UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Systematic study with the GW approach of the ionization energy and electronic affinity of small molecules

Relatore	Laureando
Prof. Paolo Umari	Simone Cigagna

Anno Accademico 2021/2022

Abstract

This thesis aims to verify the capabilities of a new algorithm for computing energetic levels. The algorithm is based on the many body perturbation theory through the *GW* approximation. This allows for ab-initio simulations through a two-step process. First, a calculation based on density functional theory is performed, followed by a *GW* elaboration. These methods are available in the software package Quantum ESPRESSO.

The student will apply the *GW* method to a series of small molecules known as *GW*100, for which the HOMO energies (which correspond to the vertical ionization energies) computed through various codes and approximations are reported.

First of all, the student will learn the basic theory of the DFT and *GW* methods and he will learn to use the codes in the Quantum ESPRESSO package. The student will then obtain the parameters required for the calculations in order to guarantee the convergence of the results.

Then, the student will interface with the *GW* code through AiiDA, which allows the automatic execution of a computation, including the gathering of the data. This will lead to the development of python code.

This will allow the student to obtain the energy levels for the reference set *GW*100.

The student will also have the opportunity to assemble and use a small cluster based on Raspberry Pi 4B+, to assess the scaling capabilities of the new method.

Contents

Chapter 1

Introduction

1.1 The basics of ab initio simulations

Understanding the electronic structure of a system is a necessary step to consistently predict it's electrical, optical and magnetic properties. Some of the most ambitious and promising approaches for this task are ab-initio calculations, which attempt to predict the behavior of matter solely on the basis of quantum mechanics considerations, without the introduction of additional models or parameters. In practice, this requires solving the many body Schrödinger equation, which, with the exception of a few simple problems such as for Hydrogen-like atoms, necessitates the aid of modern computers. However, given the complexity of the many body terms, even the most powerful calculators are often unable to solve the problem without the aid of meaningful approximations.

Since the second half of the nineteenth century, numerous improvements in this regard have been made. In particular, density functional theory (DFT) [1, 2] is now able to accurately predict the energy values for the ground state of small molecules in a few seconds, even while running on portable devices. Section 1.1.1 introduces the theory and approximations on which density functional theory is based.

However, DFT fails to compute the energy of exited states, which is required to determine some fundamental material properties such as electronic band gaps and optical responses. For this purpose, more advanced techniques have been developed, such as the GW method [3] and the related G_0W_0 approximation [4], which are explained in Section 1.1.2. The downside of these new methods is their much higher computational demand, and in recent years numerous steps forward have been made to ease this problem. An example aimed at exploiting parallelism of modern supercomputers is briefly shown in Section 1.1.2, while Section 3.2 tests its implementation [5] in the QUANTUM ESPRESSO [6] package on a small cluster of Raspberry Pis.

Finally, Section 4.1 compares the result obtained with QUANTUM ESPRESSO with other well established implementations of the G_0W_0 approximation, by analysing the HOMO energies of the 100 molecules included in the $GW100$ database [7].

1.1.1 Density Functional Theory

The Hohenberg-Kohn Functional

The starting point for all first principle calculations is the many body Hamiltonian for a system of N electrons and M nuclei. Using Hatree atomic units, the Hamiltonian can be written as

$$\hat{H} = \underbrace{-\frac{1}{2}\sum_i \nabla_i^2}_{\hat{T}_e} \underbrace{-\sum_{i,I} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}}_{\hat{V}_{\text{ext}}} + \underbrace{\sum_{i,j>i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}_{\text{int}}} \overbrace{-\sum_I \frac{1}{2M_I}\nabla_I^2}^{\hat{T}_N} + \overbrace{\sum_{I,J>I} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}}^{\hat{V}_{II}} \tag{1.1}$$

where the lower case subscripts refer to the electrons, while the upper case subscripts denote the nuclei of charge Z_I and mass M_I. The first term $\hat{T}_e$ is the kinetic energy operator for the electrons, $\hat{V}_{\text{ext}}$ is the interaction between the electrons in the field of the nuclei and $\hat{V}_{\text{int}}$ represents the interaction between the electrons. Finally, the last two terms consist in the kinetic energy $\hat{T}_N$ of the nuclei and the Coulomb interaction between the nuclei $\hat{V}_{II}$. Since atoms are much heavier than individual electrons the problem can be split into two parts: first, the nuclear kinetic energy $\hat{T}_N$ is neglected and the Schrödinger equation is solved while considering the nuclei positions $\mathbf{R}_I$ as parameters. This is known as the Born-Oppenheimer or adiabatic approximation [8], and allows the interaction between the nuclei to be treated as a classical addictive term. The nuclei part of the problem can be solved at a later time by using the previously computed electronic eigenvalues. With this approximation the many-body time-independent Schrödinger equation is reduced to the following expression:

$$\left(\hat{T}_e + \hat{V}_{\text{int}} + \hat{V}_{\text{ext}}\right)\psi(\{\mathbf{r}_i\}) = E\psi(\{\mathbf{r}_i\}). \tag{1.2}$$

Solving Equation 1.2 for $\psi(\{\mathbf{r}_i\})$ is still an impossible task, mainly because the many body term $\hat{V}_{\text{int}}$ requires the simultaneous knowledge of all the single electron wave functions ψ_i. Density functional theory approaches the problem by considering as fundamental variable the ground state density

$$n(\{\mathbf{r}_i\}) = \sum_{i=1}^{N} |\psi_i(\{\mathbf{r}_i\})| \tag{1.3}$$

rather than the wavefunction $\psi(\{\mathbf{r}_i\})$. Indeed, it can be proven that for any system of interacting particles in an external potential V_{ext}, the external potential is uniquely determined[1] by the ground state particle density $n_0(\mathbf{r})$ [9]. This is the first Hohenberg-Kohn theorem and a direct consequence is that, since the hamiltonian is fully determined, the wavefunctions for all states are determined. The second Hohenberg-Kohn theorem states that a universal functional for the energy can be defined in terms of the density $n(\{\mathbf{r}_i\})$, valid for any external potential V_{ext}. The exact ground state energy of the system is the global minimum value of this functional and the density $n(\mathbf{r})$ that minimizes this functional is the exact ground state density $n_0(\mathbf{r})$. The many body problem can thus be reformulated as the minimization of a single functional of the density

$$E_{\text{HK}}[n] = T[n] + E_{\text{int}}[n] + \int d^3\mathbf{r} V_{\text{ext}}(\mathbf{r})n(\mathbf{r}). \tag{1.4}$$

The Kohn-Sham Functional

However, no exact functional is known for any system of more than one electron, and the practical usefulness of density functional theory derives instead from the ansatz made by Kohn and Sham in 1965 [2]. Their idea was to replace the original interacting system with an easier to solve independent particle system, with the same ground state density. All of the difficult many body terms are then incorporated into an exchange-correlation functional E_{xc}, and the accuracy of the solution of the Kohn-Sham equations depends only on the accuracy of the approximation used for the exchange correlation functional. The Hohenberg-Kohn functional can then be rewritten in the form

$$E_{\text{KS}} = T_s[n] + \int d^3\mathbf{r} V_{\text{ext}}(\mathbf{r})n(\mathbf{r}) + E_{\text{H}}[n] + E_{\text{xc}}[n] \tag{1.5}$$

where the first term is the kinetic energy functional of the non interacting system with density $n(\mathbf{r})$ and E_{H} represents the well defined Hartree energy

$$E_{\text{H}}[n(\mathbf{r})] = \frac{1}{2}\int \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2. \tag{1.6}$$

[1] except for a constant

The minimization of the functional showed in Equation 1.5 leads to the Kohn-Sham Schrödinger-like single particle self-consistent equations

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + \underbrace{\frac{\delta E_{\text{H}}}{\delta n(\mathbf{r})}}_{V_{\text{H}}} + \underbrace{\frac{\delta E_{\text{xc}}}{\delta n(\mathbf{r})}}_{V_{\text{xc}}} \right) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \tag{1.7}$$

Solving Equation 1.7 numerically requires the expansion of the wave function in a basis set (such as plane waves). This would require a summation over an infinite number of states, which is obviously not practical. For this reason, it is convenient to truncate the sum to include only solutions with an energy inferior to an arbitrary value. Choosing a small energy cutoff can lead to non converged results, so caution must be taken when preparing a DFT calculation. To further simplify computations, it is useful to observe that many properties of materials are not particularly dependent on the more tightly bound core electrons, but rather they are determined by valence electrons. As a consequence, the Coulomb potential of the nuclei and the effects of the core electrons can be replaced by an effective potential (called pseudopotential) that can be computed once and reused in future calculations.

1.1.2 Many Body Perturbation Theory

The *GW* approach can be considered as a post-processing of a DFT calculation. This implies that the accuracy of the final result will depend on the choices for the exchange and correlation functional used in the DFT as well as the basis used for expanding the wavefunction.

Green's Function

The starting point of the *GW* approach is the Green's Function. In general, a Green's Function is defined as a solution of a linear differential equation with a Dirac delta inhomogeneous function. In quantum mechanics, the differential linear equation we are interested in is the Schrödinger equation

$$\left[i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\nabla^2 \right] \Psi(\mathbf{r}, t) = V(\mathbf{r}, t)\Psi(\mathbf{r}, t) \tag{1.8}$$

and a Green's Function is a solution of the inhomogeneous problem

$$\left[i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\nabla^2 \right] G(\mathbf{r}, t; \mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'). \tag{1.9}$$

A solution $\Psi(\mathbf{r}, t)$ of the original problem (Equation 1.8) can be written in terms of the solution $G(\mathbf{r}, t; \mathbf{r}', t')$ that satisfies Equation 1.9

$$\Psi(\mathbf{r}, t; \mathbf{r}', t') = \int G(\mathbf{r}, t; \mathbf{r}', \mathbf{t}')\Psi(\mathbf{r}', t')d^3\mathbf{r}'. \tag{1.10}$$

By writing the wavefunction in terms of the time evolution operator $U(t, t') = e^{-\frac{i}{\hbar}H(t-t')}$ and using the closure relation we get

$$\Psi(\mathbf{r}, t) = \int \langle \mathbf{r}|e^{-\frac{i}{\hbar}H(t-t')}|\mathbf{r}'\rangle \langle \mathbf{r}'|\Psi(t')\rangle \, d^3\mathbf{r}'. \tag{1.11}$$

Comparing Equation 1.10 with Equation 1.11 reveals that the Green's function can be seen as the probability amplitude of finding a particle at position $\mathbf{r}'$ at time t' given that it started at $\mathbf{r}$ at time t. In other words, the Green's function is a propagator.

Switching to the second quantization formalism, the single particle electron Green's function is defined as

$$G_{ij}(t, t') = -i\langle T[c_i(t)c_j^\dagger(t')]\rangle \tag{1.12}$$

where T is the time-ordering operator

$$T[c_i(t)c_j^\dagger(t')] = \theta(t - t')c_i(t)c_j^\dagger(t') - \theta(t' - t)c_j^\dagger(t')c_i(t) \tag{1.13}$$

and c_i and $c_j^\dagger$ are the creation and annihilation operators.

It can be proven that the frequency representation of the one particle green's function can be written as

$$G(\mathbf{r}_1, \mathbf{r}_2; \omega) = \sum_i \frac{A^i(\mathbf{r}_1)A^{i*}(\mathbf{r}_2)}{E^{N+1,i} - E^{N,0} - \omega - i\eta_i} - \sum_j \frac{B^j(\mathbf{r}_1)B^{j*}(\mathbf{r}_2)}{E^{N-1,i} - E^{N,0} + \omega - i\eta_j} \tag{1.14}$$

in which $E^{N,i}$ is the energy of the system with N electrons in its i-th quantum state and A^i, B^i are the so called quasi-particle amplitudes. Equation 1.14 clearly shows that the Green's function exhibits peaks when the frequency ω equals the quasi-particle energies $\mathcal{E} = E^{N+1,i} - E^{N,0}$ and $\mathcal{E}' = E^{N-1,j} - E^{N,0}$. Finally, the terms η_i and η_j represent the inverse life-time of the i and j transitions.

The G_0W_0 approximation

To ease the computation of the Green's function it is common to start from the one-particle Green's function of a non interacting system

$$G_0(\mathbf{r_1}, \mathbf{r_2}; \omega) = \sum_i \frac{\phi_i(\mathbf{r_1})\phi_i^*(\mathbf{r_2})}{\omega - \epsilon_i - i\eta} - \sum_j \frac{\phi_j(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)}{\omega - \epsilon_j + i\eta} \tag{1.15}$$

where $\phi_i(\mathbf{r})$, $\phi_j(\mathbf{r})$ are the valence and conduction Kohn-Sham orbitals and ϵ_i, ϵ_j are their respective energies. This new Green's function is related to the interactive problem through Dyson's equation

$$G(\mathbf{r_1}, \mathbf{r_2}; \omega) = G_0(\mathbf{r}_1, \mathbf{r}_2; \omega) + \int d\mathbf{r}' d\mathbf{r}'' G_0(\mathbf{r_1}, \mathbf{r}'; \omega)\Sigma(\mathbf{r}', \mathbf{r}''; \omega)G(\mathbf{r}'', \mathbf{r}_2; \omega) \tag{1.16}$$

where the self-energy operator Σ can be approximated as a frequency convolution of the Green's function G with the screened coulomb interaction W

$$\Sigma(\mathbf{r_1}, \mathbf{r_2}; \omega) \sim \frac{i}{2\pi} \int d\omega' e^{i\omega'\eta} G(\mathbf{r_1}, \mathbf{r_2}; \omega + \omega')W(\mathbf{r}_1, \mathbf{r_2}; \omega') \tag{1.17}$$

with η representing a positive infinitesimal. The screened coulomb interaction W can be written as a set of self-consistent integrals of G^0, G and the coulomb interaction v. These equations form the so called GW method. However, its self-consistent nature combined with the large non-diagonal matrices required for working with W necessitate the introduction of further approximations. A common technique consists of entirely replacing G with the non interactive green's function G_0, which greatly reduces the computational burden, while relying on the choice for the exchange and correlation functional used in the DFT calculation. The new screened coulomb interaction is referred as W_0 and gives the name G_0W_0 to the approximation. In addition, the quasi-amplitudes can be approximated with the Kohn and Sham orbitals which allow to work with diagonal matrices. Finally, the quasi-particle energies can be obtained with the following self-consistent one-variable equation

$$\mathcal{E}_i = \epsilon_i - \langle\phi_i|V_{xc}|\phi_i\rangle + \langle\phi_i|\Sigma(\mathcal{E})|\phi_i\rangle \tag{1.18}$$

Exploiting parallelisation

It is useful to split Σ_{xc} into the correlation part Σ_c and the exchange one Σ_x and to define the function

$$S_i(\mathbf{r}; \omega) = \sum_{\omega'} \int d\mathbf{r}' \psi_i^*(\mathbf{r}')G_r(\mathbf{r}'; \omega - \omega')W_r(\mathbf{r}'; \omega) \tag{1.19}$$

where the notation $G_\mathbf{r}(\mathbf{r}'; \omega) = \langle\mathbf{r}'|G_0(\omega)|\mathbf{r}\rangle$ and $W_\mathbf{r}(\mathbf{r}'; \omega) = \langle\mathbf{r}'|W_0(\omega)|\mathbf{r}\rangle$ is used for a generic point $\mathbf{r}$ in the simulation cell. The expectation value of Σ_c can then be written as

$$\langle\psi_i|\Sigma_c(\omega)|\psi_i\rangle = \int d\mathbf{r} S_i(\mathbf{r}; \omega)\psi_i(\mathbf{r}) = \frac{1}{N_{\mathrm{r,tot}}} \sum_{\alpha=1,N_{\mathrm{r,tot}}} S_i(\mathbf{r}_\alpha; \omega)\psi_i(\mathbf{r}_\alpha) \tag{1.20}$$

where the last member is the discretization of the integral on a grid of $N_{r,\alpha}$ equally spaced points, labelled as $\mathbf{r}_\alpha$. The computation can be simplified by observing that points for which $\psi(\mathbf{r}_\alpha) \sim 0$ can be neglected. Thus, only points that satisfy the condition $|\psi_i(\mathbf{r}_\alpha)| > s$ are considered, with an opportune threshold s. In addition, a coarser grid can be taken by choosing only one point every n. The total number of grid points is therefore reduced to an integer $N_{n,s}$ which depends on the chosen parameters. The convergence of the expected value for the self energy operator $\langle\psi_i|\Sigma_c(\omega)|\psi_i\rangle_{n,s}$ can be eased by weighting each grid point according to $\psi_i(\mathbf{r}_\alpha)$

$$\langle\psi_i|\Sigma_c(\omega)|\psi_i\rangle_{n,s} = \frac{\sum_{\alpha=1}^{N_{n,s}} S_i(\mathbf{r}_\alpha;\omega)\psi_i(\mathbf{r}_\alpha)}{\sum_{\alpha'=1}^{N_{n,s}} \psi_i^*(\mathbf{r}_{\alpha'})\psi_i(\mathbf{r}_{\alpha'})}. \tag{1.21}$$

Convergence can be achieved with just $N_{n,s} \sim 100$. Moreover, each point can be computed independently from the other, allowing the algorithm to scale linearly with the number of nodes in a cluster.

Chapter 2

Preparing an ab-initio calculation

2.1 Quantum ESPRESSO

Quantum ESPRESSO is a collection of open source software for electronic structure calculations. One of the most important package included is `PWscf`, which performs various density functional theory calculations using a plane waves basis set and pseudopotentials. In particular, the `pw.x` executable requires an input file where parameters such as the simulation cell size must be set, and stores the charge density and Kohn-Sham orbitals in the binary files `charge-density.dat` and `wfcN.dat`. Additional information regarding the calculation is stored in the xml file `data-file-schema.xml`. Excitations can then be studied using the `GWL` package: first, `pw4gww.x` performs the preparations for the G_0W_0 approximation starting from a previously computed DFT calculation. Then, the `gww.x` code computes the expectation values of the correlation part of the self-energy operator which are used to calculate the quasi particle energies through analytical continuation. A development version of Quantum ESPRESSO implements the G_0W_0 approximation using the technique described in Section 1.1.2 to exploit parallelisation and requires a python script called `easyanalyser.py` to merge each point of the grid in a single file that can be read from `gww.x`.

2.2 AiiDa

"AiiDA is an open-source Python infrastructure to help researchers with automating, managing, persisting, sharing and reproducing the complex workflows associated with modern computational science and all associated data"[1]. In practice, AiiDa allows to easily create complex and modular workflows for multiple local and remote computers, while automatically recording inputs, outputs and metadata in a database [10, 11]. To interface with external codes such as Quantum ESPRESSO, AiiDa requires the installation of additional plugins, which inform AiiDa on how to actually prepare the required input files, run the code with the correct parameters and parse the output.

Calculations

The rules for creating an input file, as well as the definition of the necessary parameters for executing the code are specified in a Calculation. In particular, a Calculation is defined as a *process that creates new data*, and can be implemented either as function (called calculation function) or as a class derived from the base-class `CalcJob`, included in the AiiDa package. The latter is the most common implementation thanks to its increased flexibility. A `CalcJob` must override the `define` class-method of the parent-class and list the expected inputs and outputs of the calculation. In addition, a `CalcJob` requires the `prepare_for_submission` method, where the calculation inputs are transformed in one or more input files and AiiDA is informed on which files to retrieve after the calculation is finished.

[1] `www.aiida.net` Automated Interactive Infrastructure and Database for Computational Science

Parsers

The output data of a calculation can then be parsed by an optional class, derived from the `Parser` base-class, where basic information can be extracted from the calculation. Usually, a parser at least detects the exit status of job as well as the running time.

Workflows

A workflow is defined as a process that calls other workflows and calculations. Simple workflows can be implemented in a function through the decorator `@workfunction` but a much more powerful implementation can be created with a class which inherits from the `WorkChain` base-class. These type of workflows are known as `WorkChains` and must override the `define` class-method, which lists the required inputs, outputs and, most importantly, the outline of the calculation. This is a list of methods that must be implemented inside the `WorkChain` class and each of them launches one or more `CalcJob`s and `WorkChain`s, using the output of a previous calculation as part of their inputs when needed.

aiida-quantumespresso

The QUANTUM ESPRESSO team provides an official AiiDa plugin[2] for some of the codes included in the package. For instance, the `quantumespresso.pw` CalcFunction allows to easily start a `pw.x` calculation and, since everything is done in python, the results can be accessed and analysed with ease. The plugin provides additional tools to further automate the execution of `pw.x`, such as auto-restarting `WorkChains`, custom `DataTypes` to work with pseudopotentials and command line utilities. However, the official plugin does not support the new version of `pw4gww.x` which exploits parallelisation. For this reason, I created a fork[3] of the plugin which adds a `CalcJob` and a `Parser` for launching the `pw4gww.x` code, a `CalcJob` which wraps `easyanalyser.py` and a `WorkChain` to run `pw.x`, `pw4gww.x` and `easyanalyser.py` one after another. The fork also provides a special `WorkChain` designed for running `pw4gww.x` on N computers: each machine only computes $1/N$ of the total required points and the results are merged in a single file.

[2] `https://github.com/aiidateam/aiida-quantumespresso`
[3] available at `https://github.com/simonecig/aiida-quantumespresso`

Chapter 3

Running an ab-initio calculation

3.1 GW100

The $GW100$ set is a collection of ionization potentials and electron affinities of 100 different molecules, computed using various GW techniques. It allows, for example, to quantitatively compare different basis set, the effect of pseudopotentials and the handling of unoccupied states. With the goal of comparing the new G_0W_0 approach with more established methods, I computed the HOMO energies for each of the 100 molecules with the new version of `pw4gww.x`. Using AiiDa's workflows, running the same calculation multiple times while changing only some parameters is as easy as inserting the code for running a `WorkChain` once inside a loop: at each step, AiiDa will automatically create the input files, start the calculation and store everything in an internal database.

3.2 Calculation Parameters

After some initial testing with various pseudopotential families, I chose the PseudoDojo [12, 13] set of norm conserving PBE pseudopotential, as they appeared to provide slightly better results overall. As anticipated in Section 1.1.1, running a `pw.x` calculation requires setting a cutoff value for the wavefunciton. PseudoDojo provides a recommended value for each element of the periodic table and I assumed the cutoff for a given molecule to match the highest value of the constituent atoms (usually around 82 Ry).

With the goal of reducing computation times and memory requirements, the simulation cell was restricted to 10.5 Å. However, with such a small cell, all energies must be corrected with a proper offset to account for boundary conditions, as I will explain in Section 3.2.1.

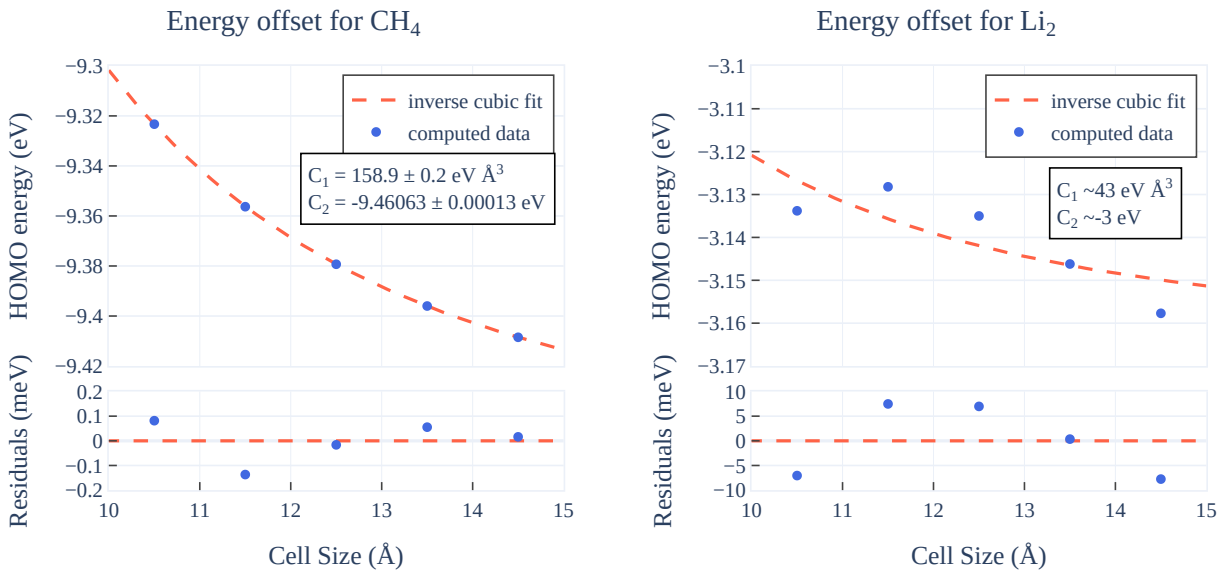
A `pw4gww.x` calculation is mainly characterized by the parameters introduced in Section 1.1.2: the grid spacing n and the threshold s. For each molecule I set the former to the arbitrary value of $n = 8$ and I wrote a simple script to "bruteforce" the corresponding value for s such that the total number of grid points is $90 \leq N_{n,s} \leq 110$, which should lead to converged results.

3.2.1 Energy offset

To find the energy value far away from the molecule, I run multiple `pw.x` calculations while increasing the cell size of 1 Å, from $a_{min} = 10.5$ Å to $a_{max} = 14.5$ Å, and recorded the resulting ionization energies. Since the energy is expected to depend on the inverse of the volume of the molecule, the data was fitted with the formula $y(a) = C_1 a^{-3} + C_2$ and the offset

$$q = C_2 - y(10.5) \tag{3.1}$$

was used to correct the energies obtained through G_0W_0. Different molecules obviously resulted in different correction values. For instance, Figure 3.1a shows the fit for CH_4, which resulted in an offset



(a) Successful energy offset fit for CH_4.

(b) Failed energy offset fit for Li_2

Figure 3.1

of $q = -0.14\,\text{eV}$. For 17 out of the 100 molecules, however, the produced energies significantly deviate from the inverse cubic formula, probably due the small cell which prevents the DFT calculation from reaching full convergence. An example is shown in Figure 3.1b.

Overall, this is the limiting factor for the accuracy of the results which will be shown in Section 4.1. Indeed, different choices for the cell size often lead to significantly different fit parameters. The only way to completely eliminate this problem is to use a much bigger cell size, resulting in drastic increases in computation times and memory requirements. With the hardware at disposal, however, this was not possible.

3.3 Hardware

Depending on the hardware and the parameters of the calculation, a G_0W_0 simulation can take many hours to finish. To speed things up, the job was split into two machines: a laptop and a cluster of four Raspberry Pi 4B+ connected to a LAN switch. Some technical details on the hardware used are presented in Table 3.1. Using AiiDa's command line, a `WorkChain` can be run on any remote or local computer, provided that an SSH connection can be established and the required code is installed on the target machine.

Machine	CPU	Base Clock	Overclock	Ram Capacity
Huawei Matebook D	AMD Ryzen 5 2500U	2.0 GHz	-	8 GB
Raspberry Pi 4B+ (x4)	Cortex A72	1.5 GHz	2.0 GHz	8 GB

Table 3.1: Brief overview on the hardware used.

Performance and price Comparisons

The cluster is particularly important since it shows the advantages of parallel computing. In fact, a single Raspberry requires 26 minutes and 33 seconds to compute a single point of a `pw4gww.x` calculation for CH_4, while a cluster of 4 Pis can complete 4 points in the same time.

Still, the laptop performs much better as it is able to complete the same calculation in just 2 minutes and 40 seconds. Even taking into account price, this particular cluster is sub-optimal. Figure 3.2 shows

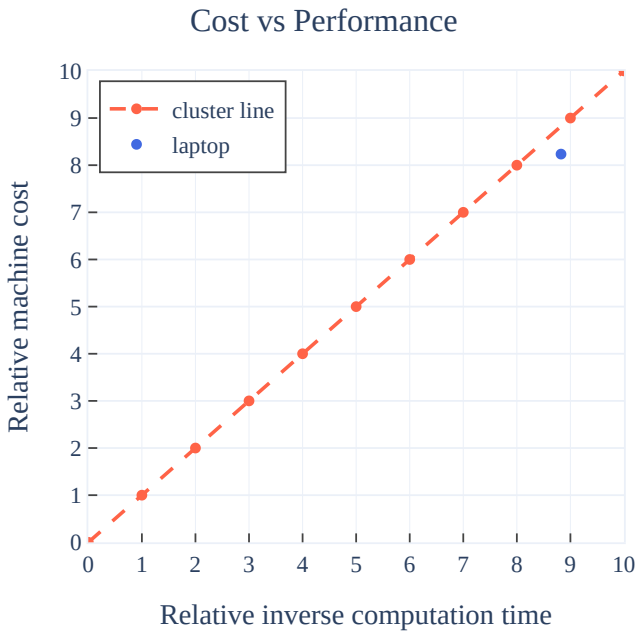


Figure 3.2: Cost of a machine relative to the price of a single Raspberry Pi 4B+ (85 €) versus the inverse of the computation time relative to the value for a single Raspberry ($1/26.5\,\mathrm{min}^{-1}$). In addition to the point (blue) corresponding to the laptop used in this thesis, the Figure also shows the predicted line for a cluster (red) made of up to 10 nodes, assuming a linear scaling.

the price versus the inverse of the computation times for a single point of a `pw4gww.x` calculation, with both quantities expressed with respect to the values of a single Raspberry. Thus, a machine that appears below the cluster line is able to extract more performance per money than a cluster of Raspberry Pi. This is the case for the laptop that I used, which costs 8.2 times as much as a Raspberry Pi 4B+, but is almost 9 times as powerful.

Finally, even considering the fact that ARM processors are more power efficient, the laptop still comes ahead: in my testing a single Raspberry can draw up to 8.9 W, while the laptop consumes 63.4 W under full load. Therefore, the 9 Pis required to match the laptop performance would consume $\sim 10\%$ more power.

Chapter 4

Calculation results

4.1 Comparison with the *GW*100 database

The *GW*100 set consists of many ionization potentials obtained with different techniques. Here, I focus on a subset which includes only energies obtained with algorithms similar to Quantum ESPRESSO (i.e. PBE G_0W_0). It is important to observe that, for a given molecule, even energies from this subset can differ significantly, with an average standard deviation of 0.2 eV. My results for the ionization energies of the 83 molecules for which the energy offset could be computed, are presented in Appendix A.1, together with the average values of the PBE subset. In addition, Table A.1 also reports the CCSD(T)[1] data-set [14], which is a highly accurate yet computational expensive Quantum Chemistry approach, as well as the experimental ionization energies [15] as a reference. The agreement of my results with the PBE average strongly depends on the molecule. For instance, the absolute errors for BF, CO and Cs_2 are all significantly smaller than 0.02 eV, while many fluorine compounds such as FH, F_2 and LiF have absolute errors of 1.0 eV. This is probably caused by the strongly localized $3d$ and $2p$ orbitals of fluorine, which are particularly difficult to treat with plane wave approaches. Overall, the average absolute difference with respect to the PBE subset is 0.4 eV, which is still significantly higher than the average standard deviation for the subset.

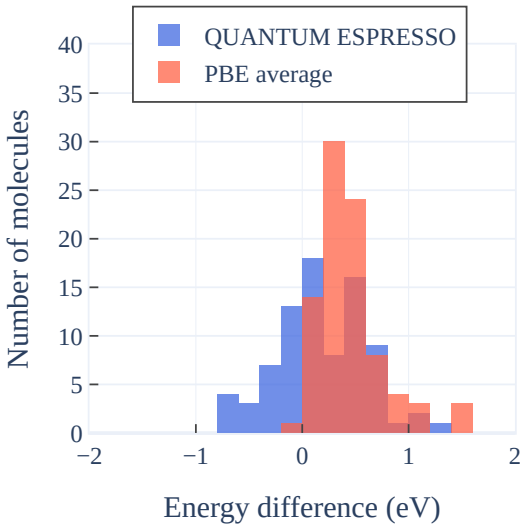


Figure 4.1: Histograms of the differences of the Quantum ESPRESSO (blue) and PBE average (red) sets, with respect to the CCSD(T) values. The overlapping area of the two histograms is portrayed in a darker shade of red.

Surprisingly, the agreement of my data with respect to the CCSD(T) data-set is substantially better: while the mean absolute difference (MAD) is 0.3 eV, the mean difference (MD) is only 0.16 eV. For reference, both the MAD and MD between the PBE and CCSD(T) sets are 0.4 eV, as shown in Table 4.1a. This suggests that the new `pw4gww.x` might have more in common with the coupled cluster approach than other PBE methods, given the smaller systematic error. This is likely due to the fact that other codes expand the screened coulomb interaction W within a basis set which appears to be too limited. Figure 4.1 sums up the previous considerations: while Quantum ESPRESSO's histogram spans a wider area than the PBE set, the latter is limited to positive values since CCSD(T) energies are systematically lower.

Comparisons with the experimental values provide a similar picture: looking at either Table 4.1b or Figure 4.2, it is clear that CCSD(T) is the most compatible set with the experimental values and

[1] Coupled Cluster with Single, Double, and an estimate of Triple substitutions

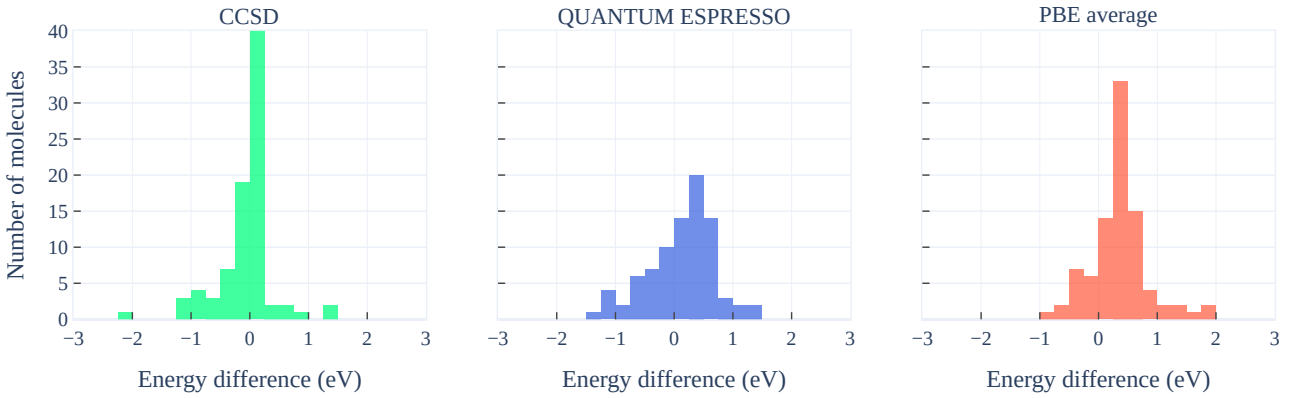


Figure 4.2: Histograms of the ionization energies differences for the CCSD(T) (left), Quantum ESPRESSO (middle), and PBE average (right) sets, with respect to the experimental values.

Comparison with CCSD(T)		
	MD (eV)	MAD (eV)
QE	0.16	0.3
PBE	0.4	0.4

(a) Mean difference (MD) and mean absolute difference (MAD) for Quantum ESPRESSO and PBE average, with respect to the CCSD(T) set.

Comparison with Experiments		
	MD (eV)	MAD (eV)
CCSD(T)	0.06	0.3
QE	0.10	0.5
PBE	0.4	0.5

(b) Mean difference (MD) and mean absolute difference (MAD) for Quantum ESPRESSO, PBE average and CCSD(T), with respect to the experimental values.

Table 4.1

that the PBE set is affected by a systematic error. `pw4gww.x`, ranks between the two, with a mean difference comparable to CCSD(T) and an absolute mean difference identical to the PBE set.

4.2 Conclusions

In this thesis, after a small overview on the theoretical basis on density functional theory and many body perturbation theory, I have tested the the new version of Quantum ESPRESSO on a small cluster. However, as shown in Section 3.3, to fully exploit its parallelisation capabilities, a cluster's node must be more powerful than a Raspberry Pi 4B+, which in my testing did not provide a meaningful advantage over an average laptop. In Section 3.2 I have introduced the main parameters used for the calculations, with a focus on the energy offset needed to compensate for the small simulation cell. The technique explained in Section 3.2.1 proved to be somewhat inconsistent, but it was the only option with the hardware at disposal. Overall, the results I have shown in Section 4.1 highlight the distinction between the new version of `pw4gww.x` and other PBE G_0W_0 methods. Indeed, the latter uses a limited basis for W, leading to slightly worse comparisons with the the coupled cluster set, as well as the experimental values.

I believe that the accuracy of the results could be increased with a larger simulation cell, which should lower the random error that derives from the energy fit, leading to histograms with higher peaks and lower tails than the one shown in Figure 4.2. Consequently, to confidently assess the accuracy of the new version of Quantum ESPRESSO, further tests with more powerful hardware are required. It is worth noting that using the new AiiDa's plugin I have developed for this thesis, repeating all calculations on a super computer with a different cell size is extremely simple: after exchanging the SSH keys needed to establish a safe connection to the remote machine, the user is only required to change the variable related to the cell size, while AiiDa will take care of creating the input files, launching `pw.x`, `pw4gww.x` and `easyanalyser.py` as well as retrieving the results.

Appendix A

Numerical Results

A.1 Results Table

Molecule	QUANTUM ESPRESSO (eV)	PBE average (eV)	CCSD (eV)	Experimental (eV)
C_8H_{10}	-8.66	-8.59 ± 0.02	-8.85	-8.77
O_3	-12.10	-11.69 ± 0.08	-12.55	-12.73
C_4H_{10}	-11.79	-11.46 ± 0.02	-11.57	-11.09
C_7H_8	-8.30	-8.65 ± 0.02	-8.90	-8.82
C_6H_5OH	-8.27	-8.39 ± 0.03	-8.70	-8.75
C_5H_5N	-9.75	-9.12 ± 0.03	-9.66	-9.66
C_4	-11.09	-10.91 ± 0.07	-11.26	-12.54
P_2	-10.57	-10.29 ± 0.03	-10.47	-10.62
Ag_2	-7.31	-7.67 ± 0.08	-7.49	-7.66
CO_2	-13.54	-13.31 ± 0.04	-13.71	-13.77
BeO	-10.25	-9.14 ± 0.11	-9.94	-10.10
MgO	-8.25	-6.94 ± 0.10	-7.49	-8.76
BH_3	-13.20	-12.88 ± 0.04	-13.28	-12.03
H_2	-16.43	-15.83 ± 0.03	-16.40	-15.43
BF	-10.60	-10.62 ± 0.04	-11.09	-11.00
Si_2H_6	-10.48	-10.37 ± 0.03	-10.64	-10.53
OCSe	-11.10	-10.25 ± 0.03	-10.78	-10.37
GaCl	-10.17	-9.75 ± 0.06	-9.77	-10.07
PN	-11.48	-11.24 ± 0.05	-11.74	-11.88
B_2H_6	-12.14	-11.87 ± 0.02	-12.26	-11.90
As_2	-10.44	-9.50 ± 0.03	-9.78	-10.00
H_2NNH_2	-10.18	-9.29 ± 0.03	-9.72	-8.98
Na_4	-4.79	-4.17 ± 0.01	-4.22	-4.27
OCS	-11.09	-10.98 ± 0.04	-11.17	-11.19
H_2CO	-11.24	-10.40 ± 0.04	-10.84	-10.88
CI_4	-8.59	-8.87 ± 0.07	-9.27	-9.10
C_5H_6	-8.25	-8.39 ± 0.02	-8.68	-8.53
CBr_4	-9.77	-10.00 ± 0.05	-10.46	-10.54
CCl_4	-11.56	-11.11 ± 0.05	-11.56	-11.69
CH_4N_2O	-10.13	-9.39 ± 0.04	-10.05	-9.80
C_2H_3Br	-8.78	-9.08 ± 0.06	-9.27	-9.90
C_2H_3I	-8.84	-9.20 ± 0.06	-9.33	-9.35
$C_4H_{10}O$	-10.20	-9.36 ± 0.03	-9.82	-9.61
$C_6H_5NH_2$	-7.38	-7.69 ± 0.03	-7.99	-8.05
C_8H_8	-8.28	-8.1 ± 0.03	-8.35	-8.43
CO	-13.54	-13.64 ± 0.03	-14.21	-14.01
C_2H_6O	-11.32	-10.20 ± 0.03	-10.68	-10.64
CH_2O_2	-10.86	-10.81 ± 0.04	-11.42	-11.50
$C_5H_6N_2O_2$	-8.63	-8.76 ± 0.03	-9.08	-9.20
$C_4H_4N_2O_2$	-9.36	-9.27 ± 0.04	-10.12	-9.68

CH_4O	-11.23	-10.60 ± 0.03	-11.04	-10.96
$C_4H_5N_3O$	-8.27	-8.35 ± 0.04	-9.51	-8.94
C_6H_6	-8.70	-9.02 ± 0.02	-9.29	-9.23
$C_5H_5N_5O$	-7.72	-8.03 ± 0.03	-8.33	-8.48
CH_4	-14.42	-13.94 ± 0.02	-14.37	-13.60
C_2H_6	-12.87	-12.38 ± 0.02	-13.04	-11.99
C_2H_4	-10.25	-10.36 ± 0.02	-10.67	-10.68
C_2H_2	-10.97	-11.07 ± 0.02	-11.42	-11.49
NCH	-13.65	-13.27 ± 0.03	-13.87	-13.61
C_3H_8	-12.13	-11.80 ± 0.02	-12.05	-11.51
Kr	-13.66	-13.65 ± 0.05	-13.94	-14.00
Ne	-20.15	-19.8 ± 0.4	-21.32	-21.56
Ar	-15.65	-15.24 ± 0.04	-15.54	-15.76
Xe	-11.59	-12.31 ± 0.10	nan	-12.13
SO_2	-12.46	-11.91 ± 0.04	-13.49	-12.50
C_2H_3Cl	-9.89	-9.83 ± 0.03	-10.09	-10.20
C_2H_3F	-10.27	-10.25 ± 0.03	-10.55	-10.63
C_2H_4O	-9.63	-9.62 ± 0.04	-10.21	-10.24
CS_2	-9.84	-9.85 ± 0.04	-9.98	-10.09
C_3H_6	-10.81	-10.59 ± 0.02	-10.86	-10.54
CF_4	-16.26	-15.45 ± 0.04	-16.30	-16.20
I_2	-8.95	-9.49 ± 0.09	-9.51	-9.36
ClH	-12.60	-12.35 ± 0.04	-12.59	-12.79
NaCl	-8.97	-8.35 ± 0.09	-9.03	-9.80
FH	-16.33	-15.36 ± 0.04	-16.03	-16.12
NH_3	-10.50	-10.34 ± 0.03	-10.81	-10.82
HOOH	-11.63	-11.04 ± 0.04	-11.59	-11.70
Br_2	-10.11	-10.29 ± 0.05	-10.54	-10.51
N_2	-15.08	-14.95 ± 0.03	-15.57	-15.58
H_2O	-12.50	-12.00 ± 0.04	-12.56	-12.62
BrK	-7.74	-7.53 ± 0.07	-8.13	-8.82
F_2	-15.99	-15.02 ± 0.04	-15.71	-15.70
Cl_2	-11.49	-11.21 ± 0.04	-11.41	-11.49
GeH_4	-12.42	-12.07 ± 0.04	-12.50	-11.34
HN_3	-10.24	-10.44 ± 0.03	-10.68	-10.72
SH_2	-10.40	-10.11 ± 0.04	-10.31	-10.50
SF_4	-11.80	-12.17 ± 0.04	-12.59	-11.69
TiF_4	-14.60	-13.97 ± 0.05	-15.48	-13.30
AlI_3	-9.09	-9.53 ± 0.08	-9.82	-9.66
AsH_3	-11.00	-10.19 ± 0.03	-10.40	-10.58
$MgCl_2$	-11.44	-11.15 ± 0.05	-11.66	-11.80
LiF	-11.21	-10.16 ± 0.09	-11.32	-11.30
PH_3	-10.70	-10.33 ± 0.03	-10.52	-10.59

Table A.1: Ionization potentials calculated with QUANTUM ESPRESSO, compared with the mean value for all PBE G_0W_0 sets in $GW100$, the CCSD(T) set and the experimental values.

Bibliography

[1] Richard M Martin. *Electronic structure: basic theory and practical methods*. Cambridge university press, 2020.

[2] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[3] Lars Hedin. New method for calculating the one-particle green's function with application to the electron-gas problem. *Phys. Rev.*, 139:A796–A823, Aug 1965.

[4] Mark S. Hybertsen and Steven G. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B*, 34:5390–5413, Oct 1986.

[5] Paolo Umari. A fully linear response g 0 w 0 method that scales linearly up to tens of thousands of cores. *The Journal of Physical Chemistry A*, 2022.

[6] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, sep 2009.

[7] Michiel J. van Setten, Fabio Caruso, Sahar Sharifzadeh, Xinguo Ren, Matthias Scheffler, Fang Liu, Johannes Lischner, Lin Lin, Jack R. Deslippe, Steven G. Louie, Chao Yang, Florian Weigend, Jeffrey B. Neaton, Ferdinand Evers, and Patrick Rinke. Gw100: Benchmarking g0w0 for molecular systems. *Journal of Chemical Theory and Computation*, 11(12):5665–5687, 2015. PMID: 26642984.

[8] Max Born and W Heisenberg. Zur quantentheorie der molekeln. In *Original Scientific Papers Wissenschaftliche Originalarbeiten*, pages 216–246. Springer, 1985.

[9] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[10] Martin Uhrin, Sebastiaan P. Huber, Jusong Yu, Nicola Marzari, and Giovanni Pizzi. Workflows in aiida: Engineering a high-throughput, event-based engine for robust and modular computational workflows. *Computational Materials Science*, 187:110086, 2021.

[11] Sebastiaan P. Huber, Spyros Zoupanos, Martin Uhrin, Leopold Talirz, Leonid Kahle, Rico Häuselmann, Dominik Gresch, Tiziano Müller, Aliaksandr V. Yakutovich, Casper W. Andersen, Francisco F. Ramirez, Carl S. Adorf, Fernando Gargiulo, Snehal Kumbhar, Elsa Passaro, Conrad Johnston, Andrius Merkys, Andrea Cepellotti, Nicolas Mounet, Nicola Marzari, Boris Kozinsky, and Giovanni Pizzi. Aiida 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Scientific Data*, 7(1):300, Sep 2020.

[12] D. R. Hamann. Optimized norm-conserving vanderbilt pseudopotentials. *Phys. Rev. B*, 88:085117, Aug 2013.

[13] M.J. van Setten, M. Giantomassi, E. Bousquet, M.J. Verstraete, D.R. Hamann, X. Gonze, and G.-M. Rignanese. The pseudodojo: Training and grading a 85 element optimized norm-conserving pseudopotential table. *Computer Physics Communications*, 226:39–54, 2018.

[14] Katharina Krause, Michael E. Harding, and Wim Klopper. Coupled-cluster reference values for the gw27 and gw100 test sets for the assessment of gw methods. *Molecular Physics*, 113(13-14):1952–1960, 2015.

[15] SG Lias, JE Bartmess, JF Liebman, JL Holmes, RD Levin, and WG Mallard. Nist chemistry webbook, nist standard reference database number 69. *National Institute of Standards and Technology, Gaithersburg MD2003*, 2003.