 APS March Meeting

9. D. Lu, Y. Li, **D. Rocca**, H. V. Nguyen, F. Gygi, and G. Galli, *Efficient evaluation of dielectric response functions and calculations of ground and*

excited state properties beyond local Density Functional approaches, 2010 APS March Meeting

10. M. Voros, A. Gali, **D. Rocca**, G. Zimanyi, and G. Galli, *Ab initio theory of impact ionization applied to silicon nanocrystals*, 2011 APS March Meeting
11. **D. Rocca**, D. Lu, H.-V. Nguyen, and G. Galli, *Ab initio calculations of optical absorption spectra: Solution of the Bethe-Salpeter equation within density matrix perturbation theory*, 2011 APS March Meeting
12. H.-V. Nguyen, T. H. Pham, **D. Rocca**, and G. Galli, *Efficient GW calculations using eigenvalue-eigenvector decomposition of the dielectric matrix*, 2011 APS March Meeting
13. Y. Ping, **D. Rocca**, D. Lu, and G. Galli, *Ab-initio calculations of absorption spectra of nanowires by solving the Bethe-Salpeter Equation*, 2011 APS March Meeting
14. M. Voros, **D. Rocca**, G. Zimanyi, G. Galli, and A. Gali, *Impact ionization rates from ab initio calculations*, 2011 DPG Meeting
15. **D. Rocca**, Y. Ping, D. Lu, and G. Galli, *Solution of the Bethe-Salpeter equation without empty electronic states: Applications to solids, nanostructures and molecules*, 2012 APS March Meeting
16. **D. Rocca**, Z. Bai, R.-C. Li, and G. Galli, *Iterative diagonalization of non-Hermitian eigenproblems in time-dependent density-functional and many-body perturbation theory*, 2012 APS March Meeting
17. T. H. Pham, H.-V. Nguyen, **D. Rocca**, and G. Galli, *Calculation of quasi particle energies using a spectral decomposition of the static dielectric matrix: application to molecules and nanoparticles*, 2012 APS March Meeting
18. Y. Ping, Y. Li, **D. Rocca**, F. Gygi, and G. Galli, *Band gap engineering and optical properties of tungsten trioxide*, 2012 APS March Meeting
19. S. Wippermann, T. Li, **D. Rocca**, G. Zimanyi, and G. Galli, *Si Nanoparticles embedded in solid matrices for solar energy conversion: electronic and optical properties from first principle calculations*, 2012 APS March Meeting
20. M. Voros, A. Gali, **D. Rocca**, G. Galli, and G. Zimanyi, *Multiple exciton generation in silicon nanoparticles*, 2012 APS March Meeting

21. S. Wippermann, M. Voros, **D. Rocca**, A. Gali, G. Zimanyi, and G. Galli, *High pressure core structures of Si nanoparticles for solar energy conversion*, 2013 APS March Meeting

22. M. Voros, S. Wippermann, **D. Rocca**, G. Galli, A. Gali, and G. Zimanyi, *Germanium nanoparticles for solar energy conversion*, 2013 APS March Meeting

23. T. A. Pham, H.-V. Nguyen, **D. Rocca**, and G. Galli, *Ab initio calculations of quasiparticle energies of solids, liquids and molecules using a spectral decomposition of the dielectric matrix*, 2013 APS March Meeting

24. **D. Rocca**, *Computational spectroscopy using many-body perturbation theory: Large scale calculations without virtual orbitals*, Réunion Générale du GDRCoDFT, May 21-24, 2013, Guidel-Plages.

25. **D. Rocca**, *Computational spectroscopy using many-body perturbation theory: Large scale calculations without virtual orbitals*, Rencontre des chimistes théoriciens du grand-est, June 7-8, 2013, Reims.

26. **D. Rocca**, *Computational spectroscopy using many-body perturbation theory: Large scale calculations without virtual orbitals*, 18th ETSF Workshop on Electronic Excitations, October 1-4, 2013, Luxembourg.

27. **D. Rocca**, *Random-phase approximation correlation energies from Lanczos chains and an optimal basis set: Theory and applications to the benzene dimer*, Réunion générale du GDR CORREL, November 27-29, 2013, Paris.

28. J. Morbec, **D. Rocca**, B. Pinaud, T. Jaramillo, and G. Galli, *Opto-electronic properties of* Ta_3N_5*: a joint experimental and theoretical study*, 2014 APS March Meeting

29. S. Wippermann, M. Voros, Bálint Somogyi, A. Gali, **D. Rocca**, F. Gygi, G. Galli, and G. T. Zimanyi *High-pressure Phase Ge nanoparticles and Si-ZnS nanocomposites: New Paradigms to Improve the Efficiency of MEG Solar Cells*, 2014 APS March Meeting

30. M. Voros, **D. Rocca**, A. Gali, and G. Galli, *Opto-electronic properties of silicon nanoparticles: Excitation energies, sum rules, and Tamm-Dancoff approximation*, 2014 APS March Meeting

31. **D. Rocca**, *Efficient calculation of correlation energies within the random phase approximation using Lanczos chains and an optimal basis set*, 2014 APS March Meeting

32. **D. Rocca** and G. Galli, *Ab Initio Modeling and Optimization of Solar Energy Materials Using Newly Developed Many-Body Perturbation Theory Approaches*, 10th International Conference of Computational Methods in Sciences and Engineering, April 4-7, 2014, Athens
33. **D. Rocca**, *Efficient calculation of correlation energies within the random phase approximation and beyond using Lanczos chains and an optimal basis set*, The 9th International Conference on Computational Physics, January 7-11, 2015, Singapore
34. **D. Rocca**, B. Mussard, G. Jansen, and J. Angyan, *Dielectric matrix formulation of correlation energies within the Random Phase Approximation: Inclusion of screened exchange effects*, 2015 APS March Meeting
35. **D. Rocca**, B. Mussard, G. Jansen, and J. Angyan, *Dielectric matrix formulation of correlation energies within the Random Phase Approximation: Inclusion of screened exchange effects*, Psi-k 2015 Conference, September 6-10, 2015, San Sebastian
36. **D. Rocca**, *Theoretical Spectroscopy Methods for Ground State Correlation Energies*, ETSF Workshop 2015, December 14-16, 2015, Palaiseau
37. **D. Rocca**, A. Abboud, S. Lebègue, and G. Vaitheeswaran, *Two-dimensional silicon and carbon monochalcogenides with the structure of phosphorene*, 2016 APS March Meeting, March 14-18, 2016, Baltimore
38. **D. Rocca**, *Improving the accuracy of ground-state correlation energies within a plane-wave basis set: The electron-hole exchange kernel*, Interregional Workshop, October 13-14, 2016, Strasbourg
39. A. Dixit, M. Badawi, T. Bucko, S. Lebègue, J. Angyan, and **D. Rocca**, *Accurate ground-state correlation energies within the RPA and beyond: Theory and applications to molecules and zeolites*, Chemical Physics Conference, October 17-20, 2016, Nancy
40. **D. Rocca**, A. Dixit, and J. Angyan, *Improving the accuracy of ground-state correlation energies within a plane-wave basis set: The electron-hole exchange kernel*, 2017 APS March Meeting, March 13-17 2017, New Orleans
41. A. Dixit, M. Badawi, T. Bucko, S. Lebègue, J. Angyan, and **D. Rocca**, *Zero and finite temperature adsorption energies in zeolites: A random phase approximation study*, 2017 APS March Meeting, March 13-17 2017, New Orleans

42. **D. Rocca**, *Accurate ground-state correlation energies within the RPA and beyond: Theory and applications to molecules and zeolites*, XXIX IUPAP Conference in Computational Physics (CCP2017), July 9-13 2017, Paris
43. M. Badawi, T. Bucko, J. G. Angyan, A. Dixit, S. Lebègue, and **D. Rocca**, *Adsorption enthalpies of hydrocarbons in zeolites: Assessment of van der Waals correction schemes in DFT and AIMD simulations*, 7th FEZA Conference "The Zeolites: Materials with Engineered Properties", July 3-7 2017, Sofia
44. Y. Ping, R. Sundararaman, **D. Rocca**, and F. Wu, *First-principles Engineering of Charged Defects for Two-dimensional Quantum Technologies*, 2018 APS March Meeting, Los Angeles
45. **D. Rocca**, A. Dixit, J. Claudot, and S. Lebègue, *Improving the efficiency of beyond-RPA methods within the dielectric matrix formulation: Algorithms and applications to the A24 and S22 test sets*, 2018 APS March Meeting, Los Angeles
46. **D. Rocca**, *First-Principles Engineering of Charged Defects for Two-Dimensional Quantum Technologies*, GDR-REST: The second general meeting, May 22-25 2018, Porquerolles
47. **D. Rocca**, A. Dixit, M. Badawi, S. Lebègue, T. Gould, and T. Bucko, *Bridging molecular dynamics and correlated wave-function methods for accurate finite-temperature properties*, 2019 APS March Meeting, Boston
48. F. Wu, **D. Rocca**, and Y. Ping, *Dimensionality Dependence of Radiative Recombination in Black Phosphorus from First-principles*, 2019 APS March Meeting, Boston

Posters (presenter if first author)

1. **D. Rocca**, R. Gebauer, Y. Saad, and S. Baroni, *A novel approach to Time-Dependent DFT based on Lanczos chains*, Modeling Materials at the Nanoscale, Palau (Italy), September 21-23, 2006
2. **D. Rocca**, F. De Angelis, R. Gebauer, and S. Baroni, *Time-Dependent DFT study of squaraine dye-sensitized solar cells*, Gordon Conference on Time-Dependent Density-Functional Theory, Waterville (ME, USA), July 15-20, 2007; Time-Dependent Density-Functional Theory: Prospects and Applications, Benasque (Spain), September 10-15, 2008

3. **D. Rocca**, D. Lu, Y. Ping, A. Aliano, and G. Galli, *Optical properties of nanostructures from novel density matrix based methods*, 21th Annual Workshop on Recent Developments in Electronic Structure Methods, University of California, Davis, June 23-25, 2009; Gordon Conference on Time-Dependent Density-Functional Theory, New London (NH, USA), July 5-10, 2009
4. Y. Li, D. Lu, H.-V. Nguyen, **D. Rocca**, and G. Galli, *Van der Waals interactions in weakly-bonded molecular systems from first principles: An adiabatic connection fluctuation-dissipation theorem approach*, 21th Annual Workshop on Recent Developments in Electronic Structure Methods, University of California, Davis, June 23-25, 2009; Gordon Conference on Time-Dependent Density-Functional Theory, New London (NH, USA), July 5-10, 2009
5. Y. Ping, **D. Rocca**, D. Lu, Y. Ping, and G. Galli, *Ab initio calculations of optical spectra of silicon nanostructures*
6. **D. Rocca**, Y. Ping, D. Lu, T. A. Pham, H.-V. Nguyen, and G. Galli, *Ab initio calculations of optical absorption spectra: Solution of the Bethe-Salpeter equation within density matrix perturbation theory*, Gordon Conference on Time-Dependent Density-Functional Theory, Biddeford (ME, USA), August 14-19, 2011
7. Y. Ping, **D. Rocca**, D. Lu, Y. Ping, and G. Galli, *Ab-initio calculations of optical spectra of silicon nanowires by solving the Bethe-Salpeter equation*, Gordon Conference on Time-Dependent Density-Functional Theory, Biddeford (ME, USA), August 14-19, 2011
8. H.-V. Nguyen, T. A. Pham, **D. Rocca**, and G. Galli, *Efficient GW calculations using eigenvalue-eigenvector decomposition of the dielectric matrix*, Gordon Conference on Time-Dependent Density-Functional Theory, Biddeford (ME, USA), August 14-19, 2011
9. S. Wippermann, M. Voros, **D. Rocca**, A. Gali, G. Zimanyi, and G. Galli, *Multiple Exciton Generation in Si Nanoparticles under pressure*, 2012 APS March Meeting
10. **D. Rocca**, *Ab initio calculations of electronic excitations within many-body perturbation theory*, Journée Jean Barriol, October 30, 2012
11. **D. Rocca**, *Random-phase approximation correlation energies from Lanczos chains and an optimal basis set: Theory and applications to the benzene dimer*, 6th Workshop on Time-Dependent Density-Functional Theory:

Prospects and Applications, Benasque (Spain), January 14-17, 2014; Low-scaling and Unconventional Electronic Structure Techniques Conference (LUEST), Telluride (CO, USA), June 1-5, 2014

12. **D. Rocca**, B. Mussard, G. Jansen, and J. Angyan, *Dielectric matrix formulation of correlation energies within the Random Phase Approximation: Inclusion of screened exchange effects*, 17th International Workshop on Computational Physics and Materials Science: Total Energy and Force Methods, Trieste, January 15-17, 2015; Green's function methods: The next generation II, Lausanne, May 4-7, 2015

13. **D. Rocca**, A. Dixit, J. Claudot, and S. Lebègue, *A novel implementation to compute MP2 correlation energies without basis set superposition errors and complete basis set extrapolation*, APS March Meeting, March 5-9 2018, Los Angeles

Awards and Fellowships

- Prime d'encadrement doctoral et de recherche - PEDR (2017-2021): Salary bonus obtained upon evaluation of research and supervising activities.
- Lorraine Université d'Excellence (LUE) Future Leader - ERC track (2018 and 2019)
- Chaire d'Excellence CNRS (2012-2017)
- Two travel grants from the Materials Computation Center (MCC) at the University of Illinois Urbana-Champaign (2008, 2011)
- Undergraduate fellowship of the University of Pavia (1998-2002)
- Collegio Borromeo fellowship for undergraduate student (1998-2002)
- Undergraduate fellowship of the Institute for Advanced Study (IUSS), Pavia (1998-2002).

Memberships

American Physical Society, 2008-present

Professional Services

- Peer-reviewer for major scientific journals, including Physical Review Letters, Physical Review B, Journal of Chemical Physics, Applied Physics Letters, Journal of Physical Chemistry, Journal of Chemical Theory and Computation, Journal of Molecular Modeling, Journal of Computational Chemistry, Molecular Physics, Nanoscale, ACS Omega, and Scientific Reports (Nature).
- Co-organizer of the CECAM workshop “Density- and response density-based models for Intermolecular Interactions in Molecular Assemblies and in Solids” in 2016.

Appendix B

Publications

Citation metrics available at `http://scholar.google.com/citations?user=BzKIlMEAAAAJ`

1. S. Sorella, M. Casula, and **D. Rocca**, *Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods*, J. Chem. Phys. 127, 014105 (2007)

2. **D. Rocca**, R. Gebauer, Y. Saad, and S. Baroni, *Turbo charging time-dependent density-functional theory with Lanczos chains*, J. Chem. Phys. 128, 154105 (2008)

3. L. Huang, **D. Rocca**, S. Baroni, K. E. Gubbins, M. Buongiorno Nardelli, *Molecular design of photoactive acenes for organic photovoltaics*, J. Chem. Phys. 130, 194701 (2009)

4. **D. Rocca**, R. Gebauer, F. De Angelis, M. K. Nazeeruddin, and S. Baroni, *Time-dependent density functional theory study of squaraine dye-sensitized solar cells*, Chem. Phys. Lett. 475, 49 (2009)

5. D. Lu, Y. Li, **D. Rocca**, G. Galli, *Ab initio Calculation of van der Waals Bonded Molecular Crystals*, Phys. Rev. Lett. 102, 206411 (2009)

6. A. Gali, M. Voros, **D. Rocca**, G. T. Zimanyi, and G. Galli, *High-Energy Excitations in Silicon Nanoparticles*, Nanolett. 9, 3780 (2009); highlighted on MRS Bulletin 35, January 2010.

7. **D. Rocca**, D. Lu, and G. Galli, *Ab initio calculations of optical absorption spectra: Solution of the Bethe-Salpeter equation within density matrix perturbation theory*, J. Chem. Phys. 133, 164109 (2010); selected for the November 1, 2010 issue of Virtual Journal of Biological Physics Research.

8. O. B. Malcioglu, R. Gebauer, **D. Rocca**, and S. Baroni, *turboTDDFT - A code for the simulation of molecular spectra using the Liouville-Lanczos approach to time-dependent density-functional perturbation theory*, Comput. Phys. Commun. 182, 1744 (2011)
9. **D. Rocca**, Z. Bai, R.-C. Li, and G. Galli, *A block variational procedure for the iterative diagonalization of non-Hermitian random-phase approximation matrices*, J. Chem. Phys. 136, 034111 (2012)
10. **D. Rocca**, Y. Ping, R. Gebauer, and G. Galli, *Solution of the Bethe-Salpeter equation without empty electronic states: Application to bulk systems*, Phys. Rev. B 85, 045116 (2012)
11. Y. Ping, **D. Rocca**, D. Lu, and G. Galli, *Ab-initio calculations of absorption spectra of semiconducting nanowires within many-body perturbation theory*, Phys. Rev. B 85, 035316 (2012)
12. H.-V. Nguyen, T. A. Pham, **D. Rocca**, and G. Galli, *Improving accuracy and efficiency of calculations of photoemission spectra within many-body perturbation theory*, Phys. Rev. B: Rapid Communications 85, 081101(R) (2012)
13. S. Wippermann, M. Voros, **D. Rocca**, A. Gali, G. Zimanyi, and G. Galli, *High-Pressure Core Structures of Si Nanoparticles for Solar Energy Conversion*, Phys. Rev. Lett. 110, 046804 (2013)
14. Y. Ping, **D. Rocca**, and G. Galli, *Electronic excitations in light absorbers for photoelectrochemical energy conversion: first principles calculations based on many body perturbation theory*, Chem. Soc. Rev. 42, 2437 (2013)
15. T. A. Pham, H.-V. Nguyen, **D. Rocca**, and G. Galli, *GW calculations using the spectral decomposition of the dielectric matrix: Verification, validation, and comparison of methods*, Phys. Rev. B 87, 155148 (2013)
16. M. Voros, **D. Rocca**, G. Galli, G. T. Zimanyi, and A. Gali, *Increasing impact ionization rates in Si nanoparticles through surface engineering: A density functional study*, Phys. Rev. B 87, 155402 (2013)
17. Y. Ping, **D. Rocca**, and G. Galli, *Optical properties of tungsten trioxide from first principles calculations*, Phys. Rev. B 87, 165203 (2013)
18. **D. Rocca**, *Random-phase approximation correlation energies from Lanczos chains and an optimal basis set: Theory and applications to the benzene dimer*, J. Chem. Phys. 140, 18A501 (2014), Special issue Advances in Density Functional Theory.

19. M. Voros, S. Wippermann, Bálint Somogyi, Adam Gali, **D. Rocca**, G. Galli, and G. T. Zimanyi, *Germanium nanoparticles with non-diamond core structures for solar energy conversion*, J. Mater. Chem. A 2, 9820-9827 (2014)

20. X. Ge, S. J. Binnie, **D. Rocca**, R. Gebauer, and S. Baroni, *turboTDDFT 2.0 - Hybrid functionals and new algorithms within time-dependent density-functional perturbation theory*, Comput. Phys. Commun. 185, 2080-2089 (2014)

21. **D. Rocca**, M. Voros, A. Gali, and G. Galli, *Ab initio opto-electronic properties of silicon nanoparticles: Excitation energies, sum rules, and Tamm-Dancoff approximation*, J. Chem. Theory Comput. 10, 3290-3298 (2014)

22. A. Mesbah, J. Prakash, **D. Rocca**, Sébastien Lebègue, J. C. Beard, B. A. Lewis, and J. A. Ibers, *Syntheses, crystal structure, and electronic properties of the five ABaMQ$_4$ compounds RbBaPS$_4$, CsBaPS$_4$, CsBaVS$_4$, RbBaVSe$_4$, and CsBaVSe$_4$*, J. Solid State Chem. 233, 217-220 (2016)

23. J. Prakash, A. Mesbah, J. Beard, **D. Rocca**, Sébastien Lebègue, CD Malliakas, and J. A. Ibers,, *Two new ternary chalcogenides Ba$_2$ZnQ$_3$ (Q = Se, Te) with chains of ZnQ$_4$ tetrahedra: syntheses, crystal structure, and optical and electronic properties*, Z. Naturforsch. B 71, 425-429 (2016)

24. F. Kaoui and **D. Rocca**, *Random phase approximation correlation energy using a compact representation for linear response functions: Application to solids*, J. Phys. Condens. Matter 28, 035201 (2016)

25. B. Mussard, **D. Rocca**, G. Jansen, and J. G. Ángyán, *Dielectric matrix formulation of correlation energies in the Random Phase Approximation: inclusion of exchange effects*, J. Chem. Theory Comput. 12, 2191-2202 (2016)

26. A. Dixit, J. G. Ángyán, and **D. Rocca**, *Improving the accuracy of ground-state correlation energies within a plane-wave basis set: The electron-hole exchange kernel*, J. Chem. Phys. 145, 104105 (2016)

27. Mohamed Barhoumi, **Dario Rocca**, Moncef Said, and Sébastien Lebègue, *A first principle study of graphene functionalized with hydroxyl, nitrile, or methyl groups*, J. Chem. Phys. 146, 044705 (2017)

28. **D. Rocca**, A. Abboud, G. Vaitheeswaran, and S. Lebègue, *Two-dimensional silicon and carbon monochalcogenides with the structure of phosphorene*, Beilstein J. Nanotechnol. 8, 1338 (2017)

29. A. Dixit, J. Claudot, Sébastien Lebègue, and **Dario Rocca**, *A novel implementation to compute MP2 correlation energies without basis set superposition errors and complete basis set extrapolation*, J. Chem. Phys. Communications 146, 211102 (2017)

30. P. Giannozzi *et al.*, *Advanced capabilities for materials modelling with Quantum ESPRESSO*, J. Phys. Condens. Matter 29, 465901 (2017)

31. S. Piccinin, **D. Rocca**, and M. Pastore, *The role of solvent in the energy level alignment of dye-sensitized NiO interfaces*, J. Phys. Chem. C 121, 22286 (2017)

32. A. Dixit, J. Claudot, S. Lebègue, and **D. Rocca**, *Improving the efficiency of beyond-RPA methods within the dielectric matrix formulation: Algorithms and applications to the A24 and S22 test sets*, J. Chem. Theory Comput. 13, 5432 (2017)

33. F. Wu, A. Galatas, R. Sundararaman, **D. Rocca**, and Y. Ping, *First-principles Engineering of Charged Defects for Two-dimensional Quantum Technologies*, Phys. Rev. Materials 1, 071001(R) (2017)

34. J. Claudot, W. J. Kim, A. Dixit, H. Kim, T. Gould, **D. Rocca**, and S. Lebègue, *Benchmarking different methods for the description of intermolecular interactions*, J. Chem. Phys. 148, 064112 (2018)

35. A. Dixit, J. Claudot, T. Gould, S. Lebègue, and **D. Rocca**, *Methods for converging correlation energies within the dielectric matrix formalism*, Phys. Rev. B 97, 115104 (2018)

36. E. P. Hessou, W. G. Kanhounnon, **D. Rocca**, H. Monnier, C. Vallières, S. Lebègue, and M. Badawi, *Adsorption of NO, NO_2, CO, H_2O and CO_2 over isolated monovalent cations in faujasite zeolite: a periodic DFT investigation*, Theor. Chem. Acc. 137, 161 (2018)

37. M. Barhoumi, A. Abboud, L. Debbichi, M. Said, T. Björkman, **D. Rocca**, and S. Lebègue, *Electronic and vibrational properties of several two dimensional halides from ab initio calculations*, Beilstein J. Nanotechnol. 10, 823-832 (2019)

38. **D. Rocca**, A. Dixit, M. Badawi, S. Lebègue, T. Gould, and T. Bucko, *Bridging Molecular Dynamics and Correlated Wave-function Methods for Accurate Finite Temperature Properties*, Phys. Rev. Materials 3, 040801(R) (2019)

39. F. Wu, **D. Rocca**, and Y. Ping, *Dimensionality and Anisotropicity Dependence of Radiative Recombination in Nanostructured Phosphorene*, J. Mater. Chem. C 7, 12891-12897 (2019)

40. S. Mondal, **D. Rocca**, S. Lebègue, and G. Vaitheeswaran, *Structural, elastic, and quasiparticle bandstructure of 4, 4'-Bis (nitramino) azofurazan from first principles theory* Comput. Mater. Sci. 169, 109081 (2019)

41. T. Ayadi, L. Debbichi, M. Badawi, M. Said, H. Kim, **D. Rocca**, and S. Lebègue, *An ab initio study of the ferroelectric In2Se3/graphene heterostructure*, Physica E: Low-dimensional Systems and Nanostructures 114, 113582 (2019)

42. W. J. Kim, S. Xing, G. Kremer, M. Sicot, B. Kierren, D. Malterre, G. Contini, J. Rault, P. Le Fevre, F. Bertran, **D. Rocca**, Y. Fagot-Revurat, and S. Lebègue, *Electronic Structure of Heavy Halogen Atoms Adsorbed on the Cu (111) Surface: A Combined ARPES and First Principles Calculations Study* J. Phys. Chem. C 123, 26309-26314 (2019)

43. B. Chehaibou, M. Badawi, T. Bucko, T. Bazhirov, and **D. Rocca**, *Computing RPA adsorption enthalpies by machine learning thermodynamic perturbation theory* J. Chem. Theory Comput. (2019) DOI: 10.1021/acs.jctc.9b00782

Appendix C

Résumé étendu en français

Ce manuscrit décrit mes activités de recherche principales (mais non exclusives) développées à partir de septembre 2012, lorsque j'ai été embauché en tant que maître de conférences à l'Université de Lorraine. Mon travail a été principalement concentré sur le développement méthodologique pour le calcul des énergies de corrélation basées sur l'approximation de la phase aléatoire (RPA) et de ses variantes.

En raison de son compromis raisonnable entre précision et efficacité, la théorie de la fonctionnelle de la densité (DFT) [70, 77] est devenue la méthode la plus utilisée pour le calcul ab initio des propriétés des matériaux. La DFT s'applique à des systèmes relativement grands comprenant jusqu'à des milliers d'atomes et son succès initial est fortement lié à des approximations assez simples pour la fonctionnelle d'échange-corrélation (xc), comme les approximations de la densité locale (LDA) [77, 108] et des gradients généralisés (GGA) [106]. Des classes de fonctionnelles plus récentes incluent la meta-GGA [143, 140] et les hybrides, qui mélangent un certain pourcentage d'échanges de Hartree-Fock avec des contributions semi-locales [7]. Toutes ces méthodes impliquent un niveau croissant de précision mais également de coût numérique du calcul. Le schéma de l'échelle de Jacob des approximations DFT (voir Fig. C.1) a été proposé par John Perdew pour visualiser et classifier ces différentes approches [107]. Bien que plusieurs formes pour la fonctionnelle d'échange-corrélation aient été proposées, la recherche dans ce domaine est toujours très active, avec le but ultime d'obtenir une "théorie finale". Des classes de systèmes problématiques pour la DFT traditionnelle incluent les matériaux où la corrélation électroniques est forte ou où les forces de van der Waals jouent un rôle important. De plus, les fonctionnelles d'échange-corrélation traditionnelles n'atteignent pas systématiquement le seuil de la précision chimique (1 kcal/mol).

Ce manuscrit se concentre sur le cinquième et plus haut échelon du schéma de Perdew (Fig. C.1). Plus particulièrement, l'énergie de corrélation électronique

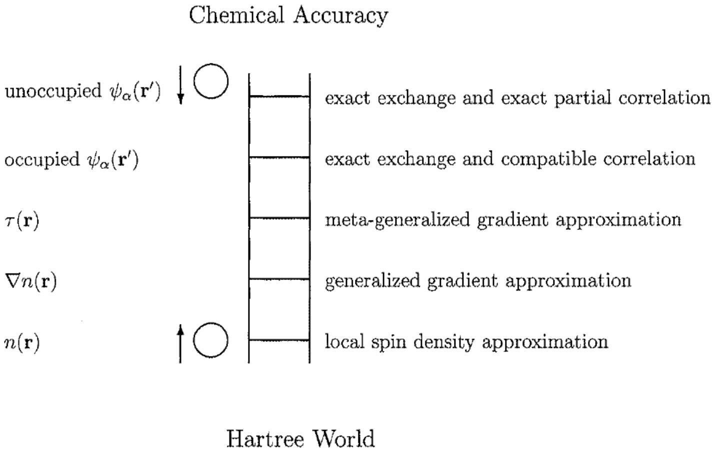


FIGURE C.1 : L'"échelle de Jacob" proposée par John Perdew pour visualiser les différents niveaux d'approximation utilisées pour les fonctionnelles d'échange-corrélation en DFT. En grimpant les échelons la précision augmente au détriment du coût de calcul, qui augmente considérablement (figure reproduite de Ref. 107).

de l'état fondamental est considérée dans l'approximation de la phase aléatoire (RPA) et ses variantes. La RPA avait déjà été introduite dans les années 50 pour décrire l'énergie de corrélation dans le gaz d'électrons dégénéré [14] et son expansion en diagrammes de Feynman est connue depuis longtemps (voir Fig. C.2) [45]. Dans le contexte de la DFT, la base théorique pour la RPA est fournie par le théorème de fluctuation-dissipation avec connexion adiabatique (ACFDT) [80, 81], qui est présenté au chapitre 2. L'ACFDT fournit en principe une expression exacte pour l'énergie de corrélation de l'état fondamental en utilisant des fonctions de réponse linéaires (c'est-à-dire la polarisabilité ou la matrice diélectrique). Dans le contexte de l'ACFDT la RPA est l'approximation pratique la plus simple qui peut être utilisée pour des applications numériques à des systèmes réalistes, tels que des molécules [39, 144, 100, 153, 101, 35, 117, 124], des solides [58, 85, 151, 60, 74], des matériaux en couches minces [92, 82] et des molécules adsorbées sur des surfaces [118, 133]. En particulier, pour les applications à la matière condensée les méthodes de chimie quantique traditionnelles (par exemple la théorie du cluster couplé) peuvent être difficilement utilisées et la RPA est probablement l'approximation corrélée la plus sophistiquée d'intérêt pratique pour ces systèmes. L'intérêt croissant suscité par la RPA est principalement (mais pas exclusivement) lié à la capacité de la RPA de décrire avec haute précision les forces de van der Waals (vdW) [32, 41]. Malgré son succès croissant, la RPA reste une approche de niche, surtout si on la compare à d'autres approximations DFT qui sont appliqués couramment dans les simulations des matériaux. Ceci est lié à différentes raisons : (1)

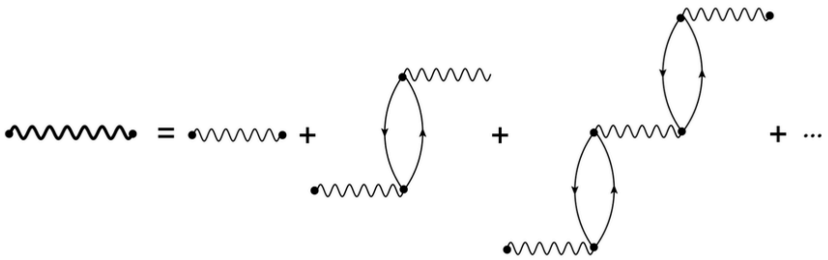


FIGURE C.2 : L'interaction effective au sein de la RPA est obtenue par la somme à l'ordre infini des diagrammes “à anneau” (figure reproduite de Ref. 75).

Bien qu'il s'agisse d'une théorie très sophistiquée, des problèmes subsistent au sein de la RPA : les courbes de liaison de certains dimères diatomiques présentent une "bosse" non physique à des distances intermédiaires [39], les énergies de corrélation totales sont généralement surestimés [78] et les énergies de liaison sont souvent sous-estimées [41, 59]. (2) Les implémentations numériques de la RPA sont plutôt complexes et les calculs pratiques relèvent plusieurs défis, tels que la diagonalisation complète de l'hamiltonien pour obtenir tous les états virtuels (de conduction), la manipulation de matrices de grande taille correspondantes aux fonctions de réponse et l'évaluation d'intégrales impropres sur la fréquence. (3) Le temps de calcul nécessaire pour la RPA est souvent supérieur de plusieurs ordres de grandeur par rapport au temps requis par les approximations (semi-)locales de la DFT ; de plus, le développement d'une implémentation pour le calcul des forces interatomiques analytiques n'est pas banale. Pour ces raisons, les applications aux matériaux réels exigent des ressources de calcul considérables et généralement sont basées sur des géométries fixes optimisées en utilisant des différentes théories.

Dans ce manuscrit, je présente des nouvelles méthodologies pour surmonter ces difficultés. Au chapitre 2, je discute de mes travaux sur le développement de nouvelles approximations qui évitent le problème d'autocorrélation de la RPA en utilisant un noyau d'échange non local. Plus précisément, j'ai développé deux nouvelles approximations pour l'énergie de corrélation de l'état fondamental basées sur l'ACFDT : le “electron-hole time-dependent Hartree-Fock” (eh-TDHF) et le “ adiabatic connection second order screened exchange” (AC-SOSEX) [99, 28, 31]. Au chapitre 3 je présente l'implémentation numérique que j'ai développé pour le calcul des énergies de corrélation pour les méthodes RPA est au-delà en utilisant une base d'ondes planes (PW) et des algorithmes itératifs. La taille de cette base peut atteindre plusieurs centaines de milliers d'éléments pour les grandes mailles. Pour cette raison, il est pratique de représenter les fonctions de réponse impliquées dans les méthodes RPA et au-delà de la RPA à l'aide d'une base auxiliaire compacte. Dans ce manuscrit on utilise les vecteurs propres d'une polarisabilité approximative ne comprenant que la contribution de l'énergie cinétique [124]. La convergence des énergies de corrélation par rapport à la base n'est pas banale et est analysée en détail au chapitre 5. Afin d'évaluer efficacement les éléments de matrice de la polarisabilité sur la base auxiliaire il est utile de recourir à des techniques numériques supplémentaires : l'élimination des états de conduction pour obtenir un formalisme qui ne dépend que des états occupés [3, 147, 126] et l'utilisation de l'algorithme de Lanczos pour calculer le résolvant de l'Hamiltonien [126, 146, 102]. Au chapitre 4 cette implémentation numérique est appliquée à une série d'ensembles moléculaires pour lesquels des résultats de haut niveau sont disponibles et à des solides. La Fig. C.3 montre un exemple d'application au calcul de l'énergie d'interaction des 22 dimères moléculaires contenus dans l'ensemble S22. Les approximations SOSEX et eh-TDHF en incluant la contribution d'échange éliminent l'effet d'auto-corrélation de la RPA et conduisent à des résultats beaucoup plus précis. Ces résultats démontrent qu'un niveau élevé de précision peut être atteint pour les énergies de l'état fondamental dans une formulation basée uniquement sur des fonctions de réponse.

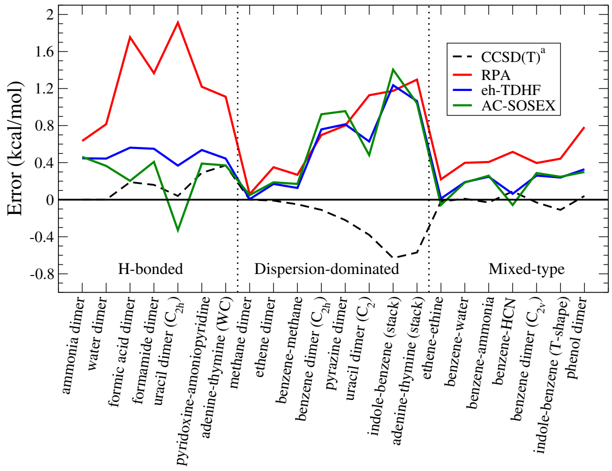


FIGURE C.3 : Différences en kcal/mol pour l'ensemble S22 pour les méthodes RPA, eh-TDHF et AC-SOSEX calculées par rapport aux valeurs de référence obtenus pas la théorie du cluster couplé CCSD(T) [142]

Les méthodologies RPA et au-delà présentées dans ce manuscrit fournissent une efficacité élevée et une grande précision pour le calcul des énergies de corrélation électronique. Cependant, en considérant que le coût numérique est considérablement supérieur à celui des approximations traditionnelles de la DFT, l'applications systématique de ces approches à des matériaux réels est particulièrement difficile. De plus, les gradients analytiques pour l'optimisation structurelle et la dynamique moléculaire impliquent des difficultés additionnelles pour l'implémentation numérique et augmentent davantage le temps de calcul. À cette fin j'ai développé une nouvelle technique basée sur l'apprentissage automatique (machine learning - ML) pour calculer les propriétés de matériaux aux températures finies à partir seulement de quelques calculs RPA. Cette méthode, appelée "machine learning thermodynamic perturbation theory" (MLPT) [20], est décrite au chapitre 6. Dans la méthode MLPT, une dynamique moléculaire *ab initio* (AIMD) est réalisée à l'aide d'une fonctionnelle DFT numériquement peu coûteuse. Cette simulation implique généralement la génération de plusieurs centaines de milliers de configurations. Parmi ces configurations un jeu d'entrainement est sélectionné pour former un modèle d'apprentissage automatique. Pour toutes les autres configurations générées par la dynamique moléculaire initiale, le modèle de ML est utilisé pour prédire de manière peu coûteuse les énergies RPA. La distribution de probabilité correspondante à la RPA est reconstruite à partir de la théorie de la perturbation thermodynamique en changeant les poids statistiques des configurations échantillonnées par la MD initiale et utilisée pour évaluer les observables à température finie. L'avantage de cette méthode est clair : en ne sélectionnant que quelques dizaines de configurations significatives sur des centaines de milliers,

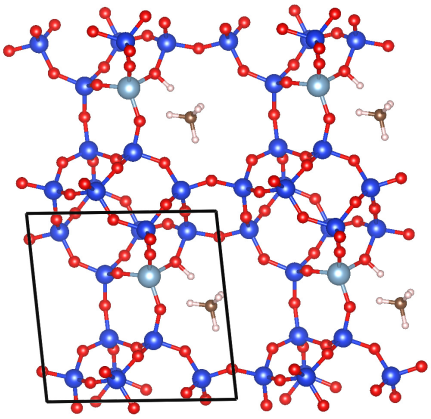


FIGURE C.4 : Modèle périodique utilisé pour CH_4 adsorbé dans la chabazite protonée. La cellule unitaire utilisée dans les calculs est représentée par des lignes continues.

TABLE C.1 : Les energies internes et enthalpie d'adsorption (kcal/mol) au niveau de théorie RPA de la chabazite protonée (HCHAB), de CH_4 et du système adsorbé (HCHAB+CH_4) selon les estimations de MLPT pour différentes taille du jeu d'entrainement.

N_{train}	HCHAB+CH_4	HCHAB	CH_4	Δ_{ads} H
10	-21398.01 ± 0.19	-20213.82 ± 0.13	-1181.05 ± 0.02	-3.74 ± 0.23
25	-21398.13 ± 0.21	-20213.94 ± 0.13	-1181.07 ± 0.02	-3.72 ± 0.25
50	-21398.17 ± 0.21	-20213.99 ± 0.13	-1181.06 ± 0.02	-3.72 ± 0.25
100	-21398.13 ± 0.21	-20213.93 ± 0.12	-1181.06 ± 0.02	-3.74 ± 0.24
200	-21398.15 ± 0.20	-20213.95 ± 0.12	-1181.06 ± 0.02	-3.74 ± 0.23

le temps de calcul nécessaire pour appliquer la théorie de la perturbation avec la RPA est réduit de plusieurs ordres de grandeur. Les résultats montrant la précision et l'efficacité de la méthode MLPT sont présentés dans le tableau C.1. Plus précisément, l'enthalpie d'adsorption du méthane dans un matériau poreux (la zéolite chabazite) a été calculée en utilisant la RPA. Le modèle périodique correspondant, qui implique 200 électrons de valence, est présenté en figure C.4. Comme le montre le tableau C.1, alors que la fonctionnelle de départ PBE+D2 donne un résultat qui dépasse largement le seuil de précision chimique (1 kcal / mol), la valeur de la RPA est en excellent accord avec la valeur expérimentale. En conséquence, l'utilisation de MLPT associé à la RPA ou à d'autres méthodes corrélées au-delà de la RPA peut atteindre un niveau de précision beaucoup plus élevé tout en limitant dans une large mesure l'investissement en temps de calcul.

Bibliography

[1] ALBRECHT, S., REINING, L., DEL SOLE, R., AND ONIDA, G. Ab initio calculation of excitonic effects in the optical spectra of semiconductors. *Phys. Rev. Lett. 80*, 20 (1998), 4510–4513.

[2] ÁNGYÁN, J. G., LIU, R.-F., TOULOUSE, J., AND JANSEN, G. Correlation energy expressions from the adiabatic-connection fluctuation–dissipation theorem approach. *Journal of Chemical Theory and Computation 7*, 10 (2011), 3116–3130.

[3] BARONI, S., DE GIRONCOLI, S., DAL CORSO, A., AND GIANNOZZI, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys. 73*, 2 (2001), 515–562.

[4] BARTLETT, R. J., AND MUSIAL, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys. 79*, 1 (2007), 291.

[5] BARTÓK, A. P., KONDOR, R., AND CSÁNYI, G. On representing chemical environments. *Phys. Rev. B 87*, 18 (2013), 184115.

[6] BATES, J. E., AND FURCHE, F. Communication: Random phase approximation renormalized many-body perturbation theory. *J. Chem. Phys. 139*, 17 (2013), 171103.

[7] BECKE, A. D. A new mixing of hartree–fock and local density-functional theories. *J. Chem. Phys. 98*, 2 (1993), 1372–1377.

[8] BEHLER, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys. 134*, 7 (2011), 074106.

[9] BEHLER, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys. 145*, 17 (2016), 170901.

[10] BEHLER, J., AND PARRINELLO, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett. 98*, 14 (2007), 146401.

[11] BENIGHAUS, T., DISTASIO JR., R. A., LOCHAN, R. C., CHAI, J.-D., AND HEAD-GORDON, M. Semiempirical Double-Hybrid Density Functional with Improved Description of Long-Range Correlation. *J. Phys. Chem. A 112* (2008), 2702–2712.

[12] BJÖRKMAN, T., GULANS, A., KRAHENINNIKOV, A. V., AND NIEMINEM, R. M. van der Waals Bonding in Layered Compounds from Advanced Density-Functional First-Principles Calculations. *Phys. Rev. Lett. 108* (2012), 235502.

[13] BLÖCHL, P. E. Projector augmented-wave method. *Physical Review B 50*, 24 (1994), 17953.

[14] BOHM, D., AND PINES, D. A collective description of electron interactions: Iii. coulomb interactions in a degenerate electron gas. *Phys. Rev. 92* (Nov 1953), 609–625.

[15] BOOTH, G. H., GRÜNEIS, A., KRESSE, G., AND ALAVI, A. Towards an exact description of electronic wavefunctions in real solids. *Nature 493*, 7432 (2013), 365.

[16] BRUNEVAL, F., SOTTILE, F., OLEVANO, V., AND REINING, L. Beyond time-dependent exact exchange: The need for long-range correlation. *The Journal of chemical physics 124*, 14 (2006), 144113.

[17] BUČKO, T., CHIBANI, S., PAUL, J.-F., CANTREL, L., AND BADAWI, M. Dissociative iodomethane adsorption on ag-MOR and the formation of AgI clusters: an ab initio molecular dynamics study. *Phys. Chem. Chem. Phys. 19*, 40 (Oct. 2017), 27530–27543.

[18] BUČKO, T., AND HAFNER, J. The role of spatial constraints and entropy in the adsorption and transformation of hydrocarbons catalyzed by zeolites. *J. Catal. 329* (Sept. 2015), 32–48.

[19] BURNS, L. A., MARSHALL, M. S., AND SHERRILL, C. D. Comparing Counterpoint-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *J. Chem. Theory. Comput. 10* (2014), 49.

[20] Chehaibou, B., Badawi, M., Bucko, T., Bazhirov, T., and Rocca, D. Computing rpa adsorption enthalpies by machine learning thermodynamic perturbation theory. *J. Chem. Theory Comput. 15* (2019), 6333–6342.

[21] Chibani, S., Chebbi, M., Lebègue, S., Bučko, T., and Badawi, M. A DFT investigation of the adsorption of iodine compounds and water in h-, na-, ag-, and cu- mordenite. *J. Chem. Phys. 144*, 24 (June 2016), 244705.

[22] Chipot, C., and Pohorille, A. *Free energy calculations: Theory and Applications in Chemistry and Biology*. Springer, 2016.

[23] Chmiela, S., Sauceda, H. E., Müller, K.-R., and Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun. 9*, 1 (2018), 3887.

[24] Chmiela, S., Tkatchenko, A., Sauceda, H. E., Poltavsky, I., Schütt, K. T., and Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv. 3*, 5 (2017), e1603015.

[25] Colonna, N., Hellgren, M., and de Gironcoli, S. Correlation energy within exact-exchange adiabatic connection fluctuation-dissipation theory: Systematic development and simple approximations. *Phys. Rev. B 90*, 12 (2014), 125150.

[26] De, S., Bartók, A. P., Csányi, G., and Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys. 18*, 20 (2016), 13754–13769.

[27] DiStasio Jr., R. A., and Head-Gordon, M. Optimized spin-component scaled second-order Møller-Plesset perturbation theory for intermolecular interaction energies. *Mol. Phys. 105* (2007), 1073–1083.

[28] Dixit, A., Ángyán, J. G., and Rocca, D. Improving the accuracy of ground-state correlation energies within a plane-wave basis set: The electron-hole exchange kernel. *J. Chem. Phys. 145*, 10 (2016), 104105.

[29] Dixit, A., Claudot, J., Gould, T., Lebègue, S., and Rocca, D. Methods for converging correlation energies within the dielectric matrix formalism. *Phys. Rev. B 97*, 11 (2018), 115104.

[30] Dixit, A., Claudot, J., Lebègue, S., and Rocca, D. Communication: A novel implementation to compute mp2 correlation energies without

basis set superposition errors and complete basis set extrapolation. *J. Chem. Phys. 146*, 21 (2017), 211102.

[31] DIXIT, A., CLAUDOT, J., LEBÈGUE, S., AND ROCCA, D. Improving the efficiency of beyond-rpa methods within the dielectric matrix formulation: Algorithms and applications to the a24 and s22 test sets. *J. Chem. Theory Comput. 13*, 11 (2017), 5432–5442.

[32] DOBSON, J. F., AND WANG, J. Successful test of a seamless van der waals density functional. *Phys. Rev. Lett. 82* (Mar 1999), 2123–2126.

[33] DORNER, F., SUKURMA, Z., DELLAGO, C., AND KRESSE, G. Melting si: beyond density functional theory. *Phys. Rev. Lett. 121*, 19 (2018), 195701.

[34] ESHUIS, H., BATES, J. E., AND FURCHE, F. Electron correlation methods based on the random phase approximation. *Theoretical Chemistry Accounts 131*, 1 (2012), 1–18.

[35] ESHUIS, H., YARKONY, J., AND FURCHE, F. Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration. *J. Chem. Phys. 132*, 23 (June 2010), 234114.

[36] FELLER, D., PETERSON, K. A., AND CRAWFORD, T. D. Sources of error in electronic structure calculations on small chemical systems. *J. Chem. Phys. 124* (2006), 054107.

[37] FLYVBJERG, H., AND PETERSEN, H. Error estimates on averages of correlated data. *J. Chem Phys. 91*, 1 (1989), 461–466.

[38] FRENKEL, D., AND SMIT, B. *Understanding molecular simulation: from algorithms to applications*. Academic press: San Diego, 2002.

[39] FUCHS, M., AND GONZE, X. Accurate density functionals: approaches using the adiabatic-connection fluctuation-dissipation theorem. *Phys. Rev. B 65* (Jun 2002), 235109.

[40] FUCHS, M., AND SCHEFFLER, M. Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems using density-functional theory. *Comput. Phys. Comm. 119* (1999), 67–98.

[41] FURCHE, F. Molecular tests of the random phase approximation to the exchange-correlation energy functional. *Phys. Rev. B 64* (Oct 2001), 195120.

[42] FURCHE, F., AND VAN VOORHIS, T. Fluctuation-dissipation theorem density-functional theory. *J. Chem. Phys. 122*, 16 (Apr. 2005), 164106.

[43] GARCÍA-GONZÁLEZ, P., FERNÁNDEZ, J., MARINI, A., AND RUBIO, A. Advanced correlation functionals: application to bulk materials and localized systems. *J. Phys. Chem. A 111*, 49 (2007), 12458–12465.

[44] GATZ, D. F., AND SMITH, L. The standard error of a weighted mean concentration-i. bootstrapping vs other methods. *Atmospheric Environ. 29*, 11 (1995), 1185 – 1193.

[45] GELL-MANN, M., AND BRUECKNER, K. A. Correlation energy of an electron gas at high density. *Phys. Rev. 106*, 2 (1957), 364.

[46] GIANNOZZI, P., ANDREUSSI, O., BRUMME, T., BUNAU, O., NARDELLI, M. B., CALANDRA, M., CAR, R., CAVAZZONI, C., CERESOLI, D., COCOCCIONI, M., COLONNA, N., CARNIMEO, I., CORSO, A. D., DE GIRONCOLI, S., DELUGAS, P., DISTASIO, R. A., FERRETTI, A., FLORIS, A., FRATESI, G., FUGALLO, G., GEBAUER, R., GERSTMANN, U., GIUSTINO, F., GORNI, T., JIA, J., KAWAMURA, M., KO, H.-Y., KOKALJ, A., KUÇUKBENLI, E., LAZZERI, M., MARSILI, M., MARZARI, N., MAURI, F., NGUYEN, N. L., NGUYEN, H.-V., DE-LA ROZA, A. O., PAULATTO, L., PONCÉ, S., ROCCA, D., SABATINI, R., SANTRA, B., SCHLIPF, M., SEITSONEN, A. P., SMOGUNOV, A., TIMROV, I., THONHAUSER, T., UMARI, P., VAST, N., WU, X., AND BARONI, S. Advanced capabilities for materials modelling with quantum ESPRESSO. *Journal of Physics: Condensed Matter 29*, 46 (oct 2017), 465901.

[47] GIANNOZZI, P., BARONI, S., BONINI, N., CALANDRA, M., CAR, R., CAVAZZONI, C., CERESOLI, D., CHIAROTTI, G. L., COCOCCIONI, M., DABO, I., DAL CORSO, A., DE GIRONCOLI, S., FABRIS, S., FRATESI, G., GEBAUER, R., GERSTMANN, U., GOUGOUSSIS, C., KOKALJ, A., LAZZERI, M., MARTIN-SAMOS, L., MARZARI, N., MAURI, F., MAZZARELLO, R., PAOLINI, S., PASQUARELLO, A., PAULATTO, L., SBRACCIA, C., SCANDOLO, S., SCLAUZERO, G., SEITSONEN, A. P., SMOGUNOV, A., UMARI, P., AND WENTZCOVITCH, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics Condensed Matter 21* (Sept. 2009), 5502.

[48] GOLL, E., LEININGER, T., MANBY, F. R., MITRUSHCHENKOV, A., WERNER, H.-J., AND STOLL, H. Local and density fitting approxima-

tions within the short-range/long-range hybrid scheme: application to large non-bonded complexes. *Phys. Chem. Chem. Phys. 10* (2008), 3353–3357.

[49] GOULD, T., TOULOUSE, J., ÁNGYÁN, J. G., AND DOBSON, J. F. Casimir–Polder Size Consistency: A Constraint Violated by Some Dispersion Theories. *J. Chem. Theory Comput. 13* (2017), 5829.

[50] GOVONI, M., AND GALLI, G. Large scale gw calculations. *Journal of chemical theory and computation 11*, 6 (2015), 2680–2696.

[51] GRÁFOVÁ, L., PITOŇÁK, M., ŘEZÁČ, J., AND HOBZA, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput. 6* (2010), 2365–2376.

[52] GRAJCIAR, L., HEARD, C. J., BONDARENKO, A. A., POLYNSKI, M. V., MEEPRASERT, J., PIDKO, E. A., AND NACHTIGALL, P. Towards operando computational modeling in heterogeneous catalysis. *Chem. Soc. Rev.* (Sept. 2018).

[53] GRIMME, S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem. 27*, 15 (2006), 1787–1799.

[54] GRUNDEI, M. M. J., AND BUROW, A. M. Random Phase Approximation for Periodic Systems Employing Direct Coulomb Lattice Summation. *J. Chem. Theory Comput. 13* (2017), 1159–1175.

[55] GULANS, A. Towards numerically accurate many-body perturbation theory: Short-range correlation effects. *J. Chem. Phys. 141* (2014), 164127.

[56] HALKIER, A., HELGAKER, T., JØRGENSEN, P., KLOPPER, W., KOCH, H., OLSEN, J., AND WILSON, A. K. Basis-set convergence in correlated calculations on Ne, N_2 and H_2O. *Chem. Phys. Lett. 286* (1998), 243.

[57] HAMANN, D. Optimized norm-conserving vanderbilt pseudopotentials. *Physical Review B 88*, 8 (2013), 085117.

[58] HARL, J., AND KRESSE, G. Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory. *Phys. Rev. B 77* (Jan 2008), 045136.

[59] HARL, J., AND KRESSE, G. Accurate bulk properties from approximate many-body techniques. *Physical review letters 103*, 5 (2009), 056401.

[60] HARL, J., SCHIMKA, L., AND KRESSE, G. Assessing the quality of the random phase approximation for lattice constants and atomization energies of solids. *Phys. Rev. B 81* (Mar 2010), 115126.

[61] HAYDOCK, R., HEINE, V., AND KELLY, M. J. Electronic structure based on the local atomic environment for tight-binding bands. *J. Phys. C 5*, 20 (1972), 2845–2858.

[62] HELGAKER, T., KLOPPER, W., KOCH, H., AND NOGA, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys. 106* (1997), 9639.

[63] HELLGREN, M., COLONNA, N., AND DE GIRONCOLI, S. Beyond the random phase approximation with a local exchange vertex. *Phys. Rev. B 98* (Jul 2018), 045117.

[64] HESSELMANN, A. Comparison of Intermolecular Interaction Energies from SAPT and DFT Including Empirical Dispersion Contributions. *J. Chem. Phys. A 115* (2011), 11321–11330.

[65] HESSELMANN, A. Random-phase-approximation correlation method including exchange interactions. *Phys. Rev. A 85*, 1 (2012), 012517.

[66] HESSELMANN, A., AND GÖRLING, A. Correct description of the bond dissociation limit without breaking spin symmetry by a random-phase-approximation correlation functional. *Phys. Rev. Lett. 106*, 9 (2011), 093001.

[67] HESSOU, E., KANHOUNNON, W., ROCCA, D., MONNIER, H., VALLIÈRES, C., LEBÈGUE, S., AND BADAWI, M. Adsorption of no, no 2, co, h 2 o and co 2 over isolated monovalent cations in faujasite zeolite: A periodic dft investigation. *Theor. Chem. Acc. 137*, 12 (2018), 161.

[68] HILL, J. G., AND PLATTS, J. A. Spin-Component Scaling Methods for Weak and Stacking Interactions. *J. Chem. Theory Comput. 3* (2007), 80–85.

[69] HIMANEN, L., JÄGER, M. O., MOROOKA, E. V., CANOVA, F. F., RANAWAT, Y. S., GAO, D. Z., RINKE, P., AND FOSTER, A. S. Dscribe: Library of descriptors for machine learning in materials science. *arXiv:1904.08875* (2019).

[70] HOHENBERG, P., AND KOHN, W. Inhomogeneous electron gas. *Physical review 136*, 3B (1964), B864.

[71] HUO, H., AND RUPP, M. Unified representation of molecules and crystals for machine learning. *arXiv:1704.06439* (2017).

[72] JÄGER, M. O., MOROOKA, E. V., CANOVA, F. F., HIMANEN, L., AND FOSTER, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *Npj Comput. Mater. 4*, 1 (2018), 37.

[73] JUREČKA, P., ŠPONER, J., ČERNÝ, J., AND HOBZA, P. Benchmark database of accurate (mp2 and ccsd (t) complete basis set limit) interaction energies of small model complexes, dna base pairs, and amino acid pairs. *Physical Chemistry Chemical Physics 8*, 17 (2006), 1985–1993.

[74] KAOUI, F., AND ROCCA, D. Random phase approximation correlation energy using a compact representation for linear response functions: application to solids. *J. Phys. Condens. Matter 28*, 3 (2016), 035201.

[75] KE, L., AND JI-SHENG, C. Dispersion relation of excitation mode in spin-polarized fermi gas. *Chinese Physics B 21*, 3 (2012), 030309.

[76] KHALIL, I., JABRAOUI, H., MAURIN, G., LEBÈGUE, S., BADAWI, M., THOMAS, K., AND MAUGÉ, F. Selective capture of phenol from biofuel using protonated faujasite zeolites with different si/al ratios. *J. Phys. Chem. C 122*, 46 (2018), 26419–26429.

[77] KOHN, W., AND SHAM, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev. 140* (1965), A1133–A1138.

[78] KURTH, S., AND PERDEW, J. P. Density-functional correction of random-phase-approximation correlation with results for jellium surface energies. *Phys. Rev. B 59* (Apr 1999), 10461–10468.

[79] KUTZELNIGG, W., AND MORGAN III, J. D. Rates of convergence of the partial-wave expansions of atomic correlation energies. *J. Chem. Phys. 96*, 6 (1992), 4484.

[80] LANGRETH, D. C., AND PERDEW, J. P. The exchange-correlation energy of a metallic surface. *Solid State Commun. 17* (Dec. 1975), 1425–1429.

[81] LANGRETH, D. C., AND PERDEW, J. P. Exchange-correlation energy of a metallic surface: Wave-vector analysis. *Phys. Rev. B 15* (Mar 1977), 2884–2901.

[82] LEBÈGUE, S., HARL, J., GOULD, T., ÁNGYÁN, J. G., KRESSE, G., AND DOBSON, J. F. Cohesive properties and asymptotics of the dispersion interaction in graphite by the random phase approximation. *Phys. Rev. Lett. 105* (Nov 2010), 196401.

[83] LOSCH, P., JOSHI, H. R., VOZNIUK, O., GRUNERT, A., OCHOA-HERNÁNDEZ, C., JABRAOUI, H., BADAWI, M., AND SCHMIDT, W. Proton mobility, intrinsic acid strength, and acid site location in zeolites revealed by varying temperature infrared spectroscopy and density functional theory studies. *J. Am. Chem. Soc. 140*, 50 (2018), 17790–17799.

[84] LU, D. Evaluation of model exchange-correlation kernels in the adiabatic connection fluctuation-dissipation theorem for inhomogeneous systems. *J. Chem. Phys. 140*, 18 (2014), 18A520.

[85] LU, D., LI, Y., ROCCA, D., AND GALLI, G. Ab initio calculation of van der waals bonded molecular crystals. *Phys. Rev. Lett. 102* (May 2009), 206411.

[86] LU, D., NGUYEN, H.-V., AND GALLI, G. Power series expansion of the random phase approximation correlation energy: The role of the third- and higher-order contributions. *J. Chem. Phys. 133*, 15 (Oct. 2010), 154110.

[87] MAATEN, L. V. D., AND HINTON, G. Visualizing data using t-sne. *J. Mach. Learn. Res. 9*, Nov (2008), 2579–2605.

[88] MAGHSOUDI, H., SOLTANIEH, M., BOZORGZADEH, H., AND MOHAMADALIZADEH, A. Adsorption isotherms and ideal selectivities of hydrogen sulfide and carbon dioxide over methane for the si-cha zeolite: comparison of carbon dioxide and methane adsorption with the all-silica dd3r zeolite. *Adsorption 19*, 5 (2013), 1045–1053.

[89] MALCIOĞLU, O. B., GEBAUER, R., ROCCA, D., AND BARONI, S. turboTDDFT - A code for the simulation of molecular spectra using the Liouville-Lanczos approach to time-dependent density-functional perturbation theory. *Computer Physics Communications 182* (Aug. 2011), 1744–1754.

[90] MARCHETTI, O., AND WERNER, H.-J. Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions. *Phys. Chem. Chem. Phys. 10* (2008), 3400–3409.

[91] MARCHETTI, O., AND WERNER, H.-J. Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2 Method. *J. Phys. Chem. A 113* (2009), 11580–11585.

[92] MARINI, A., GARCIA-GONZALEZ, P., AND RUBIO, A. First-principles description of correlation effects in layered materials. *Phys. Rev. Lett. 96* (Apr 2006), 136404.

[93] MCLACHLAN, A. D., AND BALL, M. A. Time-dependent hartree-fock theory for molecules. *Rev. Mod. Phys. 36*, 3 (1964), 844–855.

[94] MCNAMARA, J. P., AND HILLIER, I. H. Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules. *Phys. Chem. Chem. Phys. 9* (2007), 2362–2370.

[95] MØLLER, C., AND PLESSET, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev. 46*, 7 (1934), 618.

[96] MOORE, K. B., SADEGHIAN, K., SHERRILL, C. D., OCHSENFELD, C., AND SCHAEFER, H. F. C–H$\cdots$O Hydrogen Bonding. The Prototypical Methane-Formaldehyde System: A Critical Assessment. *J. Chem. Theory Comput. 13* (2017), 5379.

[97] MORGADO, C., VINCENT, M. A., HILLIER, I. H., AND SHAN, X. Can the DFT-D method describe the full range of noncovalent interactions found in large biomolecules? *Phys. Chem. Chem. Phys. 9* (2007), 448–451.

[98] MURNAGHAN, F. The compressibility of media under extreme pressures. *Proc. Natl. Acad. Sci. U.S.A. 30*, 9 (1944), 244.

[99] MUSSARD, B., ROCCA, D., JANSEN, G., AND ANGYÁN, J. G. Dielectric matrix formulation of correlation energies in the random phase approximation: Inclusion of exchange effects. *J. Chem. Theory Comput. 12*, 5 (2016), 2191–2202.

[100] NGUYEN, H.-V., AND DE GIRONCOLI, S. Efficient calculation of exact exchange and rpa correlation energies in the adiabatic-connection fluctuation-dissipation theory. *Phys. Rev. B 79* (May 2009), 205114.

[101] NGUYEN, H.-V., AND GALLI, G. A first-principles study of weakly bound molecules using exact exchange and the random phase approximation. *J. Chem. Phys. 132*, 4 (Jan. 2010), 044109.

[102] NGUYEN, H.-V., PHAM, T. A., ROCCA, D., AND GALLI, G. Improving accuracy and efficiency of calculations of photoemission spectra within the many-body perturbation theory. *Phys. Rev. B 85* (Feb 2012), 081101.

[103] NGUYEN, N. L., COLONNA, N., AND DE GIRONCOLI, S. Ab initio self-consistent total-energy calculations within the exx/rpa formalism. *Phys. Rev. B 90*, 4 (2014), 045138.

[104] OLSEN, T., AND THYGESEN, K. S. Extending the random-phase approximation for electronic correlation energies: The renormalized adiabatic local density approximation. *Phys. Rev. B 86* (Aug 2012), 081103.

[105] ONIDA, G., REINING, L., AND RUBIO, A. Electronic excitations: density-functional versus many-body green's-function approaches. *Rev. Mod. Phys. 74*, 2 (2002), 601–659.

[106] PERDEW, J. P., BURKE, K., AND ERNZERHOF, M. Generalized gradient approximation made simple. *Phys. Rev. Lett. 77* (Oct 1996), 3865–3868.

[107] PERDEW, J. P., AND SCHMIDT, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings* (2001), vol. 577, AIP, pp. 1–20.

[108] PERDEW, J. P., AND ZUNGER, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B 23* (May 1981), 5048–5079.

[109] PHAM, T. A., NGUYEN, H.-V., ROCCA, D., AND GALLI, G. G W calculations using the spectral decomposition of the dielectric matrix: Verification, validation, and comparison of methods. *Phys. Rev. B 87* (Apr 2013), 155148.

[110] PHAM, T. D., LIU, Q., AND LOBO, R. F. Carbon dioxide and nitrogen adsorption on cation-exchanged ssz-13 zeolites. *Langmuir 29*, 2 (2012), 832–839.

[111] PICCINI, G., ALESSIO, M., SAUER, J., ZHI, Y., LIU, Y., KOLVENBACH, R., JENTYS, A., AND LERCHER, J. A. Accurate adsorption thermodynamics of small alkanes in zeolites. ab initio theory and experiment for h-chabazite. *J. Phys. Chem. C 119*, 11 (2015), 6128–6137.

[112] PING, Y., ROCCA, D., AND GALLI, G. Electronic excitations in light absorbers for photoelectrochemical energy conversion: first principles calculations based on many body perturbation theory. *Chem. Soc. Rev. 42* (2013), 2437–2469.

[113] PITOŇÁK, M., NEOGRÁDY, P., ČERNÝ, J., GRIMME, S., AND HOBZA, P. Scaled MP3 Non-Covalent Interaction Energies Agree Closely with Accurate CCSD(T) Benchmark Data. *Chem. Phys. Chem. 10* (2009), 282–289.

[114] POHORILLE, A., JARZYNSKI, C., AND CHIPOT, C. Good practices in free-energy calculations. *J. Phys. Chem. B 114*, 32 (2010), 10235–10253.

[115] RAMAKRISHNAN, R., DRAL, P. O., RUPP, M., AND VON LILIENFELD, O. A. Big data meets quantum chemistry approximations: The δ-machine learning approach. *J. Chem. Theory Comput. 11*, 5 (2015), 2087–2096.

[116] RAMBERGER, B., SCHÄFER, T., AND KRESSE, G. Analytic interatomic forces in the random phase approximation. *Physical review letters 118*, 10 (2017), 106403.

[117] REN, X., RINKE, P., BLUM, V., WIEFERINK, J., TKATCHENKO, A., SANFILIPPO, A., REUTER, K., AND SCHEFFLER, M. Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New Journal of Physics 14*, 5 (May 2012), 053020.

[118] REN, X., RINKE, P., AND SCHEFFLER, M. Exploring the random phase approximation: Application to co adsorbed on cu(111). *Phys. Rev. B 80* (Jul 2009), 045402.

[119] REN, X., RINKE, P., SCUSERIA, G. E., AND SCHEFFLER, M. Renormalized second-order perturbation theory for the electron correlation energy: Concept, implementation, and benchmarks. *Physical Review B 88*, 3 (2013), 035120.

[120] REN, X., TKATCHENKO, A., RINKE, P., AND SCHEFFLER, M. Beyond the random-phase approximation for the electron correlation energy: The importance of single excitations. *Phys. Rev. Lett. 106* (Apr 2011), 153003.

[121] ŘEZÁČ, J., AND HOBZA, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the “Gold Standard,” CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput. 9* (2013), 2151–2155.

[122] ŘEZÁČ, J., AND HOBZA, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev. 116* (2016), 5038.

[123] RILEY, K. E., AND HOBZA, P. Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes. *J. Phys. Chem. A 111* (2007), 8257–8263.

[124] ROCCA, D. Random-phase approximation correlation energies from lanczos chains and an optimal basis set: Theory and applications to the benzene dimer. *J. Chem. Phys. 140*, 18 (2014), 18A501.

[125] ROCCA, D., DIXIT, A., BADAWI, M., LEBÈGUE, S., GOULD, T., AND BUČKO, T. Bridging molecular dynamics and correlated wave-function methods for accurate finite-temperature properties. *Phys. Rev. Materials 3*, 4 (2019), 040801.

[126] ROCCA, D., GEBAUER, R., SAAD, Y., AND BARONI, S. Turbo charging time-dependent density-functional theory with Lanczos chains. *J. Chem. Phys. 128*, 15 (Apr. 2008), 154105–+.

[127] ROCCA, D., LU, D., AND GALLI, G. Ab initio calculations of optical absorption spectra: Solution of the bethe–salpeter equation within density matrix perturbation theory. *J. Chem. Phys. 133*, 16 (Oct. 2010), 164109.

[128] ROCCA, D., VOROS, M., GALI, A., AND GALLI, G. Ab initio optoelectronic properties of silicon nanoparticles: Excitation energies, sum rules, and tamm–dancoff approximation. *Journal of Chemical Theory and Computation 10*, 8 (2014), 3290–3298.

[129] ROŚCISZEWSKI, K., PAULUS, B., FULDE, P., AND STOLL, H. Ab initio calculation of ground-state properties of rare-gas crystals. *Phys. Rev. B 60*, 11 (1999), 7905.

[130] RUPP, M. Machine learning for quantum mechanics in a nutshell. *Int. J. Quantum Chem. 115*, 16 (2015), 1058–1073.

[131] SAAD, Y. *Iterative Methods for Sparse Linear Systems*, second ed. SIAM, Philadelphia, 2003, p. 185 *ff.*

[132] SATO, T. Long-range corrected density functional study on weakly bound systems: Balanced descriptions of various types of molecular interactions. *J. Chem. Phys. 126* (2007), 234114.

[133] SCHIMKA, L., HARL, J., STROPPA, A., GRÜNEIS, A., MARSMAN, M., MITTENDORFER, F., AND KRESSE, G. Accurate surface and adsorption energies from many-body perturbation theory. *Nature Mater. 9* (Sept. 2010), 741–744.

[134] SCHLIPF, M., AND GYGI, F. Optimization algorithm for the generation of oncv pseudopotentials. *Computer Physics Communications 196* (2015), 36–44.

[135] SCHWABE, T., AND GRIMME, S. Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. *Phys. Chem. Chem. Phys. 9* (2007), 3397–3406.

[136] SCHWARTZ, C. Estimating Convergence Rates of Variational Calculations. In *Methods in Computational Physics*, B. Alder, S. Fernbach, and M. Rotenberg, Eds., vol. 2. Academic Press, 1963, p. 241.

[137] SHEPHERD, J. J., GRÜNEIS, A., BOOTH, G. H., KRESSE, G., AND ALAVI, A. Convergence of many-body wave-function expansions using a plane-wave basis: From homogeneous electron gas to solid state systems. *Phys. Rev. B 86* (Jul 2012), 035111.

[138] SPEYBROECK, V. V., HEMELSOET, K., JOOS, L., WAROQUIER, M., BELL, R. G., AND CATLOW, C. R. A. Advances in theory and their application within the field of zeolite chemistry. *Chem. Soc. Rev. 44*, 20 (Oct. 2015), 7044–7111.

[139] STAROVEROV, V. N., SCUSERIA, G. E., TAO, J., AND PERDEW, J. P. Tests of a ladder of density functionals for bulk solids and surfaces. *Physical Review B 69*, 7 (2004), 075102.

[140] SUN, J., RUZSINSZKY, A., AND PERDEW, J. P. Strongly constrained and appropriately normed semilocal density functional. *Physical review letters 115*, 3 (2015), 036402.

[141] SZABO, A., AND OSTLUND, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory.* Courier Corporation, 2012.

[142] TAKATANI, T., HOHENSTEIN, E. G., MALAGOLI, M., MARSHALL, M. S., AND SHERRILL, C. D. Basis set consistent revision of the S22 test set of noncovalent interaction energies. *J. Chem. Phys. 132*, 14 (Apr. 2010), 144104.

[143] TAO, J., PERDEW, J. P., STAROVEROV, V. N., AND SCUSERIA, G. E. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Physical Review Letters 91*, 14 (2003), 146401.

[144] TOULOUSE, J., GERBER, I. C., JANSEN, G., SAVIN, A., AND ÁNGYÁN, J. G. Adiabatic-connection fluctuation-dissipation density-functional theory based on range separation. *Phys. Rev. Lett. 102* (Mar 2009), 096404.

[145] TOULOUSE, J., REBOLINI, E., GOULD, T., DOBSON, J. F., SEAL, P., AND ANGYÁN, J. G. Assessment of range-separated time-dependent density-functional theory for calculating c6 dispersion coefficients. *The Journal of chemical physics 138*, 19 (2013), 194106.

[146] UMARI, P., STENUIT, G., AND BARONI, S. Gw quasiparticle spectra from occupied states only. *Phys. Rev. B 81*, 11 (Mar 2010), 115104.

[147] WALKER, B., SAITTA, A. M., GEBAUER, R., AND BARONI, S. Efficient approach to time-dependent density-functional perturbation theory for optical spectroscopy. *Phys. Rev. Lett. 96*, 11 (2006), 113001.

[148] WERNER, H.-J., KNOWLES, P. J., KNIZIA, G., MANBY, F. R., AND SCHÜTZ, M. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput. Mol. Sci. 2* (2012), 242.

[149] WERNER, H.-J., KNOWLES, P. J., KNIZIA, G., MANBY, F. R., SCHÜTZ, M., AND (AND OTHERS). *MOLPRO*, version 2012.1, a package of ab initio programs, 2012. see http://www.molpro.net.

[150] WILSON, H. F., GYGI, F., AND GALLI, G. Efficient iterative method for calculations of dielectric matrices. *Phys. Rev. B 78*, 11 (Sept. 2008), 113303.

[151] WILSON, H. F., LU, D., GYGI, F., AND GALLI, G. Iterative calculations of dielectric eigenvalue spectra. *Phys. Rev. B 79* (2009), 245106.

[152] ZHANG, L., HAN, J., WANG, H., CAR, R., AND WEINAN, E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett. 120*, 14 (2018), 143001.

[153] ZHU, W., TOULOUSE, J., SAVIN, A., AND ÁNGYÁN, J. G. Range-separated density-functional theory with random phase approximation applied to noncovalent intermolecular interactions. *J. Chem. Phys. 132*, 24 (June 2010), 244108.