UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

Dipartimento di Scienze Fisiche, Informatiche e Matematiche

DOTTORATO DI RICERCA IN PHYSICS AND NANOSCIENCES

Ciclo XXXVI

AUTOMATING CORE-LEVEL SPECTROSCOPY CALCULATIONS: TOWARDS HIGH-THROUGHPUT AND MACHINE LEARNING APPLICATIONS

Relatrici:

Prof. Elisa MOLINARI

Dott. Deborah PREZZI

Candidato:

Michael Alejandro HERNANDEZ BERTRAN

Coordinatore del Corso di Dottorato:

Prof. Marco AFFRONTE

A.A. 2022/2023

This Thesis was co-supervised by Prof. Elisa Molinari, from the Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, and Dr. Deborah Prezzi, from the National Research Council, Nanoscience Institute, Modena, Italy.

Questa Tesi ha come co-relatrici la Prof. Elisa Molinari, dell'Università di Modena e Reggio Emilia, Dipartimento di Scienze Fisiche, Informatiche e Matematiche (FIM), e la Dott. Deborah Prezzi, del Consiglio Nazionale delle Ricerche, Istituto Nanoscienze (CNR-NANO), Modena, Italia.

Abstract

Computational spectroscopy has become a vital tool in materials research, enabling both qualitative and quantitative insights into material properties in close connection with experiments. In this Thesis we present advancements in the automation of ab initio core-level spectroscopy calculations, with a focus on integrating high-throughput computing and machine learning methods to enhance the analysis of experimental spectra. By leveraging first principles simulations, this work explores the electronic and structural characterization of a prototype material, hydrogenated graphene, providing insight into its electronic and structural properties. Our study also demonstrates the use of X-ray Raman scattering (XRS) for analyzing silicon nanoparticle anodes, revealing key components of the solid-electrolyte interphase (SEI) and correlating these with observed capacity loss in lithium-ion batteries. In addition, we develop a robust, modular workflow for ab initio high-throughput X-ray Photoelectron Spectroscopy (XPS) simulations. Furthermore, machine learning models were trained to predict XPS spectra with ab initio accuracy, showcasing their potential to reduce the computational workload required in amorphous-like systems.

Sommario

La spettroscopia computazionale è diventata uno strumento importante nella ricerca sui materiali e permette di migliorare la comprensione sia qualitativa che quantitativa delle loro proprietà, in stretta connessione con gli studi sperimentali. In questa Tesi presentiamo alcuni sviluppi nel calcolo *ab initio* delle spettroscopie degli stati di core, con particolare attenzione alla loro automazione e all'integrazione di metodi di calcolo ad alto rendimento (high-throughput computing, HTC) e di apprendimento automatico per migliorare l'analisi degli spettri sperimentali. Utilizzando simulazioni da principi primi, questo lavoro esplora la caratterizzazione elettronica e strutturale di un materiale prototipo, il grafene idrogenato, offrendo una comprensione più profonda delle sue proprietà elettroniche e strutturali. Il nostro lavoro dimostra inoltre l'utilizzo della diffusione Raman di raggi X (X-ray Raman scattering, XRS) per l'analisi di anodi basati su nanoparticelle di silicio, rivelando componenti chiave dell'interfaccia solido-elettrolita e correlando questi risultati con la perdita di capacità osservata nelle batterie agli ioni di litio. Inoltre, sviluppiamo un *workflow* modulare e robusto per simulazioni HTC ab initio di spettroscopia fotoelettronica a raggi X (X-ray Photoelectron Spectroscopy, XPS). Infine, abbiamo addestrato modelli di apprendimento automatico per prevedere spettri XPS con accuratezza *ab initio*, e dimostrato la loro capacità di ridurre il carico computazionale consentendo la trattazione anche di sistemi disordinati o amorfi.

Contents

List of Figures

List of Tables

Introduction

Computational spectroscopy has become an essential tool for interpreting experimental spectra, providing both qualitative and quantitative insights into material properties. The growing synergy between experimental and theoretical approaches has created a positive feedback loop, driving significant advancements in material science. This is especially evident in the field of core-level (X-ray) spectroscopies, where high-precision ab initio calculations have become indispensable for understanding electronic and structural properties. The rise of modern X-ray facilities, such as third-generation synchrotrons and free-electron lasers, has further expanded the scope of these techniques, making core-level spectroscopies more prevalent and impactful.

Traditional ab initio methods, while highly accurate, often come with significant computational costs, especially for complex systems such as amorphous or disordered materials. This challenge is being addressed by emerging machine learning (ML) models, which offer ab initio accuracy at inexpensive computational cost as an alternative to the traditional first principles approaches. These ML techniques, whether through forward or reverse mapping, are becoming part of the standard toolkit to complement, and in some cases replace, ab initio methods by either predicting spectra directly or inferring material structures from experimental data.

In this context, high-throughput computing (HTC) has gained prominence, allowing for the large-scale generation of data needed for comprehensive materials analysis. The integration of automation infrastructures has streamlined the process, making it feasible to perform extensive simulations that were once computationally prohibitive. As computational capabilities continue to expand, the potential for ML-driven approaches in core-level spectroscopy grows, promising to enhance the efficiency and accuracy of spectral predictions and offering new avenues for both experimentalists and theorists in the study of material properties.

This work explores the automation of core-level spectroscopy calculations, highlighting the synergy between high-throughput methods and machine learning applications. By combining the precision of ab initio techniques with the computational efficiency of ML models, we aim to push the boundaries of spectroscopic analysis and open new possibilities for high-throughput material characterization. Building upon the principles of ab initio simulations for X-ray spectroscopies, the

objective of this thesis is to develop an automated tool that encodes this knowledge, enabling the efficient generation of large datasets necessary for training surrogate models with ab initio accuracy and for the interpretation of experimental results.

The Thesis is organized as follows: Chapter 1 contains the theoretical framework used in this study, summarizing the Density Functional Theory and the machine learning methods commonly applied for the prediction and interpretation of ab initio simulation data. Chapter 2 describes the core-level spectroscopy simulations in the DFT framework. In particular we will cover the basic principles of the X-ray Photoelectron Spectroscopy (XPS), X-ray Absorption Spectroscopy (XAS) and X-ray Raman scattering (XRS), along with the corresponding technical details involved in the ab initio simulations. Chapter 3 demonstrate the application of ab initio XPS simulations for probing electronic properties of hydrogenated free-standing graphene. Chapter 4 describes the application XRS in the study of Si nanoparticle anodes by complementing the experimental measurements with first principles simulations. Chapter 5 is dedicated to describe the development of an automated workflow for high-throughput XPS simulations within the AiiDA framework. Finally, in Chapter 6 we will discuss the machine learning applications to core-level spectroscopies. In particular we will focus in the development a comprehensive approach for predicting XPS spectra with ab initio acuracy, integrating the automated AiiDA workflow into the training procedure.

Chapter 1

Theoretical framework

1.1 The ab-initio approach

In quantum mechanics, a system comprising N atomic nuclei and N_e electrons presents a complex scenario where electrons interact with each other as well as the nuclei of the system. The properties and dynamics are then described by the non-relativistic Schrödinger equation,

$$\hat{H}\Psi(\mathbf{R},\mathbf{r},t) = i\hbar\frac{\partial\Psi(\mathbf{R},\mathbf{r},t)}{\partial t}, \tag{1.1}$$

where $\hat{H}$ is the Hamiltonian of the system and $\Psi(\mathbf{R},\mathbf{r},t)$ is the many-body wave function characterizing the quantum state of N nuclei and N_e electrons with coordinates $\mathbf{R} = \{\boldsymbol{R}_I S_I\}_{I=1}^{N}$ and $\mathbf{r} = \{\boldsymbol{r}_i s_i\}_{i=1}^{N_e}$ respectively. Here $\boldsymbol{R}_I$, $\boldsymbol{r}_i$ and S_I, s_i are the spatial and spin coordinates. If $\hat{H}$ does not explicitly depend on time, the states satisfying Equation 1.1 are stationary with a well-defined energy E and can be expressed as $\Psi(\mathbf{R},\mathbf{r},t) = \Psi(\mathbf{R},\mathbf{r})e^{-i\frac{E}{\hbar}t}$, leading to the time-independent Schrödinger equation:

$$\hat{H}\Psi(\mathbf{R},\mathbf{r}) = E\Psi(\mathbf{R},\mathbf{r}), \tag{1.2}$$

with eigenstates $\Psi(\mathbf{R},\mathbf{r})$ and eigenvalues E. Such Hamiltonian is composed by the kinetic energy and potential terms as follows:

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{ee} + \hat{V}_{eN} + \hat{V}_{NN}, \tag{1.3}$$

with,

$$\hat{T}_e = \sum_{i=1}^{N_e}\frac{|\hat{\mathbf{p}}_i|^2}{2m_e} = \sum_{i=1}^{N_e}\frac{\hat{p}_i^2}{2m_e}, \tag{1.4}$$

$$\hat{T}_N = \sum_{I=1}^{N} \frac{|\hat{\mathbf{p}}_I|^2}{2M_I} = \sum_{I=1}^{N} \frac{\hat{p}_I^2}{2M_I}, \tag{1.5}$$

$$\hat{V}_{ee} = \frac{1}{2} \sum_{i \neq j} \frac{q_e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}, \tag{1.6}$$

$$\hat{V}_{eN} = -\frac{1}{2} \sum_{I=1}^{N} \sum_{i=1}^{N_e} \frac{q_e^2 Z_I}{|\boldsymbol{R}_I - \boldsymbol{r}_i|}, \tag{1.7}$$

$$\hat{V}_{NN} = \frac{1}{2} \sum_{I \neq J} \frac{q_e^2 Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|}, \tag{1.8}$$

being $\hat{T}_e$, $\hat{T}_N$ the kinetic energy operators for the electrons and nuclei respectively, while $\hat{V}_{ee}$, $\hat{V}_{eN}$ and $\hat{V}_{NN}$ are the elecron-electron, electron-nuclei and nuclei-nuclei Coulomb interaction potentials in the stated order. The corresponding mass, charge and momentum for the electrons (ions) are given by m_e (M_I), $e = -|q_e|\sqrt{4\pi\epsilon_0}$ (eZ_I) and $\hat{\mathbf{p}}_i$ ($\hat{\mathbf{p}}_I$), with ϵ_0 vacuum permittivity and $\hat{\mathbf{p}}_i = -i\hbar\boldsymbol{\nabla}_i$.

Solving Equation 1.2 becomes increasingly challenging as the number of particles grows. In such cases the Born–Oppenheimer or adiabatic approximation [13] can be applied, leveraging the mass difference –approximately a factor 1837– between nucleons and electrons. This approximation allows for the separation of electronic and ionic degrees of freedom by considering fixed nuclei positions, thus treating them as parameters and, as a consequence, neglecting $\hat{T}_N$ in Equation 1.2. The adiabatic evolution of the electronic states with the nuclear displacement is supported by the typical timescales of electronic (10^{-15} to 10^{-18}s) and atomic motion (10^{-12} to 10^{-15}s) [14], confirming its wide applicability in the regime of electronic transition energies larger than nuclear motion energies. Under this constraint the electronic many-body problem becomes

$$\hat{H}_e \psi_{\{\mathbf{R}\}}(\mathbf{r}) = E_{\{\mathbf{R}\}} \psi_{\{\mathbf{R}\}}(\mathbf{r}), \tag{1.9}$$

where $\psi_{\{\mathbf{R}\}}(\mathbf{r})$, $E_{\{\mathbf{R}\}}$ and $\hat{H}_e$ represent the electronic many-body state, its energy and the Hamiltonian of the N_e interacting electrons, respectively, corresponding to the ionic configuration $\mathbf{R}$. The electronic Hamiltonian $\hat{H}_e$ now has the form

$$\hat{H}_e = \hat{T} + \hat{V}_{int} + \hat{V}_{ext} + E_{NN}, \tag{1.10}$$

with $\hat{T} = \hat{T}_e$, $\hat{V}_{int} = \hat{V}_{ee}$, $\hat{V}_{ext} = \hat{V}_{eN}$[1] and E_{NN} is a constant contribution to the total energy of the system for a given ionic configuration $\mathbf{R}$.

The dimensionality of the electronic problem remains so large that solving it can be remarkably tough. Within mean field approximations like Hartree (without particle exchange effects) [15] and Hartree-Fock [16], this problem is simplified

[1] Other external potentials can be included in the $\hat{V}_{ext}$ term

by considering non-interacting electrons within an effective potential, but neglecting electron correlation. On the other hand, post-Hartree-Fock methods such as Configuration Interaction [17] (CI), Coupled Cluster [18, 19] (CC), and Quantum Monte Carlo [20] (QMC) aim to include correlation effects but are computationally expensive and often impractical for large systems. In this scenario, Density Functional Theory (DFT) stands out as it avoids explicit many-body wave functions while providing a fair compromise between accuracy, versatility and computational complexity.

1.2 Density Functional Theory

The origins of DFT can be traced back to the seminal works of Thomas [21] and Fermi [22], where the electronic density $n(\boldsymbol{r})$ was proposed as the central quantity,

$$n(\boldsymbol{r}) = \langle \psi \left| \hat{n}(\boldsymbol{r}) \right| \psi \rangle \,, \tag{1.11}$$

where $|\psi\rangle$ is the normalized electronic many-body state and $\hat{n}(\boldsymbol{r}) = \sum_{i=1}^{N_e} \delta(\boldsymbol{r} - \boldsymbol{r}_i)$ is the density operator. Although this approximation did not succeed in accurately describing electrons in matter, it proposed a promising alternative to reduce the degrees of freedom in the problem described by Equation 1.9 to just three.

1.2.1 Hohenberg-Kohn theorems

The core principle of modern DFT is that any property of a system with many interacting particles is a functional of the ground-state electron density, as established by the Hohenberg-Kohn [23] theorems.

Theorem I: For any system of interacting particles in an external potential $V_{ext}(\boldsymbol{r})$, the ground-state particle density $n_0(\boldsymbol{r})$ uniquely determines $V_{ext}(\boldsymbol{r})$, up to an additive constant.

The first theorem validates n_0 as the main quantity in the many particle system. Since the Hamiltonian is uniquely determined by the ground-state density, the wave function of any state can, in principle, be found by solving the Schrödinger equation with this Hamiltonian.

Theorem II: For any external potential $V_{ext}(\boldsymbol{r})$, a functional for the energy $E^{HK}[n(\boldsymbol{r})]$ can be defined, such as its variational global minimum occurs at $n_0(\boldsymbol{r})$ and coincides with the ground-state energy of the system E_0:

$$E_0 = \min_{n(\boldsymbol{r})} E^{HK}[n(\boldsymbol{r})] = E^{HK}[n_0(\boldsymbol{r})] \tag{1.12}$$

The second theorem states that the exact ground-state energy and density are completely determined by $E^{HK}[n(\boldsymbol{r})]$

$$E^{HK}[n(\boldsymbol{r})] = F^{HK}[n(\boldsymbol{r})] + \int d\boldsymbol{r} V_{ext}(\boldsymbol{r}) n(\boldsymbol{r}) + E_{NN}, \tag{1.13}$$

where

$$F^{HK}[n(\boldsymbol{r})] = \langle \psi | \hat{T} | \psi \rangle + \langle \psi | \hat{V}_{int} | \psi \rangle = T[n(\boldsymbol{r})] + V_{int}[n(\boldsymbol{r})], \tag{1.14}$$

is a universal functional (the same for all electron systems), since it includes all internal kinetic and potential energies of the interacting electrons which are functionals of $n(\boldsymbol{r})$ by definition.

Although the original proofs of the Hohenberg-Kohn theorems are restricted to the space of "V_{ext}-representable" densities, which are ground-state densities of a Hamiltonian with some external potential V_{ext}, this limitation is handled by an alternative approach known as the constrained search formulation by Levy [24] and Lieb [25]. At this stage, we must address the challenge of having an exact theory for the density and energy of the interacting many-body system that depends on an unknown functional.

1.2.2 Kohn-Sham formulation

The practical application of the Hohenberg-Kohn theorems is limited by the fact that $F^{HK}[n(\boldsymbol{r})]$ is unknown. In this context, the Kohn-Sham [26] formulation offers a viable solution by transforming the many-body problem of interacting electrons into a problem of non-interacting electrons in an effective potential. The construction of such auxiliary system is based on the hypothesis that the effective potential $V^{KS}(\boldsymbol{r})$ is determined by the same ground-state density as the original interacting system. Under this assumption the Schrödinger equation (in Hartree atomic units) for the auxiliary system is

$$\left(-\frac{1}{2}\nabla^2 + V^{KS}(\boldsymbol{r}) \right) \phi_i^{KS}(\mathbf{r}) = \varepsilon_i^{KS} \phi_i^{KS}(\mathbf{r}), \tag{1.15}$$

$$\hat{H}^{KS} \phi_i^{KS}(\mathbf{r}) = \varepsilon_i^{KS} \phi_i^{KS}(\mathbf{r}), \tag{1.16}$$

where $\hat{H}^{KS}$, $\phi_i^{KS}(\mathbf{r})$ and ε_i^{KS} are the single electron Kohn-Sham Hamiltonian, states and energies, respectively. The density of this independent electron system, assumed equal to the interacting electron system one, can be written in the Kohn-Sham formulation as

$$n(\boldsymbol{r}) = \sum_{i=1}^{N_e} \left| \phi_i^{KS}(\mathbf{r}) \right|^2, \tag{1.17}$$

while the kinetic energy is defined in the following terms,

$$T^{KS}[n(\boldsymbol{r})] = -\frac{1}{2}\langle\phi_i^{KS}(\mathbf{r})|\nabla^2|\phi_i^{KS}(\mathbf{r})\rangle. \tag{1.18}$$

The universal functional $F^{HK}[n(\boldsymbol{r})]$ in Equation 1.14 can be reformulated as

$$F^{KS}[n(\boldsymbol{r})] = T^{KS}[n(\boldsymbol{r})] + V_{int}[n(\boldsymbol{r})] + \left(T[n(\boldsymbol{r})] - T^{KS}[n(\boldsymbol{r})]\right), \tag{1.19}$$

which can be expanded further by taking the Hartree part $E^H[n(\boldsymbol{r})]$ out of $V_{int}[n(\boldsymbol{r})]$

$$F^{KS}[n(\boldsymbol{r})] = T^{KS}[n(\boldsymbol{r})] + E^H[n(\boldsymbol{r})] + E_{xc}[n(\boldsymbol{r})], \tag{1.20}$$

with

$$E^H[n(\boldsymbol{r})] = \frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r}'\frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}, \tag{1.21}$$

and

$$E_{xc}[n(\boldsymbol{r})] = T[n(\boldsymbol{r})] - T^{KS}[n(\boldsymbol{r})] + V_{int}[n(\boldsymbol{r})] - E^H[n(\boldsymbol{r})]. \tag{1.22}$$

The term $E_{xc}[n(\boldsymbol{r})]$ is the exchange–correlation energy functional, which accounts for the difference between the kinetic and internal interaction energies of the many-body system and those of the auxiliary independent-particle system, where electron-electron interactions are replaced by the Hartree energy. The Equation 1.13 can be expressed in the Kohn-Sham terms as

$$E^{KS}[n(\boldsymbol{r})] = T^{KS}[n(\boldsymbol{r})] + E^H[n(\boldsymbol{r})] + E_{xc}[n(\boldsymbol{r})] + \int d\boldsymbol{r}V_{ext}(\boldsymbol{r})n(\boldsymbol{r}) + E_{NN}, \tag{1.23}$$

and its variational minimum condition for fixed particle number gives the effective potential $V^{KS}(\boldsymbol{r})$

$$V^{KS}(\boldsymbol{r}) = V_{ext}(\boldsymbol{r}) + \frac{\delta E^H[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} + \frac{\delta E_{xc}[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} = V_{ext}(\boldsymbol{r}) + V^H(\boldsymbol{r}) + V_{xc}(\boldsymbol{r}). \tag{1.24}$$

The Equations (1.15), (1.17) and (1.24), known as Kohn–Sham equations, have the form of independent-particle equations with a potential that must be found self-consistently with the resulting density.

1.2.3 Exchange–correlation functionals

The accuracy of the Kohn-Sham formulation is determined by the choice of the exchange–correlation functional contribution. The simplest approximation, proposed by Kohn and Sham [26] in its original work, is the local density approximation (LDA), which considers $E_{xc}[n(\boldsymbol{r})]$ as the exchange-correlation energy functional of a homogeneous electron gas (HEG)

$$E_{xc}^{HEG}[n(\boldsymbol{r})] = \int d\boldsymbol{r} n(\boldsymbol{r}) \epsilon_{xc}^{HEG}[n(\boldsymbol{r})], \tag{1.25}$$

where $\epsilon_{xc}^{HEG}[n(\boldsymbol{r})]$ is the exchange-correlation energy density for the HEG, containing an exchange part $\epsilon_{x}^{HEG}[n(\boldsymbol{r})] = -\frac{3}{4}(\frac{3}{\pi} n(\boldsymbol{r}))^{\frac{1}{3}}$ and a correlation contribution calculated numerically for intermediate homogeneous densities using QMC [27] and as an extrapolation of analytical low and high density limits [28] for the other density regimes.

In order to account for the inhomogeneous part of the density, a dependence on gradients has been included in the exchange-correlation functional

$$E_{xc}^{GGA}[n(\boldsymbol{r})] = \int d\boldsymbol{r} n(\boldsymbol{r}) \epsilon_{xc}^{HEG}[n(\boldsymbol{r}), \nabla n(\boldsymbol{r})]. \tag{1.26}$$

This approach, named generalized gradient approximation (GGA) [29] usually improves the results on structural and bonding properties. Besides LDA and GGA functionals, we can find other approximations with additional ingredients (meta-GGA, hybrids and generalized RPA), which are organized according to they progression towards achieving chemical accuracy [30, 31].

1.2.4 The plane-wave pseudopotential method

Solving the Kohn-Sham equations relies on numerical methods capable of handle the complexity of $V^{KS}(\boldsymbol{r})$ and the need for self-consistency. The usual approach is to expand the Kohn–Sham eigenstates $\phi_n^{KS}(\boldsymbol{r})$ into a set of basis functions, transforming the Equation 1.15 in an algebraic equation for the expansion coefficients. Plane-waves (PWs) basis sets are a common option due to mathematical simplicity and compatibility with efficient Fast Fourier Transform (FFT) algorithms which improve the scalability with the system size. In order to reduce the computational cost in extended systems, translational symmetry is imposed by constructing a minimal (primitive) cell[2] containing one or more atoms (lattice points); when repeated periodically they can approximate the original system. This procedure is the natural choice for crystals and can be extended to treat non periodic systems –such as clusters, surfaces and defects– by constructing sufficiently large cells (supercells). Under this condition the effective potential $V^{KS}(\boldsymbol{r})$ is periodic

$$V^{KS}(\boldsymbol{r} + \mathbf{T}) = V^{KS}(\boldsymbol{r}), \tag{1.27}$$

and according to the Bloch's theorem [32] each of the Kohn-Sham states can be expressed as a periodic function $u_{\mathbf{k}n}$ modulated by a plane wave with wave vector $\mathbf{k}$ lying in the first Brillouin Zone (BZ):

[2]The primitive cell is defined by the translation vectors $\mathbf{a}_i$, which generates all the possible lattice translations $\mathbf{T} = \sum_{i=1}^{3} n_i \mathbf{a}_i$ with $n_i \in \mathbb{Z}$. The primitive cell in the reciprocal space is called first Brillouin Zone.

$$\phi^{KS}_{\mathbf{k}n}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}n}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{\mathbf{k}n}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \tag{1.28}$$

where Ω, $\mathbf{G}$ and $c_{\mathbf{k}n}(\mathbf{G})$ are the cell volume, the reciprocal lattice vectors and the expansion coefficients in the PWs basis set of $u_{\mathbf{k}n}(\mathbf{r})$ respectively. The Equations (1.27) and (1.28) allow to transform the Kohn-Sham Schrödinger equation into the following matrix eigenvalue problem in reciprocal space

$$\sum_{\mathbf{G}'} H^{KS}_{\mathbf{G},\mathbf{G}'}(\mathbf{k}) c_{\mathbf{k}n}(\mathbf{G}') = \varepsilon_{\mathbf{k}n} c_{\mathbf{k}n}(\mathbf{G}), \tag{1.29}$$

with

$$H^{KS}_{\mathbf{G},\mathbf{G}'}(\mathbf{k}) = \frac{1}{2} |\mathbf{k}+\mathbf{G}|^2 \delta_{\mathbf{G},\mathbf{G}'} + V_{ext}(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}') + V_H(\mathbf{G}-\mathbf{G}') + V_{xc}(\mathbf{G}-\mathbf{G}') \tag{1.30}$$

The number of $\mathbf{G}$ vectors in the PWs representation is infinite; however, for practical calculations we must truncate the expansion at a certain point. This is achieved by introducing an energy threshold, corresponding to the kinetic energy of an electron with momentum $|\mathbf{k}+\mathbf{G}|$

$$\frac{1}{2} |\mathbf{k}+\mathbf{G}|^2 \leq E_{cut}. \tag{1.31}$$

Nevertheless, describing electron wavefunctions using a PWs basis set is highly demanding. Near the nucleus, core electron wavefunctions are nodeless and tightly confined, while valence electron wavefunctions exhibit oscillations due to orthogonality requirements. Thus, a large value of E_{cut} is usually needed in order to describe electron wavefunctions on that spatial region. A first step to solve this issue is to apply the frozen-core approximation and consider only the valence electrons in the calculations. Since core electrons do not participate in chemical bonding, in this approximation they are assumed to occupy the same states as in the isolated atom. A second step consist of replacing the ionic core potential by an effective interaction (pseudopotential) [33, 34, 35] which will lead to nodeless valence wave functions (pseudo wavefunctions) removing the orthogonality wiggles near the nucleus.

When using pseudpotentials we must provide a cutoff radius, r_{cut}, which separates the core and valence regions. In practice, the pseudo wavefunctions are equal to the corresponding all electron beyond the cut-off radius, and smoothed in the core region. Figure 2.1 shows the effects of 'pseudization' on both the wavefunctions and the potential for the isolated C atom in the region $r < r_{cut}$. Wavefunctions and potential where obtained using the `ld1.x` code contained in Quantum ESPRESSO [36, 37, 38]. Due to the frozen-core approximation, the $1s$ state (core) is excluded from the pseudization process, thus reducing the number of electrons in the Kohn-Sham equations involved in the calculations.

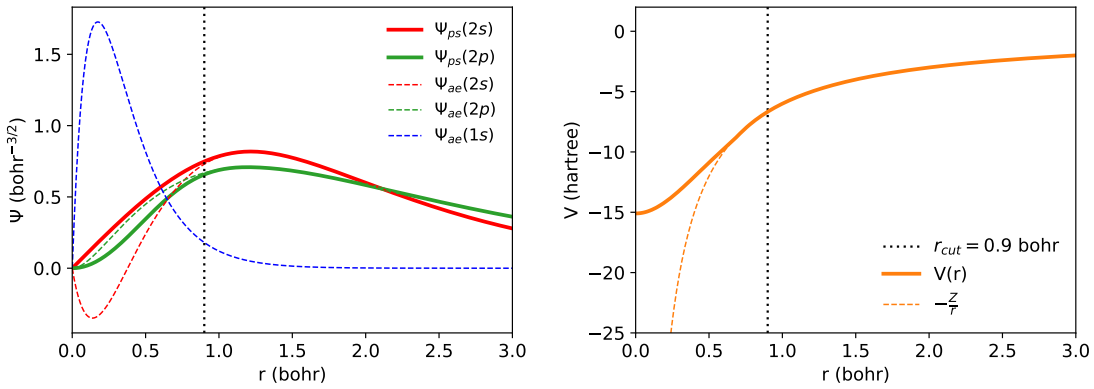

Figure 1.1: Pseudization effects on the wavefunctions (left) and the potential (right) for the isolated C atom in the region $r < r_{cut}$. Dashed (solid) lines represent the all-electron (pseudo) wavefunctions and potential.

Its worth pointing out that all the equations exposed so far are valid in the PW pseudopotential framework for a particular kind of pseudopotentials, called norm-conserving. They are constructed by imposing that the norm of the pseudo wavefunctions matches the norm of the corresponding all-electron wave functions. However, further reduction of the number of plane-wave basis states can be achieved with the same level of accuracy by augmenting the electronic charge density in the core regions using ultrasoft pseudopotentials [39, 40]. This reduction of the problem size transforms the Schrödinger equation into a generalised eigenvalue problem by introducing generalized orthonormality constraints which depend on the ionic positions. As a consequence, additional terms appear in the Kohn-Sham equations from the density-dependent terms in the total energy that should be taken into account in the calculations [41].

1.3 Machine Learning methods in ab-initio simulations

Machine Learning (ML) can be defined as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty [42]. Such data-driven approaches are classified in three categories known as unsupervised, supervised and reinforcement learning. In this work we will focus on the first two categories, by far the most widespread in the context of materials science [43].

1.3.1 Unsupervised learning

In computational materials science, data often involves numerous variables and complex relationships that are challenging to analyze using traditional methods. Unsupervised learning, also known as descriptive learning, provides a robust approach for discovering patterns within unlabeled datasets, without relying on prior knowledge or assumptions. This approach enables an effective exploration of complex material behaviors and relationships. In the following, we will introduce the principal approaches used in the thesis organized in two categories within the unsupervised methods known as dimensionality reduction and clustering.

1.3.1.1 Dimensionality reduction

A dimensionality reduction algorithm is a technique dedicated to reduce the number of features (or dimensions) in a dataset while retaining as much relevant information as possible. This process projects high-dimensional data into a lower-dimensional space, making it easier to visualize, interpret, and analyze.

Principal component analysis (PCA) [44, 45] is a widely used unsupervised technique for dimensionality reduction, feature extraction, and data visualization. This method is particularly beneficial when the original dataset includes correlated and noisy variables. The original space is projected into a new one with lower dimension defined by new coordinates, called principal components (PCs), which are orthogonal vectors formed as linear combinations of the original variables.

PCA finds the set of l orthogonal directions that provide the best approximation of the data matrix $\mathbf{X} \in \mathbb{R}^{N\times d}$ for $l < d$. This minimization process is equivalent to finding the orthogonal directions over which the variance of $\mathbf{X}$ is maximal. In this way, PCA is performed through the diagonalization of the covariance matrix $\mathbf{C}$ defined as

$$\mathbf{C} = \frac{1}{N-1}\mathbf{X}^T\mathbf{X}, \tag{1.32}$$

where the data vectors $\mathbf{X}_i$ are mean-centered. In order to reduce the dimensionality, the eigenvectors of $\mathbf{C}$ are ordered according to their corresponding eigenvalues and the number l of principal components can be chosen on the basis of the cumulative variance expressed. The original data can be projected in terms of PCs as

$$\mathbf{P} = \mathbf{X}\mathbf{V}^l, \tag{1.33}$$

where $\mathbf{V}^l \in \mathbb{R}^{d\times l}$ contains the first l eigenvectors of $\mathbf{C}$ with the larger eigenvalues. The matrix $\mathbf{P}^l \in \mathbb{R}^{N\times l}$ will contain the first PCs of the data matrix ordered according to the amount of variance captured by each of them.

CUR decomposition [46] is a matrix factorization technique utilized for dimensionality reduction either in the feature or data space. The low-rank matrix decomposition is explicitly expressed in terms of the elements of the original matrix. In practice the data matrix $\mathbf{X}$ can be approximated with a lower-rank matrix $\tilde{\mathbf{X}}$ as

$$\mathbf{X} \approx \tilde{\mathbf{X}} = \mathbf{CUR}, \tag{1.34}$$

where $\mathbf{C}$ and $\mathbf{R}$ are constructed from the selected k columns and rows of $\mathbf{X}$ respectively, and $\mathbf{U}$ is a $k \times k$ matrix.

This technique is particularly useful for tasks such as selecting optimal feature vectors or subsets, as it leverages the actual elements of the data matrix. In the case of feature selection (i.e., columns), a score π_c is computed for every column of the initial matrix

$$\pi_c = \sum_{j=1}^{k} (v_c^j)^2, \tag{1.35}$$

where v_c^j is the c-th component of the right singular vector of the single value decomposition (SVD) of $\mathbf{X}$ and k is the number of features that have yet to be selected and runs from k to 1. In the case of sample selection (rows) the scores are computed using the left singular vectors.

Farthest-point sampling [47] (FPS) is a deterministic algorithm that selects the farthest point from the already chosen points in each step. Essentially, FPS aims to select points that are as diverse as possible from the given dataset, promoting uniform sampling across the input space by sorting the data points in descending order according to diversity. The process begins by selecting an initial sample, typically the first point in the dataset $\mathbf{X} \in \mathbb{R}^{N \times d}$ to maintain its deterministic nature. In subsequent iterations, the algorithm calculates the distance between each remaining point and the selected points $\mathbf{X}_{selected}$, then picks the next point $\mathbf{X}_k$ based on the maximum distance as

$$\mathbf{X}_k = \underset{\mathbf{X}_k \in \mathbf{X}}{\operatorname{argmax}} \left(\min_{\mathbf{X}_j \in \mathbf{X}_{selected}} |\mathbf{X}_k - \mathbf{X}_j| \right) \tag{1.36}$$

The selection process continues until $K \leq d$ points are selected.

1.3.1.2 Clustering

Clustering groups data points according to their similarity. The data points are thereby organized into an efficient representation that characterizes the population being sampled. Such methods can be either applied to the original data, or to the output of a dimensionality reduction algorithm.

k-means [48] and k-medoids [49] are two partition-based clustering algorithms very similar. In both the number of clusters K in which the elements of $\mathbf{X} \in \mathbb{R}^{N\times d}$ are distributed is an input parameter. In the particular case of k-means, each data point $\mathbf{X}_i$ is assigned to a unique cluster S_k forming a data partition $S = \{S_1, ..., S_K\}_{K\leq d}$. The objective is to minimize the within-cluster scatter

$$\min_S W_S = \min_S \sum_{i=1}^{K} \sum_{\mathbf{X}_j \in S_i} ||\mathbf{X}_j - \boldsymbol{\mu}_i||^2 , \tag{1.37}$$

where $\boldsymbol{\mu}_i = \frac{1}{N_i}\sum_{j\in S_i} \mathbf{X}_j$ is the mean (centroid) of all the data points assigned to S_i. k-medoids is the generalization of k-means to non Euclidean metrics, replacing the centroid with the closest data point $\mathbf{X}_c$ to the rest of the points within the same cluster, according to the metric $d(\mathbf{X}_j, \mathbf{X}_c)$. This implies computing pairwise distances between all points in the cluster, leading to higher computational costs.

Hierarchical clustering [50] , contrary to flat clustering, offers the possibility of obtain clusters nested inside each other. There are two main approaches to hierarchical clustering: bottom-up or agglomerative clustering, and top-down or divisive clustering. Both methods take as input the dissimilarity between the data points. In the bottom-up approach, the most similar data points are merged at each step. In the top-down approach, they are split into the two most dissimilar clusters. Both approaches are heuristics, which do not optimize any well-defined objective function. Furthermore, they will always produce a clustering of the input data, removing the need to define beforehand the number of clusters. On the other hand, the clustering should be inspected always since data can have no structure at all (e.g., random noise).

1.3.2 Supervised learning

Supervised learning focuses on labeled datasets ($\mathbf{X}$, $\mathbf{y}$), meaning each input $\mathbf{X}_i$ comes with a corresponding output $\mathbf{y}_i$ (label). The tasks under this method can be broadly classified into two categories based on the nature of the label: regression, when $\mathbf{y}$ is continuous, and classification, when $\mathbf{y}$ is discrete. In practice, the main task reduces to find the functional form $f_{\boldsymbol{\omega}}$ (model) with parameters $\boldsymbol{\omega}$, connecting the inputs $\mathbf{X}$ with the outputs $\mathbf{y}$. The accuracy of the predictions $\hat{\mathbf{y}} = f_{\boldsymbol{\omega}}(\mathbf{X})$ is assessed through the so-called loss function $\mathcal{L}_{\boldsymbol{\omega}}$, allowing to find the best model parameters $\hat{\boldsymbol{\omega}}$ as

$$\hat{\boldsymbol{\omega}} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}}\ \mathcal{L}_{\boldsymbol{\omega}}(\mathbf{y}, \hat{\mathbf{y}}), \tag{1.38}$$

In the following, we will introduce the principal regression approaches used in the thesis.

1.3.3 Linear regression

Given a set of inputs $\mathbf{X} \in \mathbb{R}^{N\times d}$ and targets $\mathbf{y} \in \mathbb{R}^N$ we want to determine the linear least-squares estimator in the form

$$\hat{\mathbf{y}} = \mathbf{X}^T\boldsymbol{\omega}^{\mathbf{1}} + \omega^0\mathbf{1}_d, \tag{1.39}$$

where $\boldsymbol{\omega}^{\mathbf{1}}$, ω^0 are the parameters or weights of the model while $\mathbf{1}_d$ has dimensions $d \times 1$ and contains only ones. The values of the optimal parameters depend on the definition of the loss function. In particular, if we select $\mathcal{L}$ as the mean squared error (MSE) then we obtain

$$\hat{\boldsymbol{\omega}}^{\mathbf{1}}, \hat{\omega}^0 = \underset{\boldsymbol{\omega}^{\mathbf{1}},\omega^0}{\operatorname{argmin}} \ (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}). \tag{1.40}$$

Linear models fitted using Equation 1.40 lead to unstable solutions. One of the principal problems is the overfitting, causing a lack of generalization since the model is learning the training data instead of the function connecting the inputs with the outputs. This issue is usually addressed introducing a so-called regularization term acting on the weights $\boldsymbol{\omega}$. This approach, called ridge regression [51], introduces a loss function defined as

$$\mathcal{L}_{ridge} = (\mathbf{y} - \mathbf{X}\boldsymbol{\omega})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\omega}) + \frac{\lambda}{2}\boldsymbol{\omega}^T\boldsymbol{\omega}, \tag{1.41}$$

where the data vectors in $\mathbf{X}$ are mean-centered and $\lambda > 0$ is an hyperparameter of the model, known as regularization strength, penalizing large values of $\boldsymbol{\omega}$. The penalty term discourages model parameters from closely following the training points, introducing a manageable amount of error in the training data. This balancing, known as the bias-variance trade-off, is crucial for preventing overfitting in overparameterized models increasing the generalization capability of the model. In Equation 1.41 the intercept was excluded, as regularizing it will make the solution dependent on the outputs origin. Thus, it is estimated as the average of all the outputs and added to the final prediction. The analytical solution for the minimization of $\mathcal{L}_{ridge}$ is given by

$$\hat{\boldsymbol{\omega}} = (\mathbf{X}^T\mathbf{X} + \frac{\lambda}{2}\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}. \tag{1.42}$$

It is convenient to express the ridge regression problem in terms of inner products of the data vectors. This is usually done using the dual formulation, changing from the primal variables $\boldsymbol{\omega}$ to a new set of dual variables $\boldsymbol{\alpha}$. The dual variables are formally the Lagrange multipliers introduced to handle the constraint $\mathbf{y} = \mathbf{X}\boldsymbol{\omega}$ in the loss minimization problem. The reformulation of the problem can be done by computing $\nabla_{\boldsymbol{\omega},\boldsymbol{\alpha}}[\mathcal{L}_{ridge} - \boldsymbol{\alpha}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\omega})] = 0$, leading to $\hat{\boldsymbol{\omega}} = \mathbf{X}^T\boldsymbol{\alpha}$ with the loss expressed in terms of the dual variables as

$$\mathcal{L}_{\boldsymbol{\alpha}} = (\mathbf{y} - \mathbf{K}\boldsymbol{\alpha})^T(\mathbf{y} - \mathbf{K}\boldsymbol{\alpha}) + \frac{\lambda}{2}\boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha}, \tag{1.43}$$

where $\mathbf{K} = \mathbf{X}\mathbf{X}^T$ is the matrix of pairwise inner products (Gram matrix). The optimal dual and primal variables can be expressed then as

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{K} + \frac{\lambda}{2}\mathbf{I}\right)^{-1}\mathbf{y}, \tag{1.44}$$

$$\hat{\boldsymbol{\omega}} = \mathbf{X}^T\left(\mathbf{K} + \frac{\lambda}{2}\mathbf{I}\right)^{-1}\mathbf{y}, \tag{1.45}$$

allowing to predict the output of a new data point $\mathbf{X}_{N+1}$ as

$$y_{N+1} = \mathbf{X}_{N+1}^T\mathbf{X}^T\left(\mathbf{K} + \frac{\lambda}{2}\mathbf{I}\right)^{-1}\mathbf{y} + \hat{\omega}_0. \tag{1.46}$$

1.3.3.1 Kernels

The formulation of the ridge regression problem in terms of the Gram matrix allows to extend the ridge method to non-linear models. In the ideal case, we can transform the inputs with a non-linear function ϕ such that the new inputs $\phi(\mathbf{X})$ have a linear relation with the outputs $\mathbf{y}$. Under this assumptions, the Gram matrix becomes $\mathbf{K}_{i,j} = k(\mathbf{X}_i, \mathbf{X}_j) = \langle\phi(\mathbf{X}_i), \phi(\mathbf{X}_j)\rangle$, where k is the kernel function and $\langle\rangle$ is the inner product in the space of the new outputs. This definition allows to operate in a high-dimensional feature space without explicitly computing non-linear transformation $\phi(\mathbf{X})$. Instead, it relies on a kernel function to compute the inner product between the data points in the higher-dimensional space, using only the original feature space. This result is known as the kernel trick [52] and is only valid for Mercer or positive definite kernels. In this thesis we will use mainly the Gaussian kernel defined as

$$k(\mathbf{X}_i, \mathbf{X}_j) = \exp\left\{-\frac{||\mathbf{X}_i - \mathbf{X}_j||^2}{2\sigma}\right\}, \tag{1.47}$$

where σ is an hyperparameter of the model known as the length scale. The Gaussian kernel could be isotropic or anisotropic depending if σ is a scalar or a vector.

1.3.4 Descriptors for atomic structures

Up to this point we explored various ML methods commonly employed to address materials science challenges. However, the accuracy and performance of these approaches are conditioned, up to certain extent, by the choice of the data representation (descriptors) and their capacity to extract meaningful information

from raw data. The number of representations available is really large and can be grouped according to the strategy that is used to incorporate symmetry in their construction [53], here we will focus on the atom density-based descriptors, in particular SOAP [54].

1.3.4.1 Atom-density representations

Within the atom-density representations the atom density field ρ at the position $\mathbf{x}$ can be expressed as a sum of gaussian contributions $\mathcal{G}$ placed on top of every atom i with coordinates $\mathbf{r}_i$. This takes the following form in bra-ket notation [55]:

$$\langle \mathbf{x}|A;\rho\rangle = \sum_{i\in A}\langle \mathbf{x}|\mathbf{r}_i;\mathcal{G}\rangle \equiv \sum_{i\in A}\mathcal{G}(\mathbf{x}-\mathbf{r}_i), \tag{1.48}$$

where $|A\rangle$ represents all the structural and chemical information of the structure A. This expression can be used to express the contribution from the different local environments as

$$\langle a\mathbf{x}|A;\rho_i\rangle = \sum_{j\in A_i}\langle \mathbf{x}|\mathbf{r}_{ij};\mathcal{G}\rangle\, f_{cut}(r_{ij})\delta_{aa_j}, \tag{1.49}$$

where a defines the atomic specie and the function f_{cut} select the atomic environments within a cutoff distance from the central atom i. This representation, centered on an atom i, is translational invariant. In order to include rotational invariance, high-order correlations between atomic environments should be considered by applying the Haar integral [55]

$$\left\langle a_1\mathbf{x}_1;...;a_v\mathbf{x}_v|A;\overline{\rho_i^{\otimes v}}\right\rangle = \sum_{k=0,1}\int_{SO(3)} d\hat{R}\prod_{j=1}^{v}\left\langle a_j\mathbf{x}_j|\hat{R}\hat{i}^k|A;\rho_i\right\rangle, \tag{1.50}$$

where $\overline{\rho_i^{\otimes v}}$ represents the average over all the improper rotations $\hat{R}\hat{i}^k$ of the tensor product of v atom-centered densities. The operators $\hat{i}$ and $\hat{R}$ represent inversion and rotation respectively while the sum over k accounts for the inversion symmetry.

In this thesis the Smooth Overlap of Atomic Positions (SOAP) [54] descriptor will be used. This representation can be obtained from Equation 1.50 by expanding the translationally-invariant representation in Equation 1.49 in a basis of radial functions $R_n(x) = \langle x|n\rangle$ and spherical harmonics $Y_m^l(\hat{\mathbf{x}}) = \langle\hat{\mathbf{x}}|lm\rangle$ as

$$\langle anlm|A;\rho_i\rangle = \sum_{j\in A_i}\int d\mathbf{x}\langle n|x\rangle\langle lm|\hat{\mathbf{x}}\rangle\,\langle \mathbf{x}|\mathbf{r}_{ij};\mathcal{G}\rangle\, f_{cut}(r_{ij})\delta_{aa_j}. \tag{1.51}$$

The 3-body-order representation, called power spectrum, can be obtained by truncating the expansion in Equation 1.50 at $v = 2$ and rewriting it in the new basis as

$$\left\langle a_1 n_1; a_2 n_2; l | A; \overline{\rho_i^{\otimes 2}} \right\rangle = \frac{1}{\sqrt{2l+1}} \sum_m \langle A; \rho_i | a_2 n_2 l m \rangle \langle a_1 n_1 l m | A; \rho_i \rangle , \qquad (1.52)$$

here $|\overline{\rho_i^{\otimes 2}}\rangle$ represents the symmetrized 3-body correlation of the atom density centred on the atom i. In practice the expansion is truncated at n_{max} and l_{max} so the SOAP will contain $n_{max}^2 l_{max}$ elements. This should be taken into account in order to estimate the computational resources needed for the power spectrum representation.

Chapter 2

Core-level spectroscopy simulations in the DFT framework

Core-level spectroscopies are essential tools in material science for probing the electronic structure and chemical environment of materials. The most relevant core-level spectroscopies include X-ray Photoelectron Spectroscopy (XPS), X-ray Absorption Spectroscopy (XAS), X-ray Emission Spectroscopy (XES), Resonant Inelastic X-ray Scattering (RIXS) and X-ray Raman scattering (XRS). In this thesis we will focus on XPS, XRS and XAS (in the near edge structure XANES).

XPS is commonly used for both qualitative and quantitative surface analysis of solids up to 10 nm. It enables the identification and quantification of nearly all elements present on a surface, excluding only hydrogen and helium, thereby providing comprehensive elemental composition data [56]. XAS, on the other hand, is a versatile technique applied across soft and hard X-ray ranges to investigate coordination environments, oxidation states, and spin states in solids, liquids, and gases, offering high resolution and rapid data acquisition [57]. Meanwhile, XRS, which utilizes hard X-rays around 10 keV, offers several advantages over conventional XAS and XPS for studying soft X-ray absorption edges (e.g., C, O, and Li K-edges) with bulk sensitivity in vacuum-free conditions. This makes XRS particularly useful for examining energy storage systems under operating conditions [58].

The following sections will be focused on the theoretical description of these spectroscopies within the DFT framework, along with the corresponding technical details involved in the *ab initio* simulations.

2.1 Theoretical treatment of core-level spectroscopies

The fundamental interactions involved in core-level spectroscopies can be described by the Hamiltonian for electrons in a quantized electromagnetic field. This Hamiltonian, in a non-relativistic approximation, is given by

$$\hat{H} = \hat{H}_e + \hat{H}_{int} + \hat{H}_{field}, \tag{2.1}$$

where $\hat{H}_e$, $\hat{H}_{int}$, and $\hat{H}_{field}$ are the electronic (defined in Equation 1.10), electron-photon interaction and free photon field parts of the Hamiltonian respectively. We can move to a single particle framework by replacing the many electron Hamiltonian $\hat{H}_e$ by the Kohn-Sham Hamiltonian (Equation 1.16), then the system of non interacting electrons in a photon field reads

$$\hat{H} = \hat{H}^{KS} + \hat{H}_{int} + \hat{H}_{field}, \tag{2.2}$$

The non-electronic contributions to $\hat{H}$ are defined, in the Coulomb gauge[1], as [59]

$$\hat{H}_{int} = -\frac{e}{m_e c}\hat{\mathbf{p}} \cdot \hat{\mathbf{A}}(\mathbf{r}) + \frac{e^2}{2m_e c^2}\hat{A}^2(\mathbf{r}) = \hat{H}_1 + \hat{H}_2, \tag{2.3}$$

$$\hat{H}_{field} = \sum_{\mathbf{k},\lambda} \hbar\omega_{\mathbf{k}} \left(\hat{a}^\dagger_{\mathbf{k},\lambda}\hat{a}_{\mathbf{k},\lambda} + \frac{1}{2} \right), \tag{2.4}$$

with

$$\hat{\mathbf{A}}(\mathbf{r}) = \sum_{\mathbf{k},\lambda} A_{0\mathbf{k}}\mathbf{e}_{\mathbf{k},\lambda} \left(\hat{a}^\dagger_{\mathbf{k},\lambda} e^{-i\mathbf{k}\cdot\mathbf{r}} + \hat{a}_{\mathbf{k},\lambda} e^{i\mathbf{k}\cdot\mathbf{r}} \right), \tag{2.5}$$

where c is the speed of light in vacuum and $\hat{\mathbf{A}}(\mathbf{r})$ is the quantum-mechanical operator of the vector potential with $\hat{a}^\dagger_{\mathbf{k},\lambda}$ ($\hat{a}_{\mathbf{k},\lambda}$) being the creation (annihilation) operator for a photon with wave vector $\mathbf{k}$, polarization vector $\mathbf{e}_{\mathbf{k},\lambda}$, frequency $\omega_{\mathbf{k}}$, momentum $\hbar\mathbf{k}$, and energy $\hbar\omega_{\mathbf{k}}$. The coefficient $A_{0\mathbf{k}}$ is the corresponding amplitude of $\hat{\mathbf{A}}(\mathbf{r})$ for the photon mode $\mathbf{k}$. In the scope of this analysis, the spin-dependent terms responsible for the X-ray magnetic scattering [60] were excluded from $\hat{H}_{int}$.
The total Hamiltonian can now be rewritten as

$$\hat{H} = \hat{H}_0 + \hat{H}_{int}, \tag{2.6}$$

with $\hat{H}_0$ defined as the total Hamiltonian without electron-photon interaction with eigenfunctions

[1] The Coulomb gauge $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$ leads to $\hat{\mathbf{p}} \cdot \hat{\mathbf{A}} = \hat{\mathbf{A}} \cdot \hat{\mathbf{p}}$

$$|I\rangle = \left|\phi_i^{KS}, n_{\mathbf{k},\lambda}\right\rangle = \left|\phi_i^{KS}\right\rangle \otimes |n_{\mathbf{k},\lambda}\rangle\,, \tag{2.7}$$

were $\left|\phi_i^{KS}\right\rangle$ and $|n_{\mathbf{k},\lambda}\rangle$ are the wavefunctions of the non-interacting electronic problem and the radiation field in the occupation number $n_{\mathbf{k},\lambda}$ representation[2], respectively. By considering $\hat{H}_{int}$ as a perturbation on $\hat{H}_0$, according to the Fermi's golden rule, the transition probability per unit time from an initial state $|I\rangle$ to a final state $|F\rangle$ is given by

$$w_{if} = \frac{2\pi}{\hbar}\left|\langle F|\,\hat{O}\,|I\rangle\right|^2 \delta(E_I - E_F), \tag{2.8}$$

with

$$E_M = E_m + \sum_{\mathbf{k},\lambda}\left(n_{\mathbf{k},\lambda}^M + \frac{1}{2}\right)\hbar\omega_{\mathbf{k}}, \tag{2.9}$$

where E_m $(m = i, f)$ and $n_{\mathbf{k},\lambda}^M$ $(M = I, F)$ are the energy of the electronic system part and the occupation numbers respectively in the state $|M\rangle$. The transition operator $\hat{O}$ is defined by the Lippmann–Schwinger equation as [59]

$$\hat{O} = \hat{H}_{int} + \hat{H}_{int}\frac{1}{E_I - \hat{H}_0 + i\frac{\Gamma}{2}}\hat{O}, \tag{2.10}$$

with a photohole lifetime energy broadening Γ, and it can be solved iteratively as follows:

$$\hat{O} = \hat{H}_{int} + \hat{H}_{int}\frac{1}{E_I - \hat{H}_0 + i\frac{\Gamma}{2}}\hat{H}_{int} + ... \tag{2.11}$$

In order to determine which part of $\hat{H}_{int}$ contributes to the transition probabilities we must consider how many photons are involved in the transition process, taking into account the linearity of $\hat{\mathbf{A}}(\mathbf{r})$ in $\hat{a}_{\mathbf{k},\lambda}^\dagger$ and $\hat{a}_{\mathbf{k},\lambda}$. For one-photon processes (XPS, XAS) only the $\hat{H}_1$ term survives in the first order of the expansion and the transition operator becomes

$$\hat{O}_{one-photon} = \hat{O}_{XPS/XAS} = \hat{H}_1. \tag{2.12}$$

In the case of two-photon process only survives $\hat{H}_2$ in first order and $\hat{H}_1$ in second order (resonant process) leading to

$$\hat{O}_{two-photon} = \hat{O}_{XRS} + \hat{O}_{RIXS} = \hat{H}_2 + \hat{H}_1\frac{1}{E_I - \hat{H}_0 + i\frac{\Gamma}{2}}\hat{H}_1. \tag{2.13}$$

[2] $n_{\mathbf{k},\lambda} = 0, 1, 2, ..$ are the eigenvalues of the equation $\hat{a}_{\mathbf{k},\lambda}^\dagger\hat{a}_{\mathbf{k},\lambda}\,|n_{\mathbf{k},\lambda}\rangle = n_{\mathbf{k},\lambda}\,|n_{\mathbf{k},\lambda}\rangle$

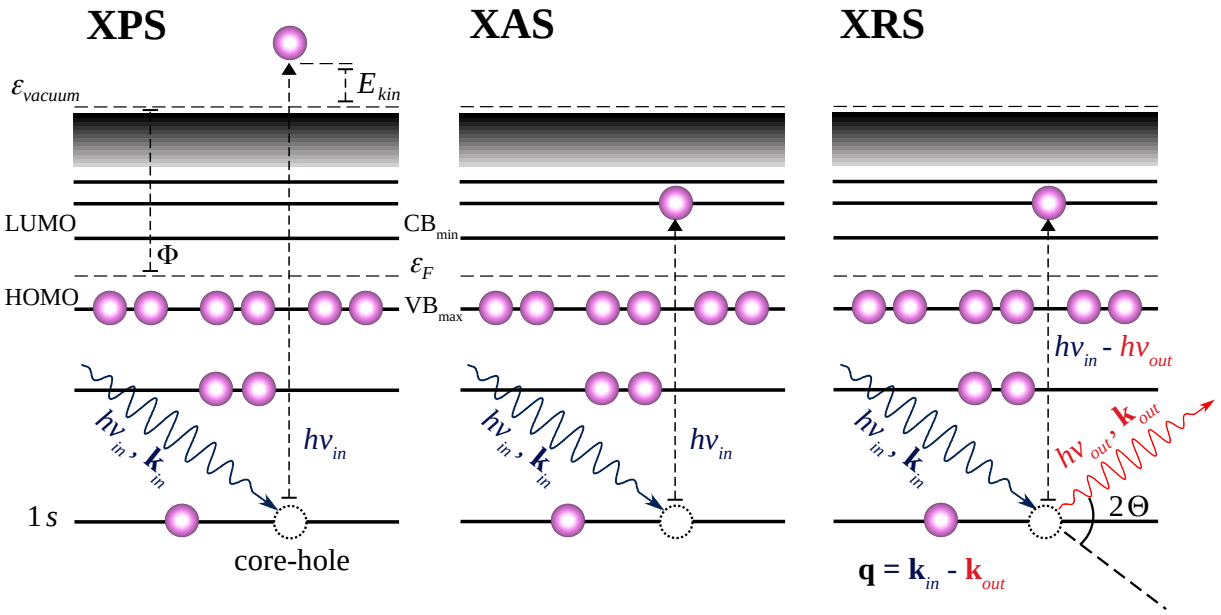

Figure 2.1: Illustration of the XPS, XAS and XRS spectroscopies. The Fermi (ε_F), vacuum (ε_{vacuum}), valence band maximum (VB_{max}), conduction band minimum (CB_{min}), higher occupied molecular orbital (HOMO) and lower unoccupied molecular orbital (LUMO) energy levels are indicated. Incident photons and scattered (in the case of XRS) photons are represented in blue and red respectively. Incident (scattered) photons are labeled according to their energy $h\nu_{in}$ ($h\nu_{out}$) and wavevector $\mathbf{k}_{in}$ ($\mathbf{k}_{out}$). The photoelectron kinetic energy E_{kin} and the work function Φ are represented by dashed segments.

2.2 X-ray Photoelectron Spectroscopy

2.2.1 Principles of XPS

The theoretical foundation of XPS lies in the photoelectric effect. As illustrated in Figure 2.1, this physical process occurs when a sample (S) is irradiated with X-rays of frequency ν, causing the emission of photoelectrons with kinetic energy after leaving the sample E^S_{kin} satisfying the energy conservation equation

$$E^S_{kin} = h\nu - BE^F - \Phi^S, \tag{2.14}$$

where $h\nu$ is the photon energy, Φ^S is the work function of the material and BE^F is the binding energy referred to the Fermi level (ε_F). Since the work function is the potential barrier at the surface, computed as the energy difference between ε_F and the vacuum level ε_{vacuum}, it is material-dependent. In practice, Equation 2.14 is reformulated in terms of the work function of the detector Φ^D [61], which is an

experimental constant, and the kinetic energy E^D_{kin} measured at the detector

$$E^D_{kin} = h\nu - BE^F - \Phi^D. \tag{2.15}$$

This relation is valid only if there is a sufficient electrical contact between the sample and the detector and if there are enough charges to align both Fermi levels. In such cases, ε_F becomes a convenient reference level for the binding energy. However, if the previous conditions are not fulfilled, as e.g., in a gas, the only common energy reference for the sample and the detector is ε_{vacuum}. Under this circumstances $E^S_{kin} = E^D_{kin}$ and the binding energies are referred to the vacuum level, leading to the equation

$$E^D_{kin} = h\nu - BE^{vacuum}. \tag{2.16}$$

The previous analysis is crucial in order to decide which simulation approach provides the closest approximation (within the practical limitations) to what is actually measured in the experiment. This point will be addressed in detail later, but as a general idea BE^F (BE^{vacuum}) quantifies the total energy difference between the system in the ground state and the system with a core-electron promoted to ε_F (ε_{vaccum}). In most cases, the difference between BE is enough to obtain information about the chemical environment, electronic structure, and oxidation states of the element under study. This quantity, called core-level shift (CLS), or chemical shift, is simply expressed as

$$CLS_{i,j} = BE_i - BE_j. \tag{2.17}$$

In a typical XPS experiment the intensity of the photoelectron current I is measured against the BE and for a given photon energy $h\nu$ it is proportional to the photoelectron cross section $\sigma_{if}(h\nu)$ as follows [62]

$$I_i(h\nu) \propto \sum_f n_i \ell A J_{h\nu} \sigma_{if}(h\nu), \tag{2.18}$$

where i is the target electron state, n_i is the density of the specie containing the i state, ℓ is the photoelectron inelastic mean free path, A is the X-ray beam spot area and $J_{h\nu}$ is the flux density of incoming photons of energy $h\nu$. The sum runs over all the final states f caracterized by a core-hole in i and a free-electron continuum state in vacuum with kinetic energy E_{kin}. Usually, in the context of simulations, only the $\sigma_{if}(h\nu)$ is meaningful for the study of XPS, since the other terms are constants for a given experimental setup. Considering that this quantity is directly related to the transition probabilities w_{if} with the corresponding transition operator $\hat{O}_{XPS}$, we can write I_i as

$$I_i \propto \sum_f n_i \sigma_{if} \propto \sum_f n_i w_{if} = \sum_f n_i \left| \langle F | \, \hat{\mathbf{p}} \cdot \hat{\mathbf{A}}(\mathbf{r}) \, | I \rangle \right|^2 \delta(E_I - E_F). \tag{2.19}$$

Taking into account the definitions of $\hat{\mathbf{A}}(\mathbf{r})$, E_I, E_F and the explicit forms of $|I\rangle$ and $|F\rangle$ for a single-photon absorption process

$$|I\rangle = \left|\phi_i^{KS}, n_{\mathbf{k},\lambda}\right\rangle, \tag{2.20}$$

$$|F\rangle = \left|\phi_f^{KS}, n_{\mathbf{k},\lambda} - 1\right\rangle, \tag{2.21}$$

Equation 2.19 transforms into

$$I_i \propto \sum_f n_i \left|\left\langle\phi_f^{KS}\right| e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{e}_{\mathbf{k},\lambda}\cdot\hat{\mathbf{p}}\left|\phi_i^{KS}\right\rangle\right|^2 \delta(E_i - E_f + \hbar\omega_{\mathbf{k}}). \tag{2.22}$$

The matrix elements in Equation 2.22 define the transition matrix, which acts on the amplitudes of the spectrum and introduces selection rules that modify the peak structure by enhancing or suppressing certain transitions. Contrary to ARPES, in XPS no angular dependence is required so the selection rules determined by the transition matrix are not considered. Given this approximation and the fact that all final states ϕ_f^{KS} are assumed to be extended plane waves with energy $E_f = E_{kin,f}$ (taking the vacuum level as energy reference), the XPS intensity in the single-particle picture reads

$$I_i \propto \sum_f n_i \delta(E_i - E_{kin,f} + \hbar\omega_{\mathbf{k}}), \tag{2.23}$$

The XPS spectrum in this approximation is represented by a series of delta peaks located at the energy $E_i = BE_i$ of the respective atom electron i with intensity modulated by the number of atoms containing the state i. In this approach, the lifetime broadening due to many-body effects such as electron–electron correlation is neglected although it can be approximated by applying an empirical line broadening to the δ-shaped binding energies. The total spectrum implies the sum over all the initial electronic states $I = \sum_i I_i$.

Finally, we should notice from Equations (2.19) and (2.22) that the photoelectron cross section for the single-photon absorption process satisfies the following relation in the independent particle approximation

$$\sigma_{if} \propto \left|\left\langle\phi_f^{KS}\right| e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{e}_{\mathbf{k},\lambda}\cdot\hat{\mathbf{p}}\left|\phi_i^{KS}\right\rangle\right|^2 \delta(E_i - E_f + \hbar\omega_{\mathbf{k}}), \tag{2.24}$$

which will be used later for the XAS theory.

2.2.2 XPS simulations

The simulation of the XPS process involves computing the BE for a group of target atoms. Various methods exist for determining BEs, categorized into two groups: initial state approximations (ISA) and final state approximations. Specifically, within the ISA BEs are estimated using the Koopmans' theorem [63] in which the

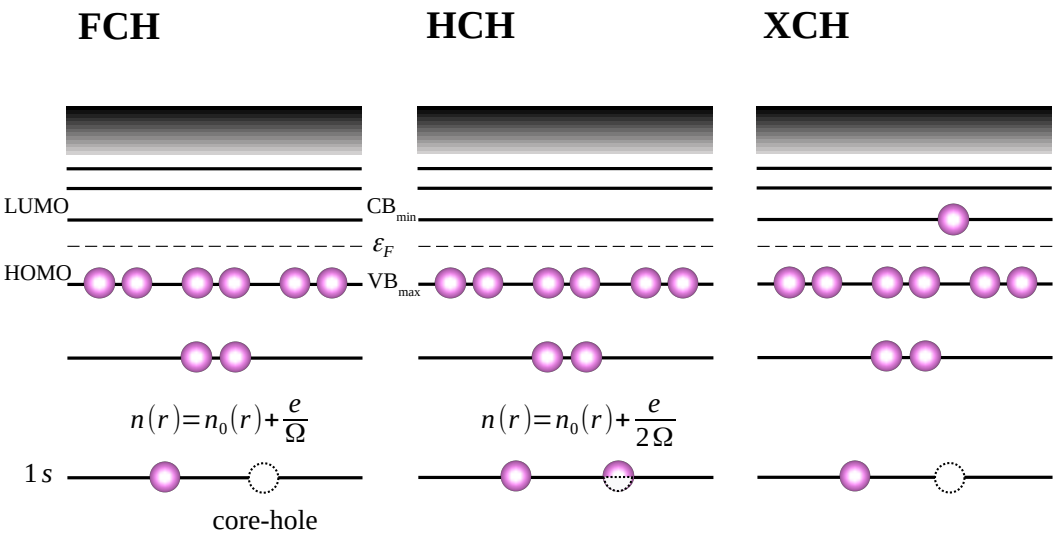

Figure 2.2: Representation of the final state according to the FCH, HCH and XCH core-hole treatments. The Fermi (ε_F), vacuum (ε_{vacuum}), valence band maximum (VB$_{max}$), conduction band minimum (CB$_{min}$), higher occupied molecular orbital (HOMO) and lower unoccupied molecular orbital (LUMO) energy levels are indicated. The charge neutrality of the cell of volume Ω is restored in the FCH and HCH by adding an uniform charge density term to the electronic density $n_0(r)$.

BE, in Hartree–Fock theory, is the negative value of the corresponding eigenstate. In the DFT framework, this statement only holds for the highest occupied state, limiting this approximation to the calculation of core-level shifts in systems with negligible core-hole relaxation effects [64, 65]. Within the final state approximation we encounter methods based on response theories, including coupled cluster (CC) [66], equation-of-motion CC (EOM-CC) [67], TDDFT [68], Green's function [69] and GW [70]. Alternatively to response theories, methods based on total energy difference were proposed, including the equivalent-core or $Z+1$, in which no core-hole is added in the calculations. Moreover, the Delta Self-Consistent Field (ΔSCF) [71, 72, 73, 62] method and its frozen core version, Delta Kohn-Sham (ΔKS) [74, 75], are other total energy difference methods, performing the explicit optimization of the electron wavefunction within the DFT framework in presence of a core-hole. For the purposes of this thesis, the ΔKS method was selected for running all the calculations based on the compromise between accuracy and computational effort.

2.2.2.1 Core-hole treatments

A common challenge arising in DFT core-level spectroscopy calculations with periodic boundary conditions (PBC) is managing the charge introduced by the explicit inclusion of the core-hole. The unit cell in a periodic calculation must be charge-neutral to avoid the divergence of the ion-ion interaction term in the total energy. The procedure adopted to maintain a neutral cell defines what is called core-hole treatment. In Figure 2.2 the most common core-hole treatments are presented. In the full core-hole (FCH) case, the core-electron is completely removed in the final state and replaced by a negative charge background distributed uniformly over the cell volume Ω. This charge correction is usually included in the DFT codes: either implicitly in the total energy evaluation (as e.g., in VASP [76, 77], CASTEP [78] and CP2K [79]) or explicitly in the electronic density (as e.g., in Quantum ESPRESSO) by adding an uniform charge density term

$$n(\mathbf{r}) = n_0(\mathbf{r}) + \frac{q}{\Omega}, \tag{2.25}$$

where $n_0(\mathbf{r})$ is the electronic density of the charged cell and $-q$ is its total charge. It is clear that in the FCH case the cell becomes neutral if we set $q = e$.

The inclusion of the background charge leads to instabilities in the calculated energies [62] which can be mitigated by adopting a supercell approach. However, the cell sizes required to achieve convergence on BE, at least in extended systems, are typically impractical and results should be extrapolated. Despite this issue, the CLSs remain quite consistent with varying cell sizes [62].

In the case of the half core-hole (HCH) method, based on the Slater-Janak transition-state model [80], the core-hole is replaced by a fractional charge $q = e/2$ and the corresponding BE is equal to the eigenvalue of the state with fixed fractional occupation obtained in a self-consistent calculation.[3] Within this approximation the corresponding core-state needs to be included in the calculation, since its eigenenergy is the quantity of interest. This constraint makes HCH impractical for pseudopotential based XPS calculations, at least for target atoms with $Z > 4$, since core states are typically excluded from the calculations. Nevertheless, HCH has proven to be useful within the pseudopotential framework for other core-level spectroscopies like XAS and XRS, where it provides an artificial but effective way of decreasing the core-hole effects in systems with poor screening.

The excited core-hole (XCH) method [82] restores the charge neutrality of the cell in the final state by placing the excited core-electron in the lowest unoccupied Kohn-Sham state. By construction this method avoids all the complications of electrostatic nature due to the background charge, leading to a faster convergence of the BE with respect to the supercell size. In addition, it allows a direct comparison with experiments in the case of metals, where the first empty state is located

[3] This method could be generalized for any fractional occupation: see Ref. 81 for more details.

at ε_F and thus the computed BEs using XCH are already refereed to ε_F as in the experiment.

2.2.2.2 The ΔKS method

Within ΔKS, the BE of a target specie i in a general system is computed as a total energy difference between the excited state E^*, with one electron removed from the core, and the ground state E^0:

$$BE_{ij}^{x} = E^*(n_{c,j} - 1, n_v + x) - E^0(n_{c,j}, n_v) + \Delta E_{corr} \tag{2.26}$$

where $n_{c,j}$ and n_v are the populations of the core state j and valence respectively, $x = 0, 1$ labels the kind of core-hole treatment, FCH and XCH, in the stated order. The ΔE_{corr} term brings the BEs to their absolute values (i.e., comparable to experiment) by introducing the core electrons relaxation effects in the presence of a core-hole [83] for a given exchange-correlation functional [75]. In practice, the correction energies in Equation 2.26 can be expressed explicitly as

$$\Delta E_{corr} = \Delta E_{core} - \delta, \tag{2.27}$$

where ΔE_{core} accounts for the core electrons relaxation in the presence of a core-hole, and it is determined performing an all-electron calculation on the target atom with and without the core-hole [83]. This is particularly needed when using pseudopotentials since the core-electrons are not included during the SCF cycle. The second term, δ, is the energy shift introduced by the exchange correlation functional [75], obtained by fitting a linear model using experimental and computed BE, as can be observed in Figure 2.3. In this procedure, the experimental BEs should correspond to systems where the reference energy level is well defined (i.e., accessible through DFT calculations), such as in molecules or metals.

It should be noticed that if the core state has non-zero orbital angular momentum ($l \neq 0$), the corresponding BE will split according to the total angular momentum quantum number j, due to the spin-orbit coupling. The spin-orbit coupling (SOC) splitting Δ_{SOC} is negligible for atoms with $Z < 10$, but it should be considered for the remaining species. This effect could be obtained directly through a constrained fully-relativistic calculation, which is not straightforward and its accuracy is limited by the quality of the pseudopotential. A simpler and more accurate approach is to use the standard scalar-relativistic calculation, which averages out the SOC effects, thus the resulting BE is just the m_j-weighted average of the corresponding j-split BEs. Here m_j is the magnetic quantum number of the total angular momentum. This allows us to express the BE with an empirical SOC correction as

$$BE_{ij}^{SOC} = BE_{il} - \frac{(2j' + 1)\Delta_{SOC}^{j'j}}{2(2l + 1)} \tag{2.28}$$

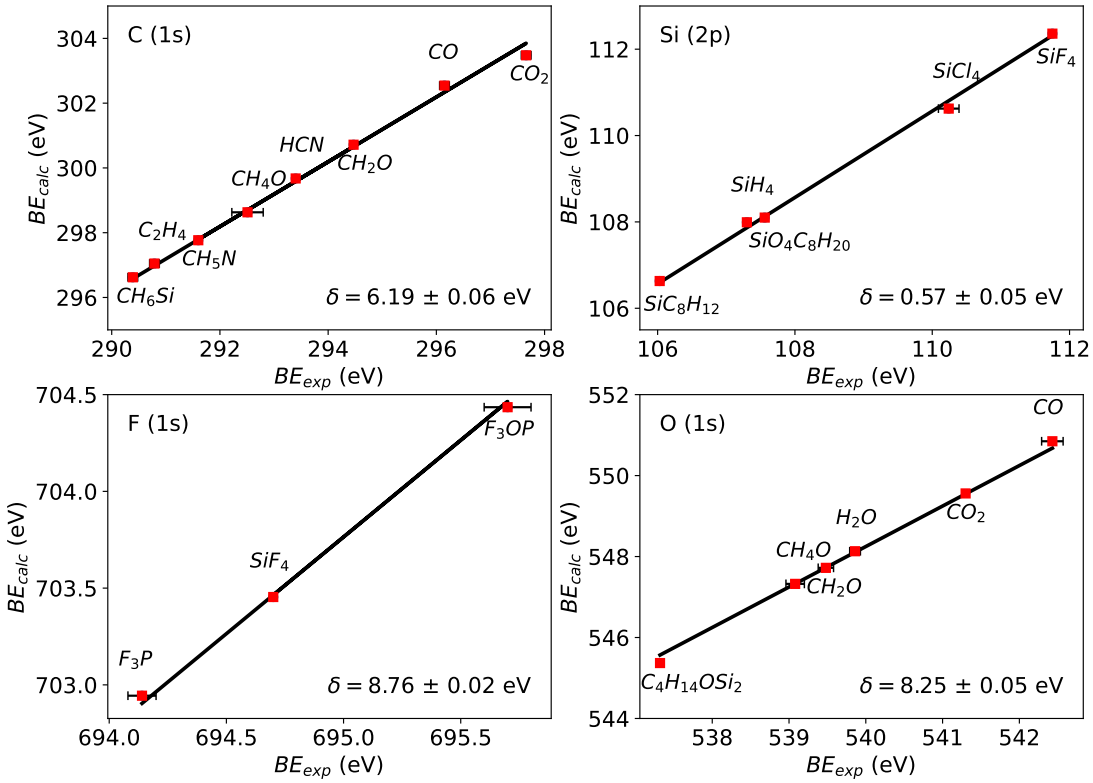

Figure 2.3: Calculated BE compared to experimental values (taken from Ref. 1). The constant offset δ is the correction to the absolute BE due to the PBE exchange correlation functional.

where j and j' are the two possibilities for the total angular momentum quantum number corresponding to l, while $\Delta_{SOC}^{j'j}$ is the experimental energy difference (for the atomic specie i) between the states with total angular momentum quantum number j' and j respectively.
The final XPS spectrum for specie i and core level state j is given by:

$$S_{ij}(E) = \sum_{n_i} V(E - BE_{ij}^{x}, \sigma, \gamma), \tag{2.29}$$

where n_i run over all the atoms of specie i, and $V(E - BE_{ij}^{x}, \sigma, \gamma)$ is the empirical line broadening centered at BE_{ij}^{x} and represented by the convolution of the Gaussian and the Lorentzian distributions, known as the Voigt profile. The Gaussian broadening σ and the Lorentzian broadening γ account for the experimental resolution and core-hole lifetime, respectively.

2.2.2.3 Energy level reference for spectrum alignment in XPS

When using Equation 2.26 the reference energy level should be taken into account, unless we are interested only in core-level shifts. By setting $x = 0$ the BEs are referred to the vacuum level, while for $x = 1$ the reference is placed at the Fermi level (ϵ_F) for metals, or at the bottom of the conduction band in the case of insulators. These considerations establish the natural core-hole treatment we should adopt to compare directly the absolute BE with experiments. For the BE calculations on isolated systems, like molecules, we should use the FCH treatment in the final state in order to compute energies refereed to ε_{vacuum}. Moreover, to obtain converged results in affordable supercell sizes, a correction for long-range electrostatic interactions, like Martyna-Tuckerman [84], needs to be included. On the other hand if the system is metallic the XCH method should be sufficient to reefer the BE to the ε_F. In the case of nonmetals ε_F lays somewhere between the valence band maximum (VB_{max}) and the conduction band minimum (CB_{min}), hindering the energy referencing procedure. In a XPS calculation using XCH the reference level, for nonmetals will be the CB_{min}, which is in general different from ε_F. An option to overcome the referencing problem is to assume that ε_F is placed in the middle of the DFT gap for the ground state system [75]

$$\varepsilon_F^0 = \frac{\varepsilon^0_{\mathrm{CB}_{min}} + \varepsilon^0_{\mathrm{VB}_{max}}}{2}, \tag{2.30}$$

leading to the following correction for the BE

$$BE_{ij}^{F} = BE_{ij}^{1} - \varepsilon_{\mathrm{CB}_{min}} + \varepsilon_F^0, \tag{2.31}$$

where BE_{ij}^{1} and $\varepsilon_{\mathrm{CB}_{min}}$ are computed within the XCH and ε_F^0 is the one obtained in the ground state. It should be notice that for metals $\varepsilon_{\mathrm{CB}_{min}} = \varepsilon_F$ since CB_{min} and VB_{max} coincide. Furthermore, $\varepsilon_F = \varepsilon_F^0$, as there are enough electrons to

screen the core-hole effects on the band structure. This two relations ensure that the BE value remains invariant for metals. In general, the validity of Equation 2.31 is limited by the accuracy of the DFT gap and should be used carefully for large gap systems.

2.3 X-ray Absorption Spectroscopy

2.3.1 XAS principles

The physical quantity of interest in an XAS experiment is the X-ray absorption coefficient $\mu(E)$, which is described, in transmission mode, by the Beer-Lambert law

$$I(E) = I_0 e^{-\mu(E)d}, \tag{2.32}$$

where I_0 and $I(E)$ are the X-ray flux, with photon energy E, before and after the beam is passed through a uniform sample of thickness d. The principal physical processes causing the attenuation of the X-ray photon flux are the photon scattering (elastic and inelastic) and photoelectric absorption. However, in the energy range of interest to XAS (0.1-100 keV)[4], only the photoelectric absorption is dominant [85]. In XAS the incident photon energy varies during the experiment generating transitions not only to the continuum, as in XPS, but also to the unoccupied electronic states. The general shape of the XAS spectrum of an atom consist of a series of core-level excitation thresholds (absorption edges) occurring near the corresponding BEs of the energy levels with principal quantum number $n = 1, 2, 3, ...$ called in XAS as the K, L, M, ... edges of the atom. Moreover, starting from the L edge an additional label is included according to the orbital angular momentum, for example the $2s$, $2p_{1/2}$ and $2p_{3/3}$ states are identified as the L_1, L_2 and L_3 edges respectively.

When the X-ray photon energy E exceeds the absorption edge energy E_0 , the corresponding photoelectron acquires a kinetic energy $E_{kin} = E - E_0$ and undergoes scattering by the neighboring atomic potentials. Since the photoelectron mean free path is determined by its kinetic energy, it is natural to divide the XAS spectrum into two regions based on the scattering regime. For small E_{kin} the scattering efficiencies are high [86] and the photoelectron suffers multiple scatterings with the neighboring atomic potentials. This energy region, covering around 50 eV above E_0, determines the part of the XAS spectrum called X-ray absorption near-edge structure (XANES) which also includes the pre-edge region determined by transitions into bound states. At higher E_k the single and weak multiple scattering events are dominant defining the part of the spectrum called extended X-ray absorption fine structure (EXAFS). An example of the two regions of the XAS spectrum for the Cu K edge is presented in the Figure 2.4.

[4]This energy range spans over the deepest core-level BE for most of the atoms

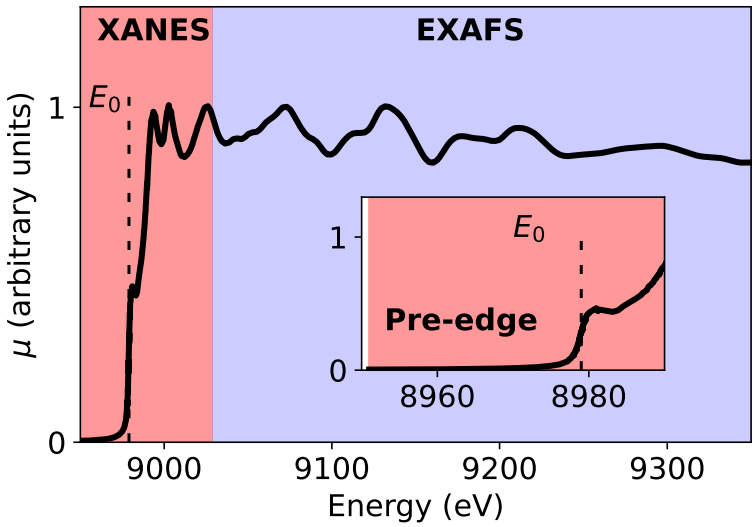

Figure 2.4: Cu K edge XAS spectrum from Ref. 2. The two fine structure regimes XANES and EXAFS are depicted. The edge energy E_0 defines the upper bound of the pre-edge region of XANES.

Since for the typical photon energies involved in XAS the photoelectric absorbtion events are dominant, the X-ray absorption coefficient is directly related to the cross section for the single-photon absorption. Then $\mu(E)$ is also proportional to the transition probabilities w_{ij} and according to the Equation 2.22 we can express

$$\mu_i(\omega_{\mathbf{k}}) \propto \sum_f \left|\left\langle \phi_f^{KS} \right| e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{e}_{\mathbf{k},\lambda} \cdot \hat{\mathbf{p}} \left| \phi_i^{KS} \right\rangle\right|^2 \delta(E_i - E_f + \hbar\omega_{\mathbf{k}}), \tag{2.33}$$

where $\mu_i(\omega_{\mathbf{k}})$ is the contribution to $\mu(\omega_{\mathbf{k}})$ for the transitions with initial state $\left|\phi_i^{KS}\right\rangle$. After expanding the transition operator in series as

$$e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{e}_{\mathbf{k},\lambda} \cdot \mathbf{p} = \mathbf{e}_{\mathbf{k},\lambda} \cdot \mathbf{p} + i(\mathbf{e}_{\mathbf{k},\lambda} \cdot \mathbf{p})(\mathbf{k} \cdot \mathbf{r}) + ..., \tag{2.34}$$

and taking into account the commutation relation $\hat{\mathbf{p}} = -i[\mathbf{r}, \hat{H}^{KS}]$, Equation 2.33 transforms into

$$\mu_i(\omega_{\mathbf{k}}) \propto \sum_f \omega_{\mathbf{k}}^2 \left|\left\langle \phi_f^{KS} \right| \hat{o} \left| \phi_i^{KS} \right\rangle\right|^2 \delta(E_i - E_f + \hbar\omega_{\mathbf{k}}), \tag{2.35}$$

with

$$\hat{o} = \mathbf{e}_{\mathbf{k},\lambda} \cdot \mathbf{r} + \frac{i}{2}(\mathbf{e}_{\mathbf{k},\lambda} \cdot \mathbf{r})(\mathbf{k} \cdot \mathbf{r}) + ... = \hat{o}_{E1} + \hat{o}_{E2} + ..., \tag{2.36}$$

where E1, E2, ... index the so-called dipole, quadrupole, ... electric transition channels. Provided that the X-ray wavelenght $\lambda = \frac{2\pi}{|\mathbf{k}|}$ is much larger than the characteristic spatial localization of the core state d_i,

$$\lambda = \frac{hc}{\mathrm{BE}_i} \gg d_i, \tag{2.37}$$

the dipole transitions $E1$ are dominant. Higher order terms, in particular E2, should be considered at higher photon energies and for heavier elements edges where Equation 2.37 does not hold. Even if the contribution from E2 is much smaller than E1, it follows different selection rules allowing to probe dipole-forbidden transitions. In the case of a single particle approximation the selection rules reduces to $\Delta l = \pm 1$ for E1 and $\Delta l = 0, \pm 2$ in the case of E2.

2.3.2 XAS simulations

Modern methods for computing XAS are classified by the level of theory and the computational approaches. Within the DFT framework we find real and reciprocal space methods. The approaches in the first category are based either on wave functions, as e.g., implemented in ADF [87] and ORCA [88], or multiple scattering in FEFF [89], while in the second category the methods rely on the band structure like e.g., in Wien2K [90] and Quantum ESPRESSO [37]. One-particle approaches provide satisfactory results in most cases, but they fail in the description of partly localized and atomic edges, where many-body and multiplet effects are important [91]. In such cases methods beyond the one-particle approximation are used including TDDFT [92], as in FDMNES [93] and Bethe Salpeter equations (BSE) as implemented in the OCEAN code [94]. Other approaches are tailored for more specific conditions, like the charge transfer multiplet model [95] which offers an accurate description of the excitonic effects in the $L_{2,3}$ edges of transition metal ions. Post Hartree–Fock schemes adopting a linear response approach [96] or an equation of motion coupled-cluster [97] yield a very good description of the initial and final state wave functions, but their applicability is computationally limited to small clusters.

2.3.2.1 Reciprocal-space pseudopotential scheme for XANES

Computing the XANES spectrum reduces to calculating the transition matrix in Equation 2.35. In a reciprocal-space pseudopotential scheme [98], as the one implemented in the XSpectra code [99, 100, 101] of Quantum ESPRESSO [36, 37, 38], this is done using the Projector Augmented-Wave (PAW) method [102]. In this procedure the all-electron final state $\left|\phi_f^{KS}\right\rangle$ in the transition matrix

$$M_{if} = \left\langle \phi_f^{KS} \right| \hat{o} \left| \phi_i^{KS} \right\rangle, \tag{2.38}$$

is expressed in terms of the pseudo wavefunctions $\left|\tilde{\phi}_f^{KS}\right\rangle$ as

$$\left|\phi_f^{KS}\right\rangle = \hat{\mathcal{T}} \left|\tilde{\phi}_f^{KS}\right\rangle, \tag{2.39}$$

where $\hat{\mathcal{T}}$ is a linear operator that acts within spherical regions centered on each atomic site $\mathbf{R}$, called augmentation regions $\Omega_{\mathbf{R}}$. This operator is expressed in the form

$$\hat{\mathcal{T}} = \hat{1} + \sum_{\mathbf{R},n} \left(|\varphi_{\mathbf{R},n}\rangle - |\tilde{\varphi}_{\mathbf{R},n}\rangle \right) \langle \tilde{p}_{\mathbf{R},n}| \,, \tag{2.40}$$

being $|\varphi_{\mathbf{R},n}\rangle$ and $|\tilde{\varphi}_{\mathbf{R},n}\rangle$ the all-electron and pseudo partial wave functions, respectively: they are identical outside $\Omega_{\mathbf{R}}$ and both form a complete basis for the corresponding all-electron and pseudo final states within the region defined by $\Omega_{\mathbf{R}}$. The vectors $|\tilde{p}_{\mathbf{R},n}\rangle$, called projector functions labeled by the index n, are equal to zero outside $\Omega_{\mathbf{R}}$ and satisfy the condition $\langle \tilde{p}_{\mathbf{R},n}|\tilde{\varphi}_{\mathbf{R}',n'}\rangle = \delta_{\mathbf{R}\mathbf{R}'}\delta_{nn'}$. Since the initial state $\left|\phi_i^{KS}\right\rangle$ is localized in the absorbing atom $\mathbf{R}_0$ and the condition $\hat{\mathcal{T}} = \hat{1}$ holds outside the augmentation region, the transition matrix can be expressed as

$$M_{if} = \left\langle \tilde{\phi}_f^{KS} | \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right\rangle, \tag{2.41}$$

with

$$\left|\tilde{\varphi}_{\mathbf{R}_0}^{KS}\right\rangle = \sum_n |\tilde{p}_{\mathbf{R}_0,n}\rangle \langle \tilde{\varphi}_{\mathbf{R}_0,n}| \, \hat{o} \left|\phi_i^{KS}\right\rangle. \tag{2.42}$$

It is important to note that the transition matrix includes, besides the final pseudo state, the initial core state, which is not part of the calculation. The core state should be provided as input to the calculation, usually within the pseudopotential of the absorbing atom. Another concern in Equation 2.35 is the summation over all the final empty states adding a significant computational effort. This is solved by transforming the summation over empty states into a summation over occupied states. Starting with the equation for μ_i arranged in a convenient form

$$\mu_i(\omega_{\mathbf{k}}) \propto \omega_{\mathbf{k}}^2 \left\langle \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right| \left(\sum_f \left|\tilde{\phi}_f^{KS}\right\rangle \delta(E_i - E_f + \hbar\omega_{\mathbf{k}}) \left\langle \tilde{\phi}_f^{KS}\right| \right) \left|\tilde{\varphi}_{\mathbf{R}_0}^{KS}\right\rangle, \tag{2.43}$$

where the object inside the parenthesis is related with the Green function operator $\hat{\tilde{G}}$ associated with the pseudo Hamiltonian $\hat{\tilde{H}}^{KS} = \hat{\mathcal{T}}^\dagger \hat{H}^{KS} \hat{\mathcal{T}}$ as

$$\sum_f \left|\tilde{\phi}_f^{KS}\right\rangle \delta(E_i - E_f + \hbar\omega_{\mathbf{k}}) \left\langle \tilde{\phi}_f^{KS}\right| = -\frac{1}{\pi} Im \left[\hat{\tilde{G}}(E)\right], \tag{2.44}$$

with

$$\hat{\tilde{G}}(E) = (E - \hat{\tilde{H}}^{KS} + i\gamma)^{-1}, \tag{2.45}$$

being $E = E_i + \hbar\omega_{\mathbf{k}}$ and γ a small positive number. The Equation 2.46 can be reformulated as

$$\mu_i(\omega_{\mathbf{k}}) \propto \omega_{\mathbf{k}}^2 \left\langle \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right| (E - \hat{\tilde{H}}^{KS} + i\gamma)^{-1} \left| \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right\rangle, \tag{2.46}$$

whose matrix elements are obtained using the recursive method of Lanczos [103, 104] by defining a new basis $|u_i\rangle$ in which $\hat{\tilde{H}}^{KS}$ is tridiagonal. Starting with the definitions

$$|u_0\rangle = \frac{\left|\tilde{\varphi}_{\mathbf{R}_0}^{KS}\right\rangle}{\sqrt{\left\langle \tilde{\varphi}_{\mathbf{R}_0}^{KS} | \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right\rangle}}, \tag{2.47}$$

$$\hat{\tilde{H}}^{KS} |u_i\rangle = a_i |u_i\rangle + b_{i+1} |u_{i+1}\rangle + b_i |u_i\rangle, \tag{2.48}$$

with $a_i = \langle u_i | \hat{\tilde{H}}^{KS} |u_i\rangle$ and $b_i = \langle u_i | \hat{\tilde{H}}^{KS} |u_{i-1}\rangle = \langle u_{i-1} | \hat{\tilde{H}}^{KS} |u_i\rangle$, leads to the following expression for the matrix elements

$$\left\langle \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right| (E - \hat{\tilde{H}}^{KS} + i\gamma)^{-1} \left| \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right\rangle = \frac{\left\langle \tilde{\varphi}_{\mathbf{R}_0}^{KS} | \tilde{\varphi}_{\mathbf{R}_0}^{KS} \right\rangle}{a_0 - E - i\gamma - \cfrac{b_1^2}{a_1 - E - i\gamma - \cfrac{b_2^2}{\ddots}}}. \tag{2.49}$$

A terminator can be applied to truncate the continued fraction, assuming that the coefficients (a_i, b_i) are equivalent to (a_N, b_N) for $i > N$. This assumption results in an analytical form of the terminator. Equation 2.49 is valid for norm-conserving pseudopotentials and its extension to ultrasoft pseudopotentials is done by taking into account the generalized orthonormality constraints on the pseudo wave functions and the corresponding transformation of the Schrödinger equation into a generalised eigenvalue problem [100].

Note that the parameter γ determines the number of iterations for a given convergence threshold: larger values will reduce the computational effort at the cost of spectral resolution. Moreover, the only constraint on γ is that it must be positive, allowing it to be either constant or energy-dependent. Its value is usually chosen to match with experimental data.

2.3.2.2 Energy level reference for spectrum alignment in XAS

As mentioned before, a sample exposed to X-rays will absorb part of the incoming photons. At a particular energy E_0 a sharp rise, known as absorption edge, will be observed (see Figure 2.4). The alignment procedure thus reduces to finding the absolute value of E_0 for the system of interest.

For metals we can assume that in most cases $E_0 = BE^F$, since the XPS binding energy pins the Fermi level which coincides with the lowest empty state. However, this assumption is no longer valid if the transition to the lowest empty state is forbidden by selection rules. In such cases E_0 is slightly larger than BE^F, thus the general relation for metals becomes

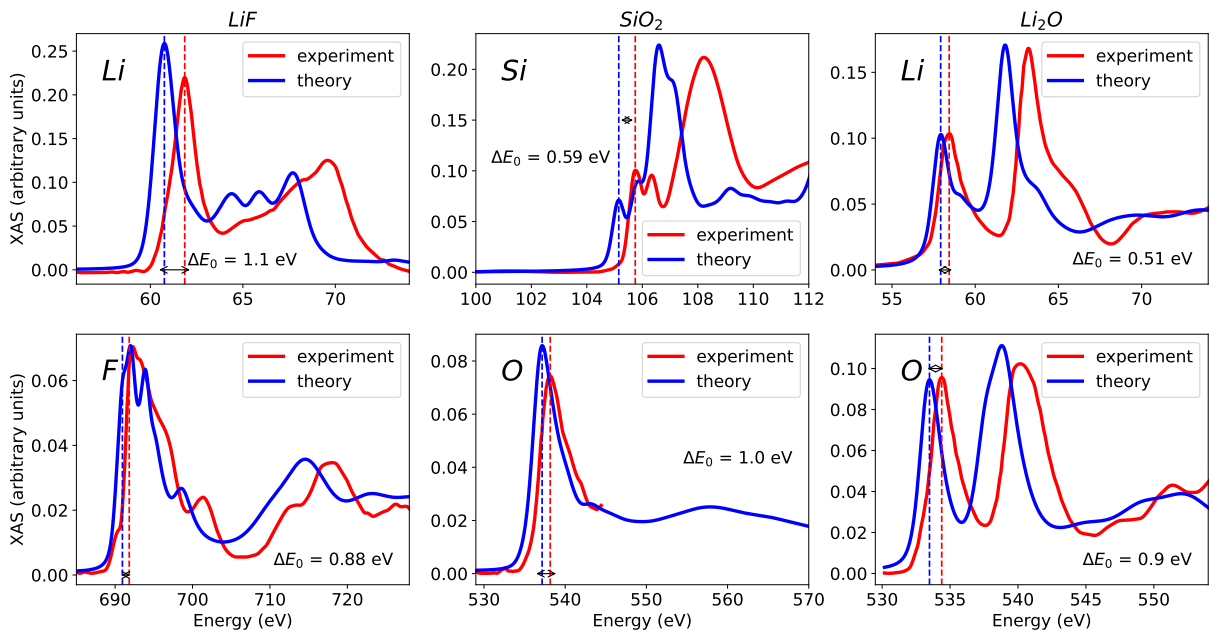

Figure 2.5: Comparison between experiments and simulations of K and $L_{2,3}$ (Only for Si in SiO_2) edges XAS spectrum for LiF (left), SiO_2 (center) and Li_2O (right). The difference between the theoretical and experimental edge energy ΔE_0 is indicated in all the cases. The experimental references were taken from Ref. 3 (LiF and Li_2O) and Ref. 4 (SiO_2). The simulations where obtained using the XSpectra code of Quantum ESPRESSO.

$$E_0 = BE^F + (\varepsilon_{f_0} - \varepsilon_F), \tag{2.50}$$

were ε_{f_0} is the energy of the first empty state allowed by selection rules. Since E_0 is referred to BE^F the final state can be represented naturally by the XCH approximation.

For non-metallic extended systems, in the absence of impurities or defects, there are no available states at the Fermi level: then E_0 is different from the BE^F and the first available state is placed at the CB_{min}. This means that usually E_0 is located at the binding energy measured from $\varepsilon_{\mathrm{CB}_{min}}$. If the selection rules are taken into account the general form for the absorption energy edge is

$$E_0 = BE^{\mathrm{CB}_{min}} + (\varepsilon_{f_0} - \varepsilon_{\mathrm{CB}_{min}}). \tag{2.51}$$

As in the case of metals, the XCH treatment of the final state will pin the binding energy in the CB_{min} of the system in the excited state. The main limitation of Equation 2.51 is the underestimation of the empty states energies, including $\varepsilon_{\mathrm{CB}_{min}}$, within the DFT framework. This issue leads to a rigid shift of the theoretical value E_0 and a contraction of the spectrum features, as apparent in Figure 2.5.

The last scenario involves isolated systems (molecules and clusters), where the binding energies are referred to ε_{vacuum}. As in the case of non-metals, E_0 differs from the binding energy since the first available empty states are located at ε_{LUMO}, measured from the vacuum level, which lays at lower energy than ε_{vacuum}. Taking into account the selection rules the general form for E_0 in finite systems is

$$E_0 = BE^{vacuum} - \varepsilon_{LUMO} + (\varepsilon_{f_0} - \varepsilon_{LUMO}), \tag{2.52}$$

where all the quantities on the right hand side are refereed to ε_{vacuum}. As was discussed in the XPS section, BE^{vacumm} is determined in a FCH calculation including a correction for long-range electrostatic interactions. However, the effect of a full core-hole is not effectively screened for some molecular systems. In this context, the HCH method is used instead for representing the final state in a XAS calculation, providing higher accuracy [105, 106, 107]. The previous considerations establish a protocol for computing E_0 in finite systems

$$E_0 = [BE^{vacuum}]_{FCH} + [\varepsilon_{f_0} - 2\varepsilon_{LUMO}]_{HCH}\,, \tag{2.53}$$

where the brackets are labeled according to the method used in each part of the equation. This hybrid approach, where the transition matrix elements computed with HCH are corrected by the BE of the core state calculated with a full core-hole was proposed by Triguero *et al* [108] and systematically used with results comparable to experiments [109, 110, 62]. Similarly to the non-metals case, for isolated systems, DFT can also underestimate the value of ε_{LUMO}, but the errors are generally less severe due to the localized nature of the electronic states (see Figure 2.6).

It is worth pointing out that, regardless of the type of system, E_0 is unique for every edge of the system. If we have non equivalent atoms, then according to the Equations (2.50) to (2.52) we will obtain a different value of E_0 for each of them. In this case the minimum of all the absorption edges across all non equivalent atoms will define the value of E_0 for the whole system. The final spectrum will be generated by adding the corresponding contributions for all the non equivalent atoms, taking into account they relative CLS and multiplicity.

The cases described above with their respective core-hole treatment establish the guidelines for the XAS spectrum alignment directly from theoretical considerations. Alternative core-hole treatments can be tested on the same system, either adapting the expressions for computing E_0 or using an experimental reference for alignment.

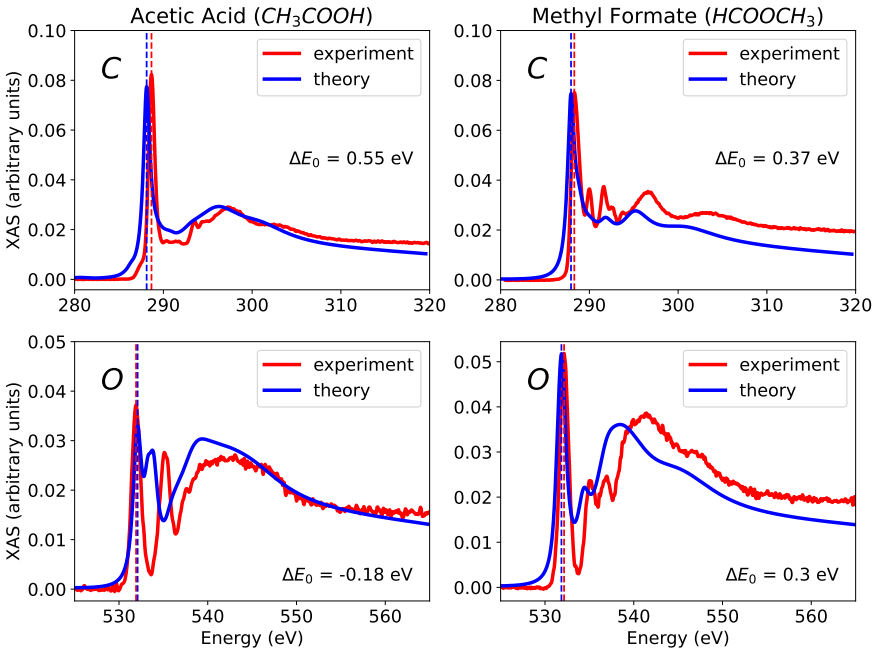

Figure 2.6: Comparison between experiments and simulations of K edges XAS spectrum for acetic acid (left) and methyl formate (right). The difference between the theoretical and experimental edge energy ΔE_0 is indicated in all the cases. The experimental references for acetic acid [5] and methyl formate [6] were taken from the Gas phase core excitation database [7]). The simulations where obtained using the XSpectra code of Quantum ESPRESSO.

2.4 X-ray Raman scattering

2.4.1 XRS principles

XRS is defined as the non resonant inelastic X-ray scattering (NIXS) from core-electron excitations. It is described as a two photon process in which an incident photon with energy $\hbar\omega_{in}$, momentum $\hbar\mathbf{k}_{in}$ and polarization $\mathbf{e}_{\mathbf{k}_{out},\lambda}$ is scattered inelastically by a core-electron reducing its energy to $\hbar\omega_{out}$, while changing its momentum and polarization to $\hbar\mathbf{k}_{out}$ and $\mathbf{e}_{\mathbf{k}_{out},\lambda'}$ respectively (see Figure 2.1). During the process, the energy $\hbar\omega$ and the momentum $\hbar\mathbf{q}$ are transferred to the system resulting in a transition from its initial state, $|I\rangle$ with energy E_I, to a final state $|F\rangle$, with energy E_F, satisfying the following energy and momentum conservation relations

$$\hbar\omega = \hbar(\omega_{in} - \omega_{out}), \tag{2.54}$$

$$\hbar\mathbf{q} = \hbar(\mathbf{k}_{in} - \mathbf{k}_{out}). \tag{2.55}$$

Moreover, $\mathbf{q}$ can be expressed in terms of the scattering angle 2Θ as

$$|\mathbf{q}| = \frac{1}{c}\left(\omega_{in}^2 + \omega_{out}^2 - 2\omega_{in}\omega_{out}\cos 2\Theta\right)^{\frac{1}{2}}, \tag{2.56}$$

which can be approximated further if we take into account that in the experiment the incident X-rays are selected such that $\omega_{in} \gg \omega$; then $\omega_{in} \approx \omega_{out}$ leading to

$$|\mathbf{q}| \approx 2\frac{\hbar\omega_{in}}{\hbar c}\sin\Theta. \tag{2.57}$$

This expression gives the absolute value of the momentum transfer in an XRS experiment for a given incident X-ray photon energy $\hbar\omega_{in}$ and scattering angle 2Θ. It arises as a consequence of the high energy constraint imposed on the incident photons, highlighting the bulk sensitivity of the technique.
The corresponding intensity for the scattered photons is proportional to the transition probabilities defined in Equation 2.8, with the transition operator $\hat{O}_{XRS}$

$$w_{if} = \frac{\pi e^2}{\hbar m_e c^2}\left|\langle F|\,\hat{A}^2(\mathbf{r})\,|I\rangle\right|^2 \delta(E_I - E_F). \tag{2.58}$$

Taking into account the definitions of $\hat{\mathbf{A}}(\mathbf{r})$, E_I, E_F and the explicit forms of $|I\rangle$ and $|F\rangle$ for the two-photon process

$$|I\rangle = \left|\phi_i^{KS}, n_{\mathbf{k}_{in},\lambda}, n_{\mathbf{k}_{out},\lambda'}\right\rangle, \tag{2.59}$$

$$|F\rangle = \left|\phi_f^{KS}, n_{\mathbf{k}_{in},\lambda} - 1, n_{\mathbf{k}_{out},\lambda'} + 1\right\rangle, \tag{2.60}$$

Equation 2.58 transforms into

$$w_{if} = C^{\lambda\lambda'}_{\mathbf{k}_{in},\mathbf{k}_{out}}(\mathbf{e}_{\mathbf{k}_{in},\lambda} \cdot \mathbf{e}_{\mathbf{k}_{out},\lambda'}) \left|\left\langle \phi_f^{KS} \right| e^{i\mathbf{q}\cdot\mathbf{r}} \left|\phi_i^{KS}\right\rangle\right|^2 \delta(E_i - E_f + \hbar\omega). \tag{2.61}$$

with

$$C^{\lambda\lambda'}_{\mathbf{k}_{in},\mathbf{k}_{out}} = \frac{\pi e^2}{\hbar m_e c^2} A_{0\mathbf{k}_{in}} A_{0\mathbf{k}_{out}} \sqrt{n_{\mathbf{k}_{in},\lambda}(n_{\mathbf{k}_{in},\lambda}+1)}. \tag{2.62}$$

The transition matrix operator $e^{i\mathbf{q}\cdot\mathbf{r}}$ can be expanded in series as

$$e^{i\mathbf{q}\cdot\mathbf{r}} = 1 + i\mathbf{q}\cdot\mathbf{r} - \frac{(\mathbf{q}\cdot\mathbf{r})^2}{2} + ..., \tag{2.63}$$

and for small momentum transfer it is dominated by the second term in the expansion. In this regime, the transition matrix elements are equivalent to those of XAS in the dipole approximation, with the polarization vector $\mathbf{e}$ replaced by the momentum transfer $\mathbf{q}$. This equivalence between XRS in the low momentum transfer regime, and XAS in the dipole approximation, justifies the use of XAS simulations for describing XRS experiments at small $|\mathbf{q}|$. In practice, the transition matrix operator is expanded in spherical harmonics in order to express the transitions in terms of contributions with a well established symmetry [3]

$$e^{i\mathbf{q}\cdot\mathbf{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l j_l(qr) Y_{lm}(\hat{\mathbf{q}}) Y_{lm}(\hat{\mathbf{r}}), \tag{2.64}$$

where j_l and Y_{lm} are the spherical Bessel functions and spherical harmonics respectively. The arguments of Y_{lm} are unitary vectors for position $\hat{\mathbf{r}}$ and momentum transfer $\hat{\mathbf{q}}$, while the index $l = 0, 1, 2, ..$ labels the monopole, dipole, quadrupole,.. terms of the expansion. We should notice than only the monopole contribution is independent of the direction of $\mathbf{r}$ and $\mathbf{q}$: as a consequence, the angular dependence of XRS can be explained exclusively by non-monopole terms in the expansion. In general, the different contributions dominate the spectrum at specific values of q [111], allowing the use of the transferred momentum as a transition-selector parameter. This characteristic can be leveraged to simplify the spectrum analysis, since for an arbitrary q in addition to the l-isotropic terms we have mixed multipole contributions modifying the peak structure. However, in powder samples, the random orientation of the grains causes the mixed multipole terms to average out, rendering their contribution to the XRS spectrum negligible compared to that of the pure transitions [3]. In Equation 2.64 only the terms with $l + l_i + l_f$ even and $|l_f - l_i| \leq l \leq l_f + l_i$ give non vanishing contributions to the transition matrix, where l_i and l_f are the orbital angular momentum of the initial and final states respectively. This determines the selection rules in XRS for monopole $\Delta l = 0$, dipole $\Delta l = \pm 1$ and quadrupole $\Delta l = \{0\ (l_i = l_f > 0), \pm 2\}$ transitions.

2.4.2 XRS simulations

Since XRS is equivalent to XAS in the low momentum transfer regime, the methods for computing XAS discussed earlier are also valid for XRS. In order to include dipole-forbiden transitions, the existing codes for computing XAS have been extended appropriately. In the independent particle picture, we find real space methods using either multiple-scattering theory within the *muffin-tin* approximation [112, 113] or full-potential DFT [114, 115, 116], and reciprocal space approaches suitable for periodic systems [3]. Beyond the single electron approximation, implementations using BSE [117, 118, 119, 94] are also available. However, the aforementioned methods are usually effective for the initial edges (K, L in less extent) and other approaches like atomic [120, 121] or ligand-field [122, 123] multiplet theories are needed in order to describe properly shallow edges.

In this work we will use the reciprocal-space pseudopotential scheme described in subsection 2.3.2 using the corresponding XRS transition operator as implemented in a modified version of the XSpectra code [3]. Moreover, the energy level reference procedure developed earlier for XAS is also valid for the spectrum alignment in XRS.

Chapter 3

Probing electronic properties of hydrogenated free-standing graphene via XPS

Graphene, unlike other two-dimensional (2D) materials, exhibits remarkable chemical versatility. Its multiple capabilities are extensively harnessed to develop customized properties and functionalities that can differ significantly from the pristine material [124]. Among its chemical derivatives, hydrogenated graphene (H-Gr) is notable for being the simplest and one of the most thoroughly studied [125, 126]. Depending on the hydrogenation conditions and configurations, H-Gr has shown potential in a wide array of applications, including hydrogen storage [127, 128, 129, 130], catalysis [131, 132], opto-electronics [133, 134, 135, 136, 137], and biosensing [138, 139, 140]. Additionally, specific hydrogenation conditions have led to predictions of fascinating optical phenomena, such as excitonic insulating phases [141, 142] and optically controlled exciton traps [143].

A crucial aspect for many of these applications is the creation of a semiconducting gap, which results from the conversion of sp^2 to sp^3 hybridization in Carbon atoms upon Hydrogen adsorption. This gap is highly tunable based on the degree of hydrogenation, potentially leading to insulating behavior in fully hydrogenated graphene, the so-called *graphane* [144, 141], where significant excitonic effects are also anticipated [141, 145]. Small-gap tunability has been demonstrated at low hydrogen coverage due to partial disruption of the π-band at the K-point of the Brillouin zone [146, 147], while recent studies have reported evidence of large gaps in highly hydrogenated samples [148, 136, 149]. However, the experimental measurement of the expected gap intrinsic to high-coverage and fully free-standing H-Gr has remained elusive so far.

Indeed, obtaining the ideal graphane structure [150, 151] with a Hydrogen atom attached to each carbon atom in the 2D hexagonal lattice, has proven to be an exceptionally challenging task. Many experimental observations are dominated by partial, inhomogeneous hydrogenation [152, 153, 154], with defects and

open edges serving as key active sites for H-C interactions [155, 156, 157, 135]. Additionally, hydrogenation is often performed on supported graphene samples [158, 159, 160], hindering the realization of two-sided hydrogenation and resulting in properties that differ significantly from those of a free-standing, fully hydrogenated graphene layer. In addition, most of these works did not attempt to connect the optical or transport measurements to specific structural motifs.

Hydrogenated graphene is thus a relevant prototype system for the application of our computational spectroscopies schemes in synergy with experiments. In the following, I present a computational characterization of the sp^2/ sp^3 hybridization of different configurations of hydrogenated graphene by means of XPS simulations. The aim is to validate the experimental observations of structural changes during the conversion from pristine, free-standing graphene to its hydrogenated phase, as revealed by XPS measurements on nanoporous graphene (NPG) samples. The experiments were performed by the group led by Prof. Maria Grazia Betti and Prof. Carlo Mariani at the Sapienza University in Rome.

The samples consist of a compact, bicontinuous interconnected three-dimensional (3D) graphene arrangement with one to few weakly interacting layers [161, 162]. A nano-porous Ni template was employed to synthesize the material through chemical vapor deposition (CVD). The resulting NPG adopts the three-dimensional morphology of the substrate and is later exfoliated by chemically dissolving the Ni template. This approach enabled the production of a high-quality, nearly defect-free, and free-standing graphane prototype. The hydrogenation (deuteration) was conducted under ultra-high-vacuum (UHV) conditions, where the sample was exposed to atomic hydrogen generated by cracking H_2 in a capillary source heated to 2100 °C. Further details on the sample preparation can be found in Ref. 163.

3.1 Methods

The theoretical predictions of the changes in the XPS spectrum of NPG upon Hydrogen adsorption were obtained by computing the C 1s CLS for prototypical systems, representing different percentages of H adsorbed on graphene. Figure 3.1 shows the DFT-PBE optimized ground-state geometries of the H-Gr structures used in the calculations spanning from one-side (f-g) to two-side (a-d,h) hydrogenation, for single (a,f-h) and double (b-d) layer H-Gr systems. In addition to the pristine graphene (e), *graphane* (a) and one side saturated graphene (f), known as *graphone* [164], were included. In the case of *graphone* we selected the most stable *boat-like* or rectangular, R, configuration. Furthermore, two-side hydrogenation of graphene bilayers (b-d) were included to replicate the diverse local structures found in a turbostratic configuration, as identified through atomic-scale analysis of the NPG microscopic structure [161]. It is worth pointing out that in the experiments the NPG samples were deuterated, while the theoretical models represents hydrogenation. This difference is not relevant for XPS, since the photoelectron

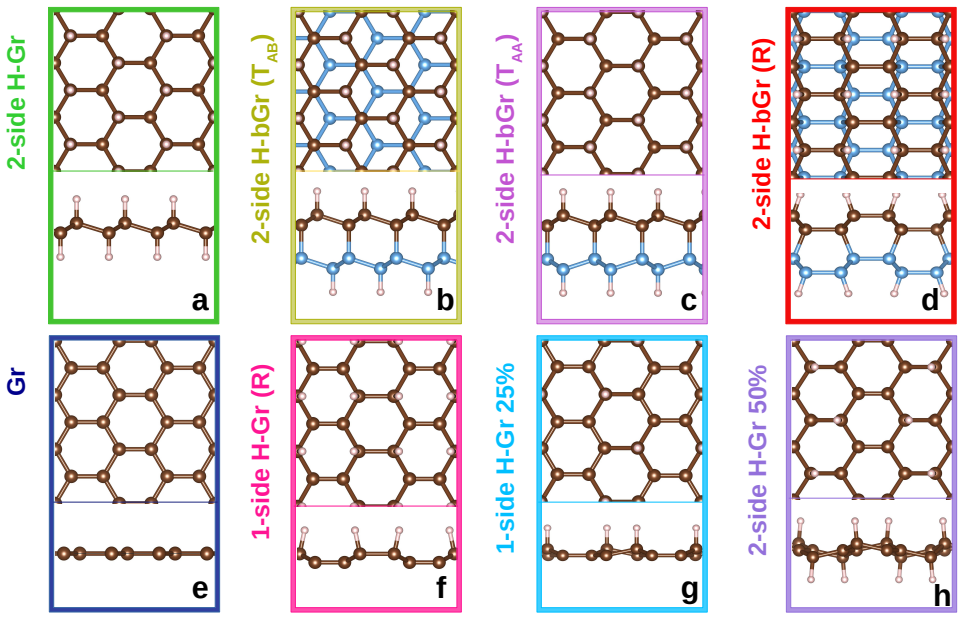

Figure 3.1: Prototypical hydrogenated graphene structures. C (H) is represented by brown (white) spheres. For the bi-layer structures (b-c), the C atoms in the bottom layer are represented in light blue.

emission timescale is much shorter than the corresponding nuclear motion.

Within the DFT-PBE theory, all the systems, except graphene, present a considerably underestimated energy gap [149]. This shortcoming hinders the direct comparison of their computed BEs, as the errors will be, at least, of the same order as the typical CLS between carbon atoms in sp^2 and sp^3 environments. To address this issue, a common reference should be included across all systems, allowing to cancel out the energy alignment errors. With this set up the CLS between a C environment in the system of interest and one in the reference becomes a quantity comparable across different systems. In our particular case, the CO_2 molecule was set as the reference system since CO_x bonds where detected on the experiment. We notice that this procedure is always valid except for the case of 3D systems.

The CLSs between the C 1s in CO_2 and in the prototypical systems were computed using the Quantum ESPRESSO package [36, 37, 38]. The so-called core-hole Delta Kohn-Sham (ΔKS) approach, described in subsubsection 2.2.2.2, was utilized to determine the core electron BE. A supercell approach was adopted to avoid spurious interactions of the C 1s core-hole with its replicas, with supercell sizes set to ensure an in-plane minimum distance of about 14Å, enough to converge the BE values within 100 meV, as shown in Figure 3.2 (bottom-right). In addition, the supercell size in the out of plane direction was set to 24Å, in order to accommodate the reference molecule at a distance from the H-Gr system sufficient to avoid interactions, as presented in Figure 3.2 (left). The periodic supercell in the final

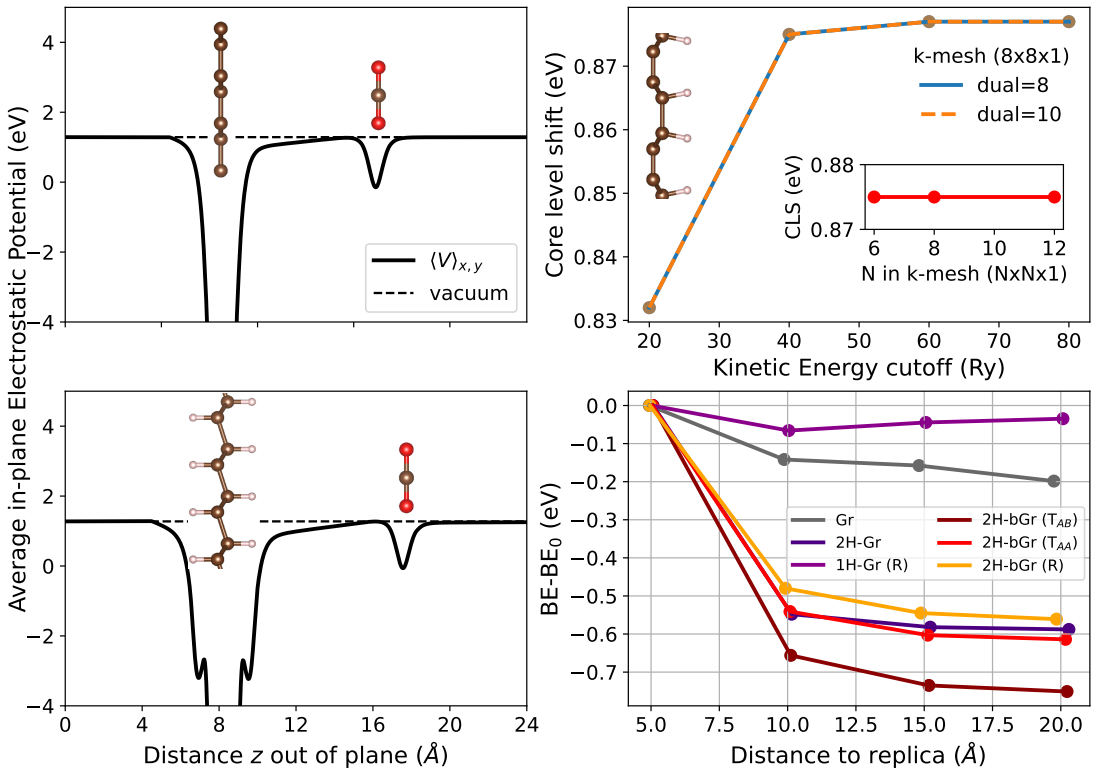

Figure 3.2: Average in-plane electrostatic potential (left) showing the effective vacuum level created by the dipole correction, removing the interaction between the H-Gr+CO_2 system and its periodic image. Convergence results (top right) for the CLS with respect to the kinetic energy (density) cutoff and the k-mesh density. Convergence of the BEs with respect to the minimum distance to the periodic image (bottom right).

state was neutralized according to the XCH treatment. Considering Rappe-Rabe-Kaxiras-Joannopoulos Ultrasoft (RRKJUS) pseudopotentials, the kinetic energy cutoff for the wave functions (charge density) was set to 60 (480) Ry, according to our convergence tests in Figure 3.2 (top-right). The Brillouin zone for the supercell was sampled by using k-points grid spacing of 0.018 $Å^{-1}$. Where present, the artificial electric field across the slab was canceled including a dipole correction [165]. The core level shift of pristine graphene is then aligned with the experimental BE value for pristine NPG in order to have the absolute binding energies for all the simulated systems.

3.2 Results and discussion

The results obtained from the simulations are summarized in Table 3.1. We reported the BE for all the non equivalent C atoms present in the considered

systems calculated with (*) and without dipole correction. In addition, the average C-C bond length, $\langle r_{C_i} \rangle_C$, for the non-equivalent atom i is also presented, ranging from 1.425 to 1.453 Å for sp^2 hybridization and from 1.510 to 1.554 Å for sp^3 hybridization. Our simulations carried on the H-Gr structures corroborate that the C 1s BE for environments in sp^2 are concentrated from 284.53 to 284.74 eV, while for the sp^3 environments they are localized from 285.42 to 286.23 eV. It is worth noting that the dipole correction to the CLS becomes more significant when the C environments are located in H-Gr systems with opposite symmetry relative to the x-y plane, i.e. one-side vs two-side hydrogenation. In such cases, this correction can reach a non negligible contribution to the CLS of ∼0.4 eV for the C environments with sp^3 hybridization.

System	BE_{C_1} (eV)	$BE^*_{C_1}$ (eV)	$\langle r_{C_1} \rangle_C$ (Å)	BE_{C_2} (eV)	$BE^*_{C_2}$ (eV)	$\langle r_{C_2} \rangle_C$ (Å)
Graphene	284.60	284.60	1.425	-	-	-
1H-Gr	284.46	284.54	1.450	285.83	285.94	1.522
2H-Gr	285.63	285.42	1.535	-	-	-
2H-bGr	285.67	285.42	1.541	285.84	285.56	1.554
1H-Gr 25%	284.61	284.74	1.435	286.10	286.23	1.510
2H-Gr 50%	284.61	284.53	1.453	286.06	285.96	1.522
2H-bGr (T_{AA})	285.84	285.52	1.536	286.01	285.67	1.548
2H-bGr (T_{AB})	285.83	285.53	1.539	286.00	285.68	1.543
CO_2	286.80	286.82	-	-	-	-

Table 3.1: Computed BEs for all the non equivalent C atoms present in the considered systems calculated with (*) and without dipole correction. In addition, we reported the average C-C bond length, $\langle r_{C_i} \rangle_C$, for the non-equivalent atoms $i = 1, 2$.

The quantification of the atomic deuterium (D) uptake was carried out by our experimental collaborators using surface-sensitive C 1s core-level photoemission spectroscopy. Figure 3.3 shows the C 1s core level spectra acquired at the clean (left) and deuterated (right) NPG samples, as compared to the DFT predictions for different H-Gr configurations (vertical lines). For clean pristine NPG, the C 1s core level spectrum presents a dominant peak at 284.6 eV binding energy (BE), which is associated to the semimetallic graphene signal with a planar sp^2 hybridization, in agreement with previous results [161, 162]. The presence of a tiny sp^3 component at 285.2 eV is associated to the distortion of the planar Gr structure due to the tubular wrinkled portion of NPG [161, 162].

In the deuterated sample, atomic deuterium uptake is revealed by the presence of an increased sp^3 component with respect to the pristine planar C-C sp^2 configuration, as can be deduced by the C 1s core level fitting [1], where the distortion

[1]Details on the fitting procedure are described in Ref. 166

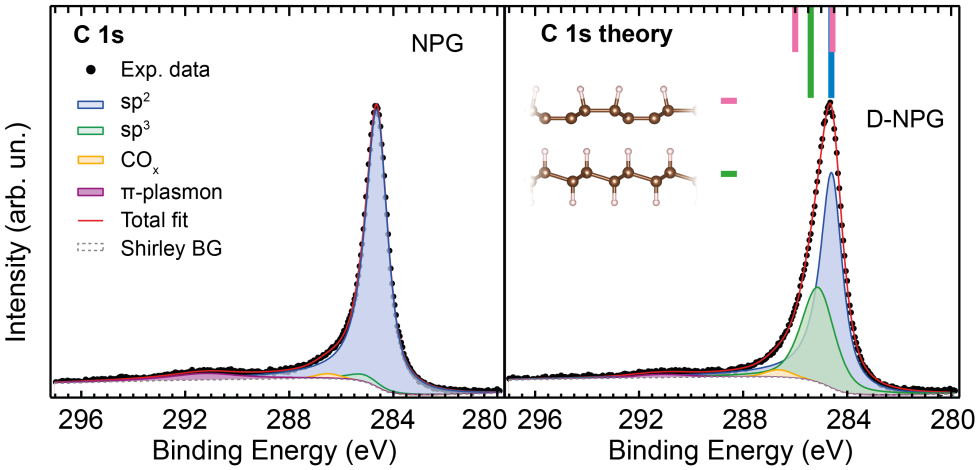

Figure 3.3: C 1s core level XPS spectra of UHV-clean NPG (left) and D-NPG (right). Experimental data (dots), sp^2 (blue peaks) and sp^3 (green peaks) fitting components, C-O_x component (yellow peak), π plasmon (violet peak), total fit (red line), and Shirley background (white) are displayed; in the inset to the right panel, simulated core-level binding energies for the H–C configurations depicted in the inset are indicated by vertical bars, as referenced to pristine Gr (blue bar).

of the sp^2 configuration towards an sp^3 hybridization is due to the chemisorption of D adatoms on top of the C sites [151, 149]. At the same time, it was possible to observe an intensity reduction of the intrinsic plasmon associated to the π collective excitations at about 291.1 eV after deuteration, attributed to the frustrated collective π plasmon modes associated with the conversion from the pristine semimetallic to a semiconducting phase [167]. The deuteration process is very clean, as only a very tiny component due to residual CO_x bonds ($< 2\%$) is appreciated at about 286.5 eV BE. Moreover, the measured C 1s spectra do not unveil any contributions at lower BE with respect to the sp^2 peak due to unsaturated C dangling bonds at vacancy sites[168] both before and after deuteration, indicating that the D chemisorption process is not inducing any damage in the NPG crystal lattice.

The increased intensity of the sp^3 component with respect to the graphene sp^2 planar configuration allowed the evaluation of the fraction of D atoms bound to C atoms in the NPG mesh. By considering the relative intensity ratio $\Theta = I(sp^3)/[I(sp^2) + I(sp^3)]$ the Θ value increases from less than 4 at.% (on the pristine NPG) to about 36 at. % for D-NPG. The persistence of a sp^2 component in the D-NPG sample is due to the presence of only a fraction of deuterium atoms in four-fold coordination with the C atoms in the graphene mesh, consistent with the permanence of the Raman G and 2D bands, which should be quenched upon deuterium adsorption due to in-plane sp^2 translation symmetry breaking.

The corresponding BEs, obtained by taking the experimental pristine Gr com-

ponent as a reference (284.6 eV, blue bar) across all computed CLSs, are depicted as vertical bars above the C 1s data of the deuterated sample (Figure 3.3 right) for 50 (*graphone*) and 100 (*graphane*)% H–C configurations (see inset; configurations are labelled according to the bar color code). The bar height indicates the relative weight of each component for a given H–C configuration. We find that the 50% single-side H–C configuration presents two inequivalent C atoms giving rise to two CL shifted components (pink bars), at -0.1 eV (i.e. 284.5 eV BE) and 1.3 eV (i.e. 285.9 eV BE), associated to sp^2 three-fold and sp^3 four-fold C coordination, respectively. The 100% 2-side H–C configuration (green bar) presents a single sp^3 component at 285.4 eV (0.8 eV CLS), in much better agreement with the C 1s experimental value for the sp^3 component (285.2 eV), as deduced by the fitting procedure.

3.3 Conclusions

The simulations provided in this study contributes to our understanding of the electronic and structural properties of hydrogenated free-standing graphene, as probed by XPS. Through the analysis of the BEs of different C environments present in a set of H-Gr prototypical structures, we confirmed the relation between the sp^2/sp^3 hybridization and the resulting core level shifts. These theoretical findings are crucial for interpreting the experimental XPS data, particularly in validating the structural transformations observed during the hydrogenation process which translates in changes in the corresponding electronic properties such as the gap opening. The inclusion of a common reference environment and the proper treatment of the long range electrostatic interaction, through a dipole correction, establishes a robust framework for comparing different hydrogenation states, thereby offering a more precise characterization of the NPG samples.

Chapter 4

XRS study of Si nanoparticle anodes

Lithium-ion batteries (LIB) degrade over time due to various physical and chemical phenomena, leading to irreversible capacity loss. Mitigating this aging process is crucial to extend the lifespan of electrochemical cells, meeting societal demands for a more sustainable energy production and storage chain. These requirements add on top of the quest for high-energy-density and cost-effective batteries, which is key for the success of electric vehicles and large grid storage.

Achieving these ambitious goals in terms of performance and sustainability requires significant advancements in materials chemistry, component manufacturing, and cell fabrication, especially for state-of-the-art technologies like silicon-graphite/NMC (or LFP) cells [169, 170], which currently dominate the market. In these cells, silicon is integrated in graphite-based composites to improve performances, given its ten times higher theoretical gravimetric capacity [171, 172, 173]. However, the amount of Si has generally to be limited to few weight percent (wt%) to avoid quick performance decay due to loss of active material, loss of cyclable lithium and other degradation mechanisms [174]. Indeed, silicon suffers from severe capacity fading issues connected to the drastic volume expansion experienced when alloying with Li [175, 176, 177], and the formation of a dynamic solid electrolyte interphase (SEI) [176, 177].

The SEI is a complex porous layer formed at the anode (or cathode, the so-called CEI) from the decomposition of carbonate-based liquid electrolytes and inorganic salts like $LiPF_6$ [178, 179]. It can play a beneficial role, acting as a passivation layer that can prevent further decomposition and structural aging of active materials [180, 181, 182, 183], but it can also obstruct diffusion pathways, reduce lithium-ion availability, and cause detrimental polarization [176]. In Silicon-based materials, the SEI building is responsible for a large irreversibility at first cycle. Moreover, as this SEI continuously evolves during further swelling/shrinking of the silicon material [176, 177], it contributes to additional particle cracking, pulverization and disconnections [184, 185, 186] that reduce the amounts of both

active material and lithium available for cycling, with an increased amount of lithium ions being trapped in the SEI and in the particles [187].

Mitigating these effects can be achieved by using amorphous, nanoporous and/or nanostructured Si [188, 189, 190, 191, 192], together with the usage of polymeric binders [193, 194] and the design of a stable SEI, which is usually obtained by the use of additives as fluoroethylene carbonate (FEC) and vinylene carbonate (VC). [195, 196, 197, 176, 198, 199] Other directions to prevent lithium loss is to use prelithiation [200] or to blend silicon with another group IV material [201].

Nevertheless, effective control-by-design strategies are limited by the difficulty to understand the origin of lithium loss in silicon compounds. This is partly due to the difficulty to probe the SEI characteristics as a function of electrode/electrolyte composition and cycling parameters, as well as to localize and quantify lithium trapped in the cycled structures at particle, electrode and cell scales [202].

Regarding the SEI, its interfacial nature, complex composition, variable thickness and evolution over cycling, represent a true challenge for experimental characterization. Techniques such as nuclear magnetic resonance (NMR) [176, 203, 204], X-ray photoelectron spectroscopy (XPS) [205, 206, 207, 208, 209], Raman spectroscopy [210], Fourier transform infrared spectroscopy (FTIR) [211], and scanning transmission electron microscopy with energy-loss spectroscopy (STEM-EELS) [212] have provided insights into the SEI composition, primarily through post-mortem analysis. It is now widely recognized that the SEI is mainly composed of Li_2CO_3, LiF, oligomer species and semicarbonates [213], in addition to other (in)organic compounds specific to the electrode and electrolyte composition, as well as to the usage of additives [205, 199, 176, 198]. Despite the significant amount of studies devoted to the SEI, fundamental questions remain regarding the organic vs inorganic compounds distribution, as well as their relative role and dynamic evolution during cycling. For instance, some authors found that LiF is quite stable over time, forming thin film patches that cover the surface of silicon particles [187], while others report that LiF is dissolved partially during cycling [214]. Nevertheless, many of these studies are limited to surface analysis (typically, probing few nanometres) and can be affected by surface contamination or intrusive sample preparation, when performed *ex situ*. Alternatively, operando characterizations are attractive because they enable the comparison between different states-of-charge of the same sample without dismounting the cells and manipulating the electrodes, using for instance standard spectroscopies as XAS or NMR, or reflectivity techniques [215], but they are not often conclusive or representative due to specific set-ups, beam damage issues, or lack of quantification[216, 217].

Beyond SEI studies, accessing lithium inventory in a cycling battery is also difficult. Combined synchrotron experiments coupling X-ray diffraction and X-ray imaging [218], or X-ray diffraction and X-ray fluorescence [219], have attempted to identify and quantify the presence and distribution of LixSi phases (x being the lithiation index) from analysing spectral signatures, combining it with total

lithium content analysis to deconvolute the various degradation processes leading to lithium loss. Nevertheless, these studies remain scarce and specific, calling for the need of dedicated bulk-level chemical-sensitive investigation in model systems to establish the nature and proportions of ions lost in silicon, particularly during the first cycle where the SEI forms.

Here we fill this gap by applying synchrotron X-ray Raman Scattering (XRS) measurements and multi-edge analysis via a combined experimental-numerical approach. Indeed, XRS offers a promising alternative to evaluate both the mechanism of lithium loss in silicon anodes and the SEI characteristics, at it provides information on the bulk electronic structures and chemical environments of materials [220, 58]. Unlike absorption techniques, XRS is a non-resonant technique that uses highly penetrating hard X-rays, which are inelastically scattered by the material allowing access to soft X-ray edges in the bulk. This enables detailed analysis of light elements like Li, C, O, and F, which are crucial components of the SEI, as well as access to Si edges (both K and $L_{2,3}$), hence providing information on silicon alloys and oxides which are formed and evolve within the electrode upon cycling.

To date, XRS has been used in battery research to study chemical environments in cathode materials and graphite [221, 222, 223, 224, 58, 225, 226, 227]. For instance, in combination with XAS measurements, XRS has successfully revealed changes in the oxidation state and lithiation level at the electrode surface and within the bulk for $LiMn_2O_4$ [228] and $LiFePO_4$ [229], respectively. Nevertheless, while XRS shows great potential, this technique is still in its infancy for applications to battery materials, due to several factors: i) quantitative analysis remains challenging due to limited reference spectra as compared to more established XAS or XPS [58] ; ii) the analysis of multi-component spectra requires DFT-aided computations to disentangle contributions in complex environments, and iii) the long measurement times required to optimize the signal-to-noise ratio make operando studies limited to restricted configurations and conditions at present [220], with real operando cells behavior remaining beyond reach for the moment. Therefore, detailed analysis of redox or degradation reactions in controlled and representative conditions by XRS require substantial improvements at the level of experimental capabilities, database construction, and modelling.

In this work, we tackle challenges i) and ii) by developing a multi-edge combined experimental-computational approach to analyse *ex situ* X-ray Raman Scattering of silicon-based anodes. Series of high-resolution synchrotron XRS data were collected by our experimental collaborators at the Li, C, O, F and Si edges in a benchmark silicon-nanoparticles pristine anode, as well as in its lithiated and delithiated states during the first cycle. XRS spectra of several key reference compounds are acquired and simulated to serve as benchmarks for data analysis. We develop the methodology, schematized in Figure 4.1, to investigate the multi-component SEI layer composition and quantify lithium loss by using computed model spectra to support the linear deconvolution of the XRS data. With

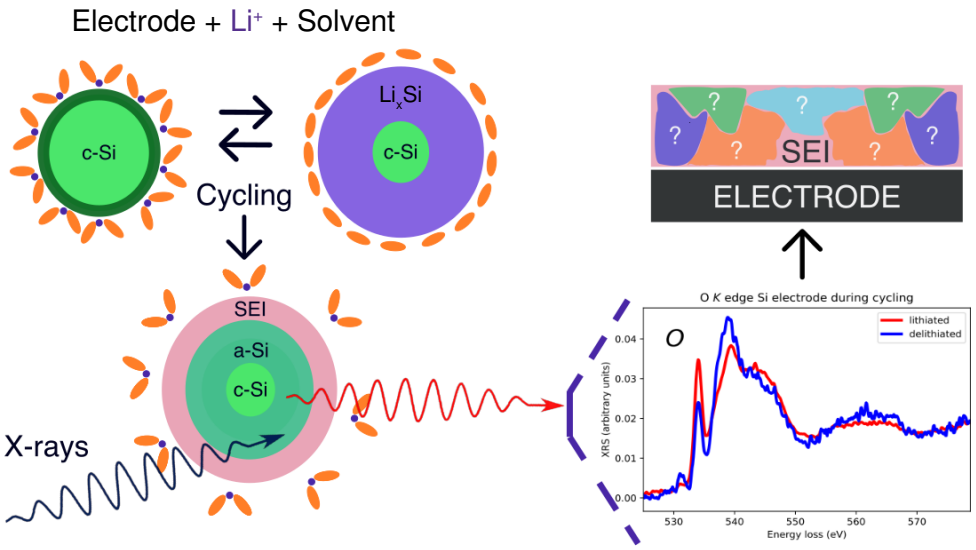

Figure 4.1: Schematic representation of the proposed methodology. The Si-Nps pristine electrode undergoes through chemical and structural changes after one cycle of charge. Such changes are probed with XRS at two different SoC. Using a multi-edge analysis combining experimental and theoretical references, the corresponding spectra is interpreted in order to reveal the compounds present in either the electrode and the SEI.

this approach, we identify the main SEI species, isolate their relative contributions, and quantitatively evaluate organic and inorganic products, such as Li_2CO_3 and LiF. Additionally, we analyse the Si environment and detect the presence of residual Li_xSi phases after delithiation, due to some trapped lithium, as confirmed by Li edge data. We evaluate the lithium lost in silicon particles and in the SEI. These results pave the way to systematic, reference data-informed, and modelling-assisted studies of SEI and CEI characteristics in the bulk of electrodes prepared under controlled state-of-charge and state-of-health conditions, as well as a better understanding of the interplay between SEI building and incomplete delithiation mechanisms during the first irreversible charge-discharge process. I contributed to this work by conducting the *ab initio* simulations of XAS/XRS spectra, performing the linear combination procedure, carrying out the quantification processes combining electrochemical cycling data and XRS measurements, and analyzing and interpreting the spectra.

4.1 Results and discussion

Our experimental collaborators used a prototypical anode, i.e., a silicon electrode composed of crystalline nanoparticles synthesized by laser pyrolysis [230] mixed

with a binder (sodium carboxymethyl cellulose, Na-CMC) and a conductive agent (carbon black, hereafter labelled Super P or SP). Details on the materials are provided in the Methods. The electrodes were cycled in half-cell using a standard carbonate-based electrolyte with $LiPF_6$ salt and vinyl carbonate (VC) additive. Samples were prepared in two states of charge (SoC) during the first charge-discharge cycle performed at C/20: fully lithiated (SoC = 100) and fully delithiated (SoC = 0) (see colored symbols in the galvanostatic data reported in Figure 4.2). Without any prior rinsing, the samples were measured by XRS at the ESRF synchrotron using a custom air-tight cell, together with the pristine electrode and pristine powder materials. A full energy scan was performed for each sample from 0 to 800 eV, followed by high resolution scans taken at all relevant edges, i.e., Li, C, O, F K-edge and Si $L_{2,3}$-edge. Details on set-up, cell, measurement protocols, XRS data reduction and corrections, are given in the Method section. The results were analysed by comparing the experimental spectra of each edge at two SoC to reference spectra acquired on purpose in the same conditions and on the same set-up, considering selected compounds of interest. The data are all represented on Figure 6.10, where top panels show fully lithiated (red) and delithiated (blue) electrodes data, and model compounds are plotted in the bottom panels. It is clear that the shapes and intensities of the spectra are affected by the SoC, some edges being more altered than others. Moreover, specific features can be visually recognized by comparing the silicon electrode data to the model materials, e.g., for instance specific low-energy C-fingerprint of Li_2CO_3, highlighting the interest of the database-informed strategy. In addition to this, modelled XRS spectra (see Method section) are also used to support the attribution of the observed changes to specific species and products, as will be discussed in details in the various sections.

In the following, we first show the capacity loss of the first charge-discharge cycle measured by means of standard galvanostatic measurements (section 4.1.1). Next, we describe the C, O and F edges, which contain information on the main organic and inorganic SEI products (section 4.1.2), as well as the Si edge data, in order to compare the presence of silicon-based alloys in the two SoC (section 4.1.4). Finally, we report on the Li edge and discuss the complex chemical environments present in the electrodes and the implications of the observed energy shifts and profiles variations in terms of lithium inventory (section 4.1.5). This multi-edge analysis enables us to integrate all numerical and experimental observations into a model of the lithium loss analysis, quantifying the different populations of lithium chemical environment.

4.1.1 Galvanostatic cycling measurement

Figure 4.2 shows the first electrochemical cycle during lithiation up to 0.01 V versus Li metal in a coin cell. The electrochemical curve features a plateau at 0.08 V, similar to that previously reported for c-Si NPs [231, 232, 233]. The discharge

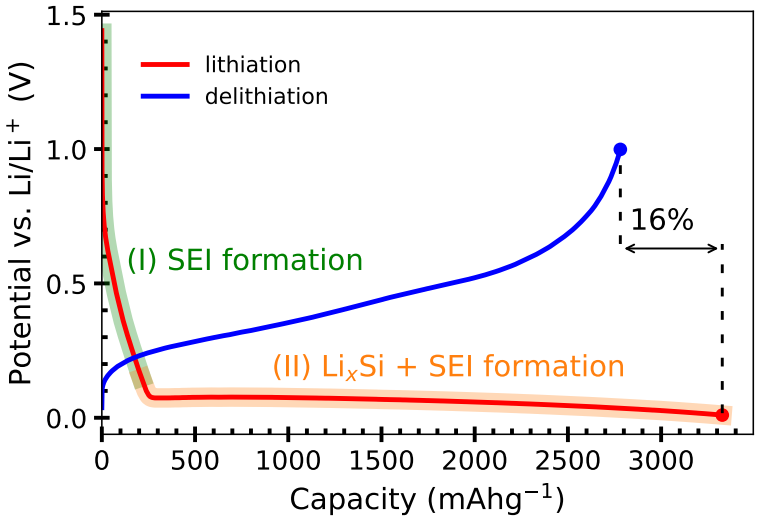

Figure 4.2: Potential vs. Li/Li^+ against capacity during the first electrochemical cycle in a coin cell at C/20 for crystalline Si-NPs based anodes. The electrode SOC analyzed post-mortem by XRS are indicated by red/blue dots for the fully lithiated/delithiated state, respectively.

capacity obtained is 3328 $mAhg^{-1}$. During delithiation, the obtained capacity reaches 2781 $mAhg^{-1}$ with a Coulombic efficiency of 84% and an irreversible capacity loss of 16%, in agreement with previous results on c-Si NPs [234, 207].

The capacity loss measured during the first cycle is different from any other, and usually related to the initial SEI formation observed for Si-based electrodes [207, 235, 187]. Indeed, during the first lithiation, two regions can be identified on the galvanostatic curve (marked in green and orange, respectively). For potentials above $\sim$ 150 mV (green), the irreversible reduction of solvents and salts in the electrolyte is known to take place, resulting in the formation of an initial, highly porous SEI, primarily composed of fluorinated compounds [236, 237, 214]. Below this lithiation potential (orange), Si lithiation starts to occur [232, 238, 12] while the SEI evolution continues, with a major contribution from Li_2CO_3 [236, 237]. Following delithiation, part of the SEI decomposes and some amount of Li remains trapped inside the Si-NPs [239, 240, 241].

4.1.2 C, O and F K-edge: tracking the early SEI formation

As said, the post-mortem XRS measurements after one cycle were made for two different SoC identified by the red (lithiated) and blue (delithiated) dot in Figure 4.2. Figure 4.3 (white region) shows the C, O and F K-edge spectra normalized to unit area of electrode samples at the end of the first charge (lithiation, top) and the first discharge (delithiation, bottom) cycles.

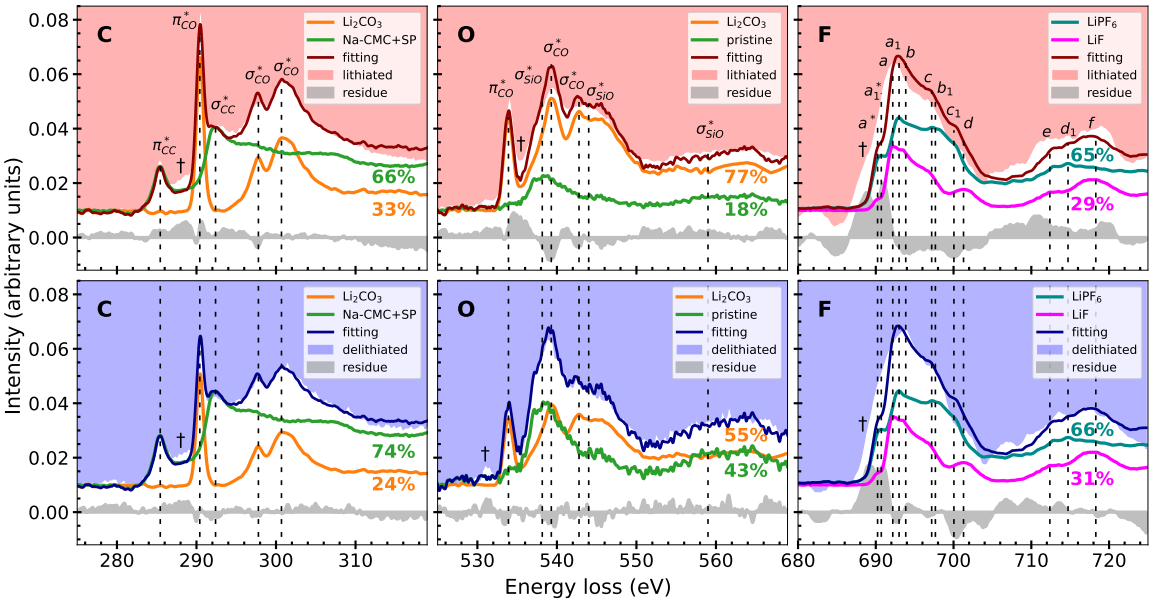

Figure 4.3: Experimental non-resonant inelastic X-ray scattering spectra of c-Si NPs electrodes in the low momentum transfer regime compared with the linear combination of selected references for the C K-edge (left), O K-edge (center) and F K-edge (right) in the lithiated (top) and delithiated (bottom) states.

We here focus on a semi-quantitative analysis of the XRS spectra in Figure 4.3 by modelling each SoC spectrum, at each given edge, with the best fit (dark red/blue curve for the lithiated/delithiated state) obtained from the linear combination (LC) of relevant experimental reference spectra (other colored solid curves) acquired in the same conditions (description of the fitting can be found in subsection 4.2.4). The weighted experimental references (see indicated %), with their principal features and the residual curve after fitting (grey area), are displayed to provide a comprehensive description. The reference compounds were selected according to the main decomposition products of the electrolyte in LIB (i.e., Li_2CO_3, $LiPF_6$ and LiF) and the initial constituents of the pristine electrode (Na-CMC + Super P and Si-NPs).

Overall, the spectra of the Si NPs electrodes at different SOC are fairly well described by the LC model. In particular, only minor differences, accounted by the residual curve, are detected for the C and O (delithiated) K-edges. The energy regions where the fitted LC model deviates appreciably from the measurements (†) reflect the incompleteness of the basis expansion. Deviations are more prominently observed in the O (lithiated) and F K-edges, and will be addressed in the following by resorting to density-functional theory (DFT) based computational modelling.

F K-edge. We start by analyzing the F edge (Figure 4.3, right panels), reminding that the pristine electrode is F-free. We notice that, for both SoC, the best fit to the XRS spectrum requires the presence of both references, i.e., $LiPF_6$ and

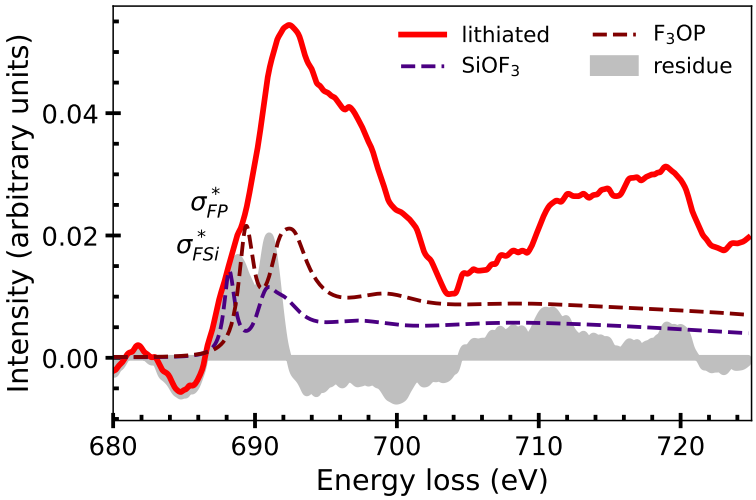

Figure 4.4: XRS F K-edge spectrum of c-Si NPs electrodes in the low momentum transfer regime. The experimental curve in the lithiated state (red line) is compared to selected theoretical references (dashed lines). The residual curve from figure 4.3 is also reported in grey.

LiF, confirming the formation of LiF already during the first cycle [176]. Whereas the relative composition obtained by the fitting remains nearly unchanged (see %) for the two SoC, we notice an upward shift in energy of 1.5 eV for the edge maximum located at 692.4 eV when the cell is delithiated. Moreover, we find for both SoC a lower energy onset at about 689 eV (†) as compared to the selected references, suggesting the presence of additional F-bearing compounds.

A previous work detected PO_xF_y and SiO_xF_y on similar systems using *ex-situ* multi-nuclear solid-state NMR ^{19}F spectra and Cross Polarization (CP) experiments [176]. These species are likely to form after the reaction of $LiPF_6$ with H_2O, and later reaction with the Si NPs surface [242, 208]. In order to explore the contribution of such species to the overall spectrum, theoretical F K-edge XAS spectra were computed for similar prototypical systems, i.e., POF_3 and F_3OSi, as reported in Figure 4.4 (dashed lines) together with the experimental lithiated curve (red solid line) and the relative residue curve (grey area).

These results suggest that the low energy feature in the electrode spectrum could indeed arise from transitions between $1s$ states and σ^* orbitals in F−P and F−Si environments, similar to those found in PO_xF_y and SiO_xF_y species, as they match fairly well with the residue curve (grey area) in this energy region. Indeed, these contributions, though minor, could solve both the model underestimation of the intensity at low energies and the overestimation between 692 and 704 eV, which both result from the absence of reference contribution in the onset energy region.

O K-edge. Moving to the O K-edge (Figure 4.3, central panels), we observe

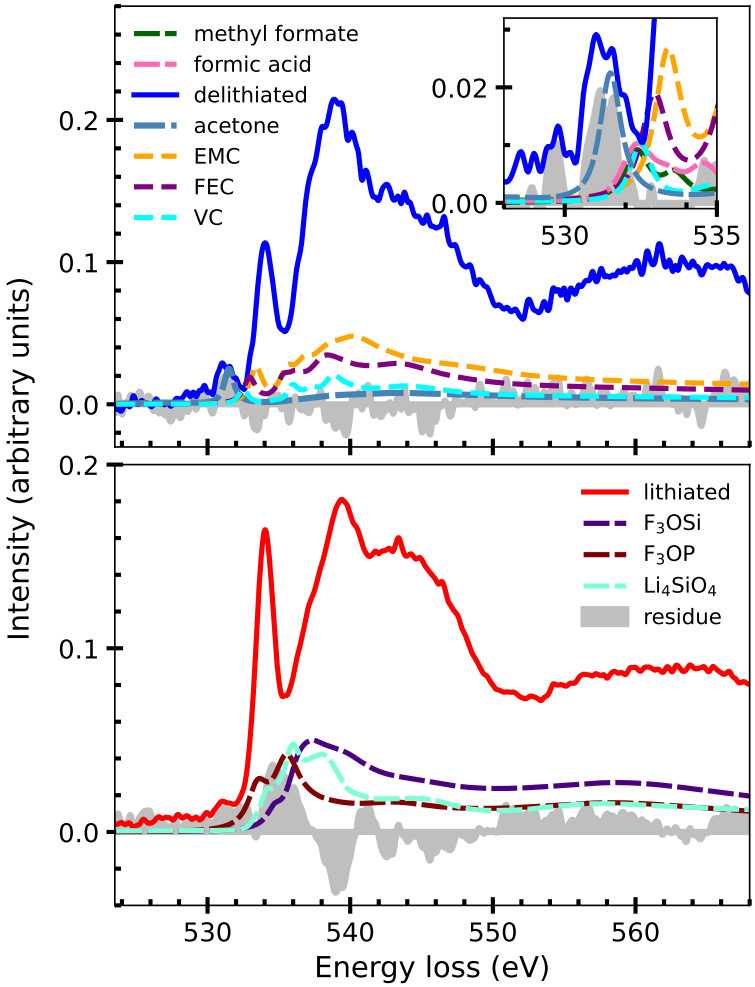

Figure 4.5: XRS O K-edge spectrum of c-Si NPs electrodes in the low momentum transfer regime. The experimental curves in the delithiated (top, blue line) and lithiated (bottom, red line) states are compared with organic (top) and inorganic (bottom) theoretical references (dashed lines). The residual curve from figure 4.3 is also reported in grey.

that the XRS spectrum departs considerably from the one of the pristine electrode (green curve) already after the first charge cycle, with a shape very much resembling that of the Li_2CO_3 reference (orange curve). This is confirmed by the best fit model, which identifies Li_2CO_3 as the main O-bearing compound in the SEI for the lithiated state (77%). Upon delithiation, we notice an intensity decrease for the peak located at 533.9 eV, which suggests a partial decomposition of the Li_2CO_3 present in the SEI. This is accompanied by an increase of the contribution due to the pristine electrode reference in the linear combination model (from 18 to 43%). Considering that the binder and the natural oxidation layer of Si NPs (SiO_2) are the main oxygenated species in the pristine electrode, both having similar O K-edge shape (see Figure 4.7 (right)), and that the binder can be considered as fixed in the first cycles, we could possibly associate the higher pristine electrode weight with a higher oxidation level of the Si NPs.

The main features not captured by the model (†) are the peak at about 531 eV (both SoC, but more prominent in the delithiated state) and the underestimation (overestimation) around 534-536 eV (above 538 eV) characterizing the lithiated state. The XRS data are compared to the computed theoretical references in Figure 4.5, split in organic (top) and inorganic (bottom) contributions. For the organic part, we considered the simplest ketone, ester and carboxyl-based compounds, to try to associate the spectral contributions to specific local chemical environments. It is clear from the inset of the top panel that the contribution of the organic species to the spectrum should be minor and localized in the 530-533 eV region, with the ketone group accounting for the feature at about 531 eV.

The inorganic compounds used as references are POF_3 and F_3OSi, as resulting from the F K-edge analysis, and Li_4SiO_4, as a prototype of Li-silicates expected to form upon lithiation of the SiO_2 natural layer of Si NPs [208]. Their contribution to the overall spectrum is distributed in a wider energy range than that of the organic ones, mostly overlapping with the underestimation/overestimation of the model in the lithiated state described above. In particular, the depth change for the dip at 535 eV could be associated to the formation of lithium silicates and PO_xF_y species. Overall, the contribution of other inorganic compounds remains relatively small when compared to Li_2CO_3, which confirms the picture drawn from the F K-edge analysis.

C K-edge. Finally, the C K-edge (Figure 4.3, left top panel) reveals the formation of the carbonaceous portion of the SEI during the initial charge cycle and its evolution upon discharge. For both SOC, the linear combination analysis indicates the presence of a variable mixture of pristine electrode (Na-CMC+SP) and Li_2CO_3, identifying the latter as the principal C-bearing compound in the SEI. The accuracy of the fitting indicates only a minor missing contribution in the 285-290 eV energy region, which can be attributed to organic species from the electrolyte according to DFT simulations, as detailed in Figure 4.6. Notably, upon delithiation, the quantity of Li_2CO_3 is reduced, confirming its partial decomposition, as indicated by the analysis of the O K-edge. This finding agrees with

previous observations by Philippe et al. [208], which reported variable-thickness SEI by monitoring the C 1s peak by soft and hard XPS during the first cycle.

4.1.3 Quantification of the initial SiO_2 content

Figure 4.7 (left) displays the XRS Si $L_{2,3}$-edge spectrum of the Si powder used in electrode preparation (violet curve), the pristine c-Si NPs electrode (green curve), as well as a reference spectrum for SiO_2, as taken from Ref. 4 (dashed brown curve). In order to quantify the amount of c-Si reverting to SiO_2 during the preparation procedure, we consider the $Si_{pristine}$ area, which accounts for the number of transitions in the energy range up to the first feature of SiO_2 for the pristine electrode, and the ΔSi area, which corresponds to the difference in number of transitions in the energy range up to the first feature of SiO_2 between the Si powder and the pristine electrode. The area ratio $\frac{\Delta Si}{\Delta Si + Si_{pristine}}$ quantifies the relative amount of c-Si that is oxidized during the preparation process.

It is also worth noting that, for this quantification, it is worth referring to the Si $L_{2,3}$ edge, where the contribution from c-Si and SiO_2 are neatly separated. Any further consideration of this sort on other edges has to be performed with care, as spectral contributions from different compounds are more overlapping and/or very similar. This is the case of the O K-edge, as exemplified in Figure 4.7 (right).

4.1.4 Si-NPs evolution

After discussing the SEI formation and evolution during the first charge and discharge cycles, we next analyse the evolution of Si NPs by tracking the changes with respect to the pristine signal on the Si $L_{2,3}$ edge. Figure 4.8 (top) shows the Si $L_{2,3}$ edge spectra of the electrode in pristine (green), lithiated (red), and delithiated (blue) conditions. In order to support the analysis, we include an experimental reference for SiO_2 [4] and a group of theoretical references (bottom panel) for c-Si, $Li_{15}Si_4$, F_3OSi and Li_4SiO_4. The spectrum for the pristine electrode (green) indicates primarily the presence of c-Si (≈94%), with a minor amount of SiO_2 (≈6%) as detailed in subsection 4.1.3, created during the electrode preparation. This maps directly to the NP morphology, which is composed by a crystalline core surrounded by a natural SiO_2 shell, as identified by the peak at 101 and by the features at 106 and 108.3 eV, respectively.

Upon charge, the features corresponding to c-Si and SiO_2 are severely dampened, with a corresponding increase of the signal below 100 eV and between 102 and 107 eV. The very same modifications are observed when comparing the computed spectrum for c-Si and a prototypical Li-silicide, i.e., $Li_{15}Si_4$. Notably, the close resemblance of the computed $Li_{15}Si_4$ to the experimental lithiated (see Figure 4.9 top left) spectrum suggests the lithiation of a major part of the NP volume, leaving only a relatively small fraction of c-Si, as we can reproduce in the LC presented in the Figure 4.9 (top right). Indeed, the lithiation of Si NPs is usually

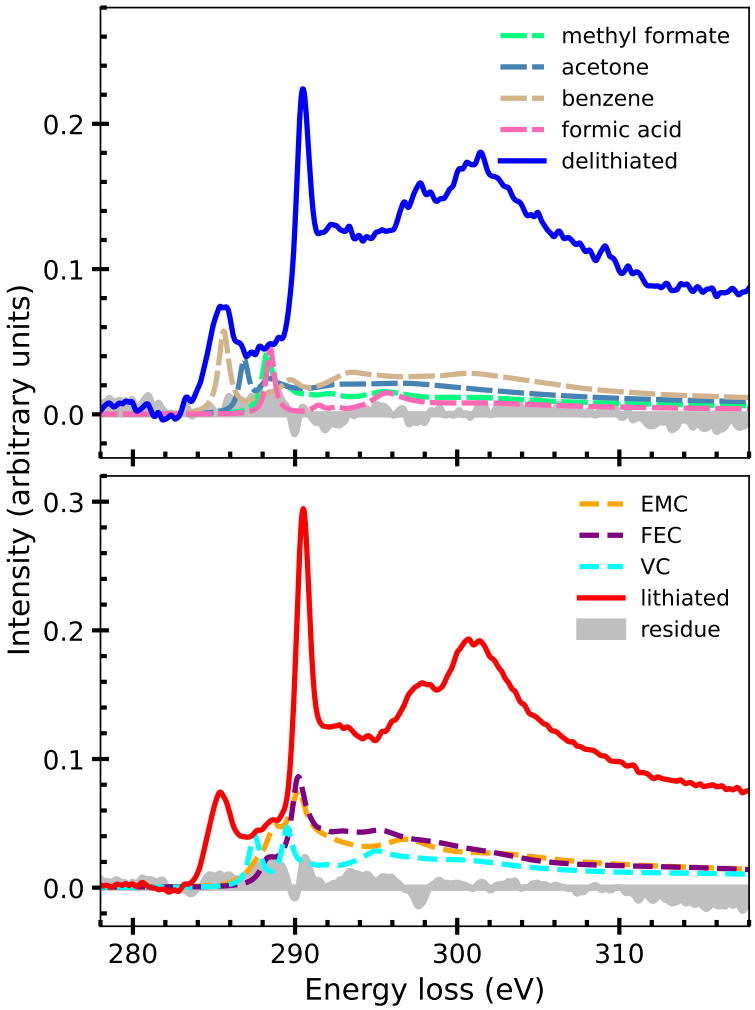

Figure 4.6: XRS C K-edge spectrum of c-Si NPs electrodes in the low momentum transfer regime. The experimental curves in the delithiated (top, blue line) and lithiated (bottom, red line) states are compared with organic theoretical references (dashed lines). The residual curve from Figure 2 of the main text is also reported in grey.

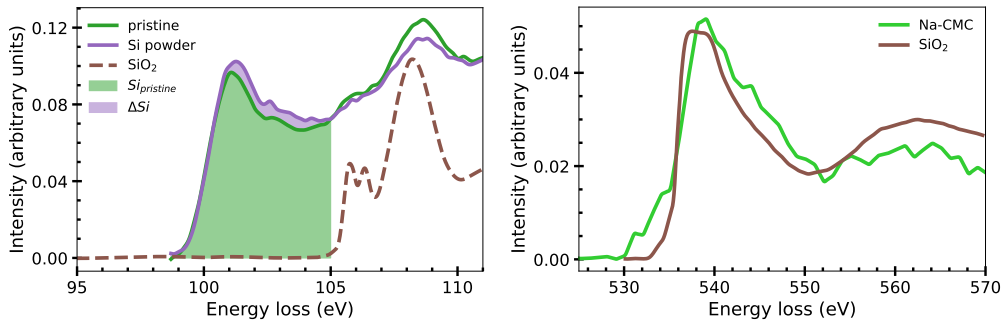

Figure 4.7: (Left) XRS Si $L_{2,3}$-edge spectrum of pristine c-Si NPs electrode, Si powder used in electrode preparation and SiO_2 from Ref. 4 (dashed lines). (Right) XRS O K-edge spectrum of Na-CMC in the low momentum transfer regime compared with SiO_2 XANES from Ref. 8.

characterized by the coexistence of a crystalline core with an outer shell highly lithiated and amorphized [243].

Upon discharge, the electrode spectrum features an attenuation of the peak at 101 eV, which could be attributed to the partial amorphization of the c-Si NPs after lithiation and delithiation, with the resulting spectrum containing characteristics of both crystalline and amorphous Si phases. Moreover, we observe the persistence of the lower onset energy (below 100 eV) found for the lithiated state, which suggests that a small quantity of Li is trapped in the NP in the form of Li_xSi. Finally, the peaks located at 106 and 108.3 eV display a higher intensity as compared to the spectrum of the pristine sample. This may be related to an increased amount of oxidized Si, as already evidenced on the O K-edge. This could be possibly produced by a native oxide regrowth [244], as we cannot exclude a short oxygen exposure in the glovebox from the sample preparation. Minor contributions from SiO_xF_y and Li_xSiO_y species are also not discarded since their main features overlap with the peak structure attributed to SiO_2, according to DFT simulations. This picture is supported qualitatively by the LC presented in Figure 4.9 (bottom right).

4.1.5 Li loss analysis

To conclude the qualitative analysis, the Li K-edge spectra of c-Si NP electrodes in the high momentum transfer regime are shown in Figure 4.10 (top) for both lithiated (red) and delithiated (blue) conditions. These curves are compared with experimental references for Li_2CO_3, LiF and $LiPF_6$ (top), as well as theoretical references for Li_xSi and Li_4SiO_4 (bottom). After the initial lithiation, the electrode very much resembles Li_2CO_3, with relatively minor amounts of LiF and $LiPF_6$ localized both at 61 eV and 69 eV. However, the LC model based on these three compounds cannot reproduce neither the lower energy onset nor the increased

intensity between 62 and 67.5 eV with respect to Li_2CO_3, as can be appreciated in Figure 4.11 (top left). The comparison with the calculated spectra of Li_xSi and Li_4SiO_4 (bottom panel) suggests that both these features could be attributed to Li-silicides (and silicates), which indeed show lower energy onset and higher intensity in the 62-67 eV region. Furthermore, these features and the ones corresponding to Li_2CO_3 undergo significant changes after discharge, when Li^+ is extracted from the NPs, thus confirming the observation from the other edges of a partial Li_2CO_3 decomposition upon delithiation. We performed a LC using only experimental references in Figure 4.11 (left) and combining experiment with theory (right). The last clearly shows that the lithiated state is composed in its major part by Li_xSi and Li_2CO_3. However, the delithiated state is not reproduced totally by the fitting indicating a composition much more diverse than the set of fingerprints we used for the LC model.

To try to classify and quantify the Li loss, we express the total capacity after

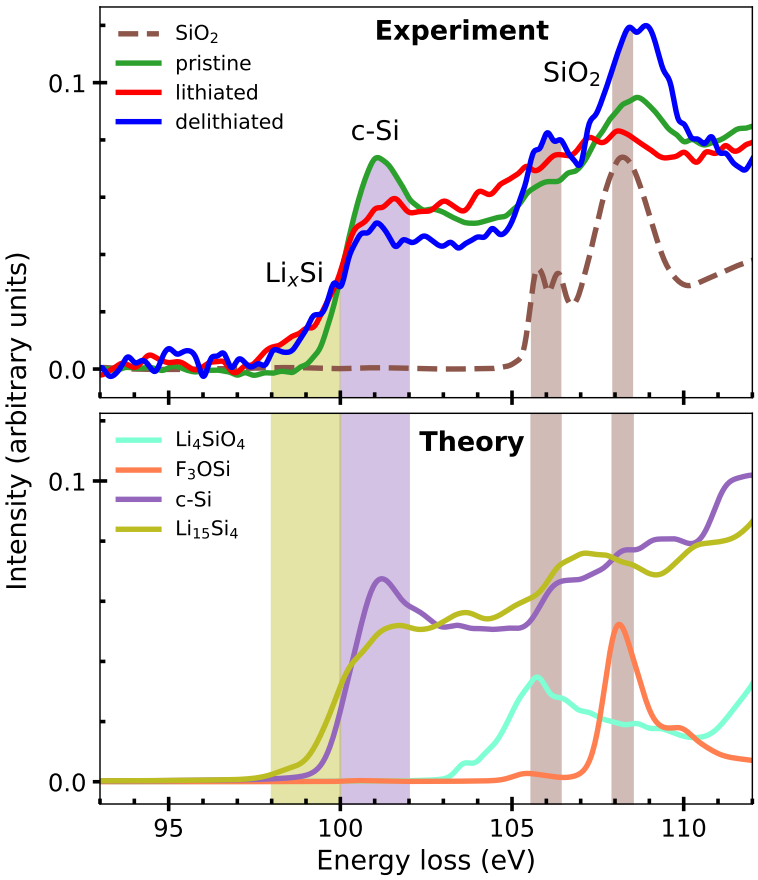

Figure 4.8: XRS Si $L_{2,3}$ edge spectrum of c-Si NPs electrodes in the low momentum transfer regime (top panel) compared with experimental SiO_2[4] (dashed brown lines) and theoretical references (bottom panel).

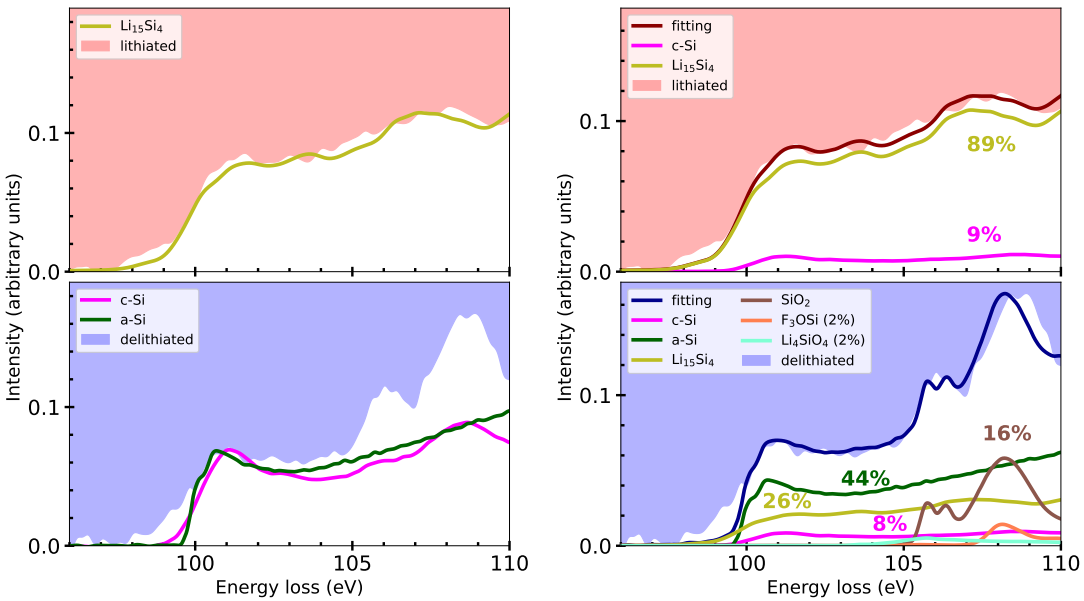

Figure 4.9: (Left) Comparison of the electrode Si $L_{2,3}$ edge spectrum in both SOC with the c-Si (this experiment), a-Si [9] (XANES experiment) and $Li_{15}Si_4$ (theory) references. (Right) LC fitting for the Si $L_{2,3}$ edge corresponding to the two SOC using as references c-Si, a-Si [9] and SiO_2 [4] from experiments and $Li_{15}Si_4$, F_3OSi and Li_4SiO_4 from theory.

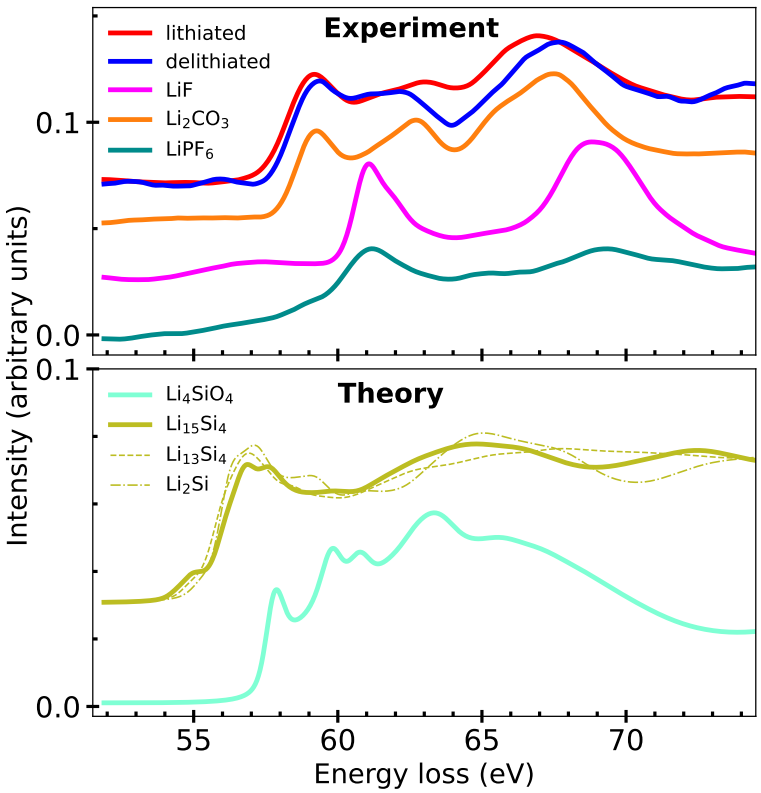

Figure 4.10: XRS Li K-edge spectrum of c-Si NPs electrodes in the high momentum transfer regime (top) compared with experimental (top) and theoretical references (bottom).

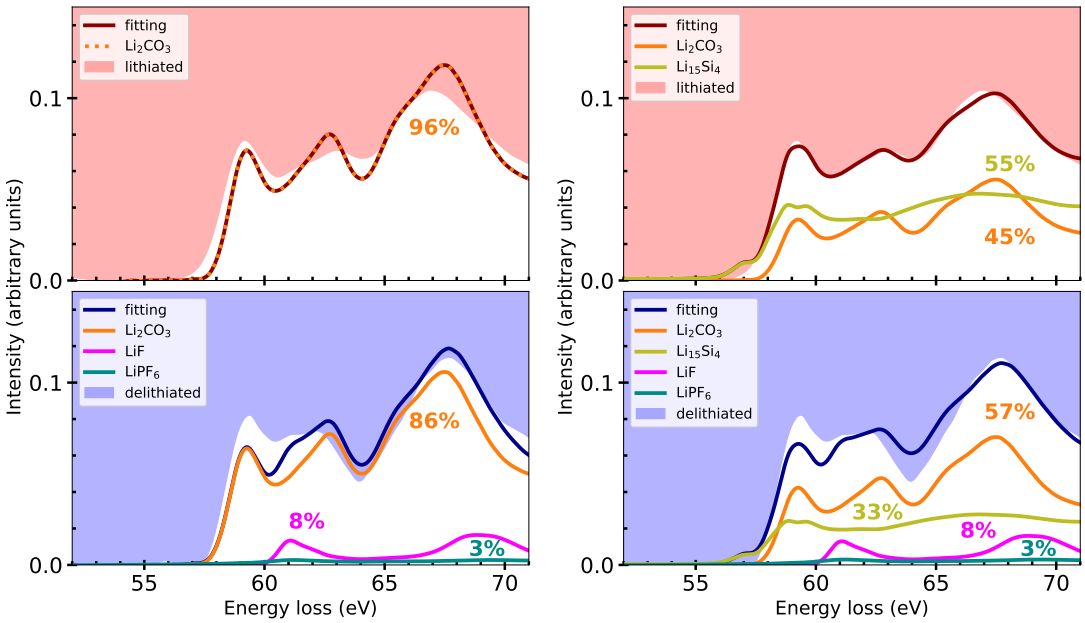

Figure 4.11: (Left) LC fitting for the Li K-edge corresponding to the two SOC using the experimental references Li_2CO_3, LiF and $LiPF_6$. The percentages of LiF and $LiPF_6$ obtained in the lithiated state are both less than 0.01. (Right) LC fitting for the Li K-edge corresponding to the two SOC using the references Li_2CO_3, LiF and $LiPF_6$ from experiment and $Li_{15}Si_4$ from theory.

the first lithiation Q^{lit} (red dot in Figure 4.2) as

$$Q^{lit} = Q^{lit}_{SEI_I} + Q^{lit}_{Li_xSi} + Q^{lit}_{SEI_{II}}, \tag{4.1}$$

where $Q^{lit}_{SEI_I}$, $Q^{lit}_{Li_xSi}$ and $Q^{lit}_{SEI_{II}}$ are the portions of the total capacity involved in the formation of fluorinated SEI, Li_xSi and Li_2CO_3-rich SEI, respectively. Similarly, following delithiation the retrieved capacity Q^{delit} (blue dot in Figure 4.2) is

$$Q^{delit} = Q^{delit}_{Li_xSi} - Q^{delit}_{SEI_I} - Q^{delit}_{SEI_{II}}, \tag{4.2}$$

being $Q^{delit}_{Li_xSi}(Q^{delit}_{SEI_I} + Q^{delit}_{SEI_{II}})$ the capacity extracted (lost) from the Si NPs (SEI evolution) upon delithiation. After combining Equation 4.1, Equation 4.2 and defining the capacity loss due to the Li trapped inside the NPs as $Q_{Li_{trap}} = Q^{lit}_{Li_xSi} - Q^{delit}_{Li_xSi}$, we can express the total capacity loss as

$$Q^{lit} - Q^{delit} = Q^{lit}_{SEI_I} + Q^{delit}_{SEI_I} + Q^{lit}_{SEI_{II}} + Q^{delit}_{SEI_{II}} + Q_{Li_{trap}}. \tag{4.3}$$

Under the previous assumptions, Q^{lit}, Q^{delit} and $Q^{lit}_{SEI_I}$ can be directly retrieved from Figure 4.2. Concerning the amount of Li trapped in Li–Si alloys, recent titration-gas chromatography experiments conducted on similar Si NPs with a Na-CMC binder [241] have found $Q_{Li_{trap}} \sim 19\%$ of the total capacity loss after the first cycle. A similar value of 17% can be obtained indirectly using a different approach described in section 6.6. These findings allow us to classify the total capacity loss, quantified at 547 $mAhg^{-1}$, by attributing 17% to the residual Li_xSi and 83% to the processes involving the formation and decomposition of the SEI.

The percentage of Li_2CO_3 decomposition upon delithiation can be quantified by correlating the linear combination weights of Na-CMC+SP and Li_2CO_3 obtained for the C K-edge with the relative quantities of carbon atoms. If we define C^i_X, w^i_X and m^i_X as the amount of C atoms, the linear combination weight, and the mass, respectively, corresponding to the compound X in the SOC $i = d, l$ (where d stands for delithiated and l for lithiated), then we can establish that $\frac{C^i_{Na-CMC+SP}}{C^i_{Li_2CO_3}} \equiv \frac{w^i_{Na-CMC+SP}}{w^i_{Li_2CO_3}}$. Assuming a negligible decomposition of the binder and the conductive additive (i.e., $C^d_{\text{Na-CMC+SP}}$=$C^l_{\text{Na-CMC+SP}}$), we obtain $\frac{w^d_{Na-CMC+SP}}{w^d_{Li_2CO_3}} C^d_{Li_2CO_3} = \frac{w^l_{Na-CMC+SP}}{w^l_{Li_2CO_3}} C^l_{Li_2CO_3}$, leading to $\frac{C^d_{Li_2CO_3}}{C^l_{Li_2CO_3}} = \frac{m^d_{Li2CO3}}{m^l_{Li2CO3}} = 0.65$. This corresponds to a 35% of decomposition of the Li_2CO_3 formed during the first lithiation. These considerations allow us to express the total capacity loss of Equation 4.3 as:

$$Q^{lit} - Q^{delit} = Q^{lit}_{SEI_I} + 1.35 Q^{lit}_{SEI_{II}} + Q_{Li_{trap}} \tag{4.4}$$

Using the results from Figure 4.2 and Equation 4.4, we can approximately split the total capacity loss during the initial cycle into 17 % from the remaining Li_xSi, 30 % from the formation of F-rich SEI, and 53 % from the Li_2CO_3-rich SEI part.

4.2 Materials and methods

4.2.1 Sample preparation

Crystalline SiNPs of size between 60 and 100 nm were synthesized by laser pyrolysis from high purity silane [230]. Electrodes were prepared by mixing 50 wt% of SiNPs, 25 wt% of sodium carboxymethyl cellulose (Na-CMC, Merck), and 25 wt% carbon black (Super P). The powders were ground and dissolved in purified water (18.2 Ω at 25 °C) to produce a slurry, which was deposited on a Cu foil (20 μm, thickness) and dried for 12 h at 80 °C.

Coin cells were prepared by cutting out circular disks of this Si-based material (0.97 mg/cm^2), assembled with carbonate-mixture based electrolyte, celgard 2400 (monolayer polypropylene) separator, and lithium metal. The electrolyte used was composed by 1 M $LiPF_6$ (3FEC/7EMC,v/v) with 2 wt% vinyl carbonate (VC). The cells were cycled and stopped at a given potential for obtaining different states of charge (SoC) conditions. Subsequently, each coin cell was disassembled and transferred without rinsing to designed airtight cell in a glovebox with Ar to probe Li, Si, C, O and F chemical environments by post-mortem XRS measurements.

4.2.2 XRS measurements

XRS experiments were performed on the beamline ID 20 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The pink beam from four U26 ondulators was monochromatized to an incident energy of 9.6837 keV, using a cryogenically cooled Si(111) monochromator and focused to a spot size of approximately 10 mm × 20 mm (V × H) at the sample position using a mirror system in Kirkpatrick–Baez geometry. The large solid angle spectrometer at ID20 was used to collect XRS data with 72 spherically bent Si(660) analyzer crystals. The data were treated with the XRStools program package as described elsewhere [245, 246].

The powder samples were pressed into a pellet, which was placed into the beam so to have a 10 grazing incident beam. All XRS measurements were collected at room temperature. Full range scans were collected from 0 to 700 eV with a 1 eV step size. After the acquisition of the broad scan, several detailed scans of specific edges were collected with a 0.2 eV step by scanning the incident beam energy to record energy losses in the vicinity of the Li K-edge (54.7 eV), C K-edge (284.2), O K-edge (525 eV), F K-edge (696.7) and Si L-edge.

Acquisition scans lasted around 2–4 h per sample. All scans were checked for consistency before averaging over them. The overall energy resolution of the XRS spectra was 0.7 eV, as estimated from the FWHM of elastic scattering from a piece of adhesive tape. Also, signals from analyzer crystals at different scattering angles were measured, covering a momentum transfer from 2.5 to 9.2 Å^{-1}. The data were integrated at high q for further analysis.

4.2.3 Computational details

XRS spectra in the low momentum transfer regime were computed with the XSpectra [99, 100, 101] code of Quantum ESPRESSO [36, 37, 38], while for the high momentum transfer XRS we used a modified version of XSpectra [3]. Crystalline and molecular structures used as theoretical references were obtained from the Materials Project [247] and PubChem [248] databases respectively. Ultrasoft pseudopotentials from the SSSP library v1.1 [249] with PBE exchange correlation functional were used for ground state calculations and non-absorbing atoms. Core-hole pseudopotentials were instead generated to describe excited atoms for the simulation of K-edge (C, O, F and Li) and L_{23} edge (Si) spectra. The energy cutoffs for the kinetic energy of the wave functions and the charge density were set to 80 Ry and 640 Ry, respectively. A supercell approach was employed to minimize interactions between excited atoms and their periodic images, maintaining a minimum distance of 8 Å between replicas in crystalline systems. In the case of molecules, the supercell size was chosen according to requirements of the Martyna-Tuckerman [84] method for treating the system as isolated. The Brillouin zone corresponding to the supercell was sampled at the Γ point for molecules and using a Monkhorst-Pack mesh with a k-points grid spacing of 0.25 $Å^{-1}$ in the case of crystals. In addition, all the spectral calculations were conducted with a 4x4x4 k-point grid with an energy-dependent broadening parameter [3]. We applied different core-hole treatments depending on the system, i.e., XCH [82] for crystals and the hybrid approach combining HCH and FCH for molecules (see subsubsection 2.3.2.2). Each spectrum was aligned with respect to the excitation onset, expressed as the binding energy of the edge core state, computed within the ΔKS [74, 75] approach and referred to the highest occupied state in the presence of the core-hole [250, 251].

4.2.4 Linear combination fitting procedure

In the linear combination (LC) procedure we express the target spectrum $|I\rangle$ in terms of a basis of known references $|I_i\rangle$ as

$$|I\rangle = \sum_{i=1}^{n} c_i |I_i\rangle , \tag{4.5}$$

where the vectors $|I\rangle$ and $|I_i\rangle$ have m components determined by the number of energy bins of the experimental data. Since the size of the basis is smaller than the number of experimental points ($n < m$) the coefficients c_i are found approximately using the least squares method as follows

$$c_j = \sum_{i=1}^{n} ||\langle I_i | I_j \rangle||^{-1} \langle I_i | I \rangle . \tag{4.6}$$

For all the cases we considered that the target spectrum $|I\rangle$ and all the references $|I_i\rangle$ are normalized in the same energy range. Furthermore, a LC is considered valid only if all the coefficients are non negative ($c_i \geq 0$).

4.3 Conclusions

The XRS spectra of anodes composed of crystalline Si NPs, probed at different states of charge, have unique chemical signatures demonstrating the SEI formation and evolution. These signatures can be differentiated when comparing qualitatively the high-resolution real electrode spectra acquired in the energy range of light elements (C, O, Li, F, Si) with well-chosen representative reference compounds as $LiPF_6$, LiF, Li metal, Li_2CO_3, carbon binder phase, silicon powder, and pristine silicon electrode.

A semi-quantitative analysis performed by linear decomposition using experimental spectra for C, O and F edges, aided by comparison with DFT-based data, evidences the presence of Li_2CO_3, $LiPF_6$ and LiF after one cycle, and potentially minor quantities of additional degradation products, such as PO_xF_y and SiO_xF.

Weight factors of relevant reference compounds are extracted, revealing how they are formed or dissolved during (de)lithiation. Specifically, we find an evident decrease in the peak intensities associated with carbonates when the electrode is delithiated, while $LiPF_6$ and LiF amounts appear to remain nearly stable during the first cycle. The SEI, which grows during lithiation due to the decomposition of organic electrolytes (EMC and FEC), additives (VC), and salt ($LiPF_6$), is thus found to partially dissolve upon delithiation, in agreement with previous studies [208].

Simulations also provide insights to interpret Si and Li edges and discuss correlations between the amounts of lithium trapped in the SEI and in the silicon. In particular, the SEI-breathing is accompanied by an incomplete delithiation, which relates to the significant capacity loss in the first cycle. Indeed, we evidence that lithium is not fully removed from the anode in the delithiated state, even after only one cycle. The irreversible lithium loss due to trapping in alloyed Si regions is quantified at 2.7% of Li unavailable for reactions, in agreement with other studies [241].

While further investigations are surely needed to better comprehend these phenomena, we here anticipate that XRS can be successfully applied to investigate in a semi-quantitative way a wide range of composite materials, enabling to gather high-fidelity bulk information, highly complementary to invasive surface-limited XPS. Spectra simulations based on ab initio DFT approaches also prove valuable to interpret and complement XRS measurements, and – validated towards more extended sets of data – could be critical to predict SEI features in a variety of materials for unexplored post-lithium ion technologies.

Chapter 5

Automated workflow for high-throughput XPS simulations

XPS experiments provide detailed information about the chemical environments within a system; however, deconvoluting the resulting spectrum presents challenges in terms of understanding the origin of each spectral feature. Conventional analysis techniques involve comparing the final spectrum to known standards of related materials obtained either from experimental [252, 7, 253] or computed spectra [254, 255] databases. In practice, XPS is a relatively slow technique, with acquisition times ranging from 30 min to several hours [256], which make the construction of such databases impractical using only experimental data. Moreover, it is challenging to experimentally probe unstable or metastable phases of a system, or to isolate the various polymorphs that contribute to the spectrum of an amorphous material. In this context, theoretical simulations, in particular those using high-throughput computing (HTC), are gaining in relevance for their capabilities to generate the large volume of data required for data mining.

Over the past decade, HTC has been systematically integrated into materials science [257, 258], driven by the automation of *ab initio* codes. This advancement has been further accelerated by the development of open-source infrastructures like AFLOW [259], AiiDA [260], ASE [261], Atomate [262], MAST [263], MatCloud [264], MPInterfaces [265], alongside commercial alternatives such as Material Studio [266], MeDA [267], Quantum ATK [268] and SimStack [269]. These frameworks have been instrumental in the creation and continuous growing of notable materials databases including AFLOW, OQMD [270], NOMAD [271], ICSD [272], Materials-Cloud [273], JARVIS [274] and Materials Project [275] (part of the Material Genome Initiative [276]). However, in the context of core-level spectroscopies we could mention only a few works dedicated to the development of automated workflows for *ab initio* simulations. For XAS we found implementations using the FEFF workflow infrastructure provided by Atomate [254, 255] and others oriented to the automation of input file generation and spectrum post-processing [251]. The particular case of XPS is even more scarce, represented

mainly by a core-hole dedicated workflow implemented in the AiiDA-FLEUR plugin [277]. This implementation relies on the all-electron DFT code FLEUR [278] with capabilities for computing absolute BE for metals, within the ΔSCF approach, and CLSs, for any system, using both ΔSCF and the ISA. The main challenges faced when using this workflow is the high computational cost of the ΔSCF method and the exclusion of screening effects for the CLSs computed within the ISA.

In the following we present a framework of automated workflow and helper functions implemented within the AiiDA-Quantum ESPRESSO (AiiDA-QE [260, 279]) module of AiiDA for the automated simulation of XPS spectra. We choose the Quantum ESPRESSO software package based on its set of sub-modules which allow for a range of properties calculations under a unified format. The purpose of this tool is to make XPS calculations simpler, more robust, easier to manage, and compatible with data reporting tools (e.g., Electronic Lab Notebooks) as well as enhanced standards for data such as the FAIR (Findable, Accessable, Interoperable, Reusable) principles [280]. Moreover, it will complete, together with an XAS plugin based in XSpectra code, a tool set for computing core-level spectroscopies within AiiDA-QE. The details of this chapter and the XAS tool are contained in the following works: "*Automated Workflows for Core-Level Spectroscopy Simulation*", Peter N. O. Gillespie, **Michael A. Hernandez Bertran**, Elisa Molinari, and Deborah Prezzi, *Preprint*, and "*AiiDAlab Quantum ESPRESSO App –The web environment for first-principles calculations*", Xing Wang, Peter Gillespie, Andres Ortega-Guerrero, Miki Bonacci, **Michael A. Hernandez Bertran**, Deborah Prezzi, Carlo A. Pignedoli, Nicola Marzari, and Giovanni Pizzi, *Preprint*.

5.1 The AiiDA infrastructure

To automatically manage the XPS workflow, the *ab initio* simulations involved, and all the generated data, the open-source framework Automated Interactive Infrastructure and Database for Computational Science (AiiDA) was utilized. AiiDA, developed under the MIT license, is based on the ADES model [281] (Automation, Data, Environment, Sharing) for computational science. This model outlines essential design criteria for scientific work environments, promoting open data provenance [282] and ensuring compliance with the FAIR principles for scientific data management.

The automation pillar in AiiDA is achieved via a Python API and a daemon. The API provides Python classes for processes, calculations, and data structures, designed to be extensible with plugins. All user tasks can be executed in Python, facilitating integration with various materials science software. The AiiDA daemon runs in the background, handling job submissions, retrieval, and management, along with interactions with computing schedulers. The communication between the daemon and its subprograms, called workers, is handled through the

message and streaming broker RabbitMQ (rabbitmq.com), allowing to sustain high-throughput workloads. Users interact with the daemon through the `verdi` command line, which offers commands for listing and inspecting calculations and workflows.

The data pillar is addressed by tracking the provenance of data and logic. AiiDA stores input and output files in a file repository, while parsed data is saved in a PostgreSQL (postgresql.org) database based in the SQL language. Data is stored in a directed acyclic graph structure, ensuring clear tracking of all calculations. AiiDA also supports caching, which prevents reruns of previously completed calculations, saving computational resources.

The environment pillar is implemented through the daemon, the plugin system, and the workflow system. Plugins include parsers, calculation classes, workflows, and `verdi` command extensions, which can be shared as Python packages via PYPI (pypi.org).

For sharing, AiiDA offers import/export capabilities for results and data, while SQL databases can be easily shared with standard tools.

5.1.1 The AiiDA plugin system

To handle with the inhomogeneity in data structures and file formats, AiiDA has a lightweight core code, with additional functionalities organized into plugins. These plugins are developed and maintained by the community. The AiiDA team provides templates for various plugins, such as those for command line interfaces, data,schedulers, parsers, calculations and workflows. Additionally, plugins for different storage backends, like PostgreSQL, can also be deployed.

To use a program with AiiDA, at least a calculation plugin and a parser plugin must be implemented. Parser plugins convert input/output file data into structures stored in the database, while calculation plugins define how to run a calculation for the given code by generating input from data structures. Pre-existing interfaces for community standards like Crystallographic Information file (CIF) [283], XCrySDen file format [284], and others facilitate this process, but new data structures can be developed via data plugins if needed.

Plugins and utilities for a specific program are packaged into an AiiDA extension, and several such packages have been created for the most popular quantum engines. All extension names are gathered in the AiiDA plugin registry (aiidateam.github.io/aiida-registry/) where users can find a list of available plugins, their content, installation instructions, and compatibility information.

5.1.2 Workflows in AiiDA

A key feature of the AiiDA framework is its ability to write, run, and share workflows. These workflows, known as workchains, automate complex, time-consuming

calculations that depend on each other, without requiring user intervention between steps. Technically, workflows are Python classes that inherit from AiiDA's base classes (`WorkChain`, `WorkFunction`), providing developers with the flexibility to incorporate any Python packages. This adaptability distinguishes AiiDA from other workflow infrastructures. Developers can encode complex protocols within workflows, while AiiDA ensures both data and logic provenance, providing in this way, a turn-key solution for launching HTC with a small Python script.

Workflows can either run in the background via the AiiDA daemon or block the Python interpreter during execution. They are robust and fault-tolerant, can handle tasks like optimizing resource usage, error handling, and restart mechanisms. They also provide, in general, simple interfaces with default values for ease execution.

5.2 XPS workflow within AiiDA-QE

In order to establish a generic workflow process for the XPS calculation, in the following, we will inspect the steps involved in the framework of Quantum ESPRESSO. Firstly, the user prepares a PwSCF (`pw.x`) calculation for the system of interest, in which one atomic site is chosen for the location of the core-hole –henceforth referred to as the "absorbing atom". This site is assigned a unique label relative to other atoms of the same element, as well as a pseudopotential for which the electron configuration of the core orbital of interest is changed accordingly to represent that of the excited state (e.g., $1s^1$ for K shell XPS). The user then prepares calculation inputs for the `pw.x` calculation taking into account the proper core-hole treatment of the final state (see subsection 2.2.2). With the core-hole approximation chosen, the user then obtains the final state wavefunction from the PwSCF calculation for the chosen absorbing atom.

To calculate the complete spectrum for orbital of interest, all atoms of the same element within the system must, in principle, be considered. However, as XPS depends on the local chemical environment of the absorbing atom, the number of sites to be considered can be reduced based on the symmetric equivalence of the sites. Thus, to calculate the complete spectrum for a given element, we need to take into account the symmetrically-nonequivalent atoms of the same element, with their respective multiplicity within the unit cell. It must be noted that, using this approach, the calculation of the final state with a localised core-hole necessarily introduces a charged defect into the system. Though this does not affect the symmetry of the crystal structure in question (insofar as concerns the XPS spectrum obtained from the calculation), the structure used in the calculation must be of sufficient size to limit the spurious interaction of the absorbing atom with its neighbouring images. Concerning XPS, a distance of 8.5 Å is sufficient in the XCH approximation, however in the FCH treatment further testing is typically required for convergence of absolute BEs [285, 73]. The final spectrum is repre-

sented as a sum of Voigtian profiles centered at the BE values (see Equation 2.29). This approximation allows a proper description of the broadening, including the effects of instrumental resolution and lifetime of the core-electron excitation.

Using this approach we are therefore able to define a generalised workflow to treat any system of interest. The following steps can be identified: (1) Check input requirements (structure, core-hole treatment, pseudopotentials, etc), (2) analyse the symmetry of the input structure and determine the number of symmetrically-nonequivalent atomic sites for each element, (3) create the supercell according to the symmetry analysis and the core-hole treatment, (4) set up pw.x calculation inputs for each nonequivalent absorbing atom and run calculations, (5) calculate the BE for each absorbing atom, and (6) calculate the final spectra weighting all the contributions by the multiplicity of each absorbing atom site. Excluding specific considerations required for treating systems with non-periodic dimensions (such as truncating the Coulomb potential), we find that this workflow logic can be effectively applied to systems with any level of periodicity. The structure of the described workflow is presented in Figure 5.1.

5.2.1 Technical details

Using the general process flow defined (Figure 5.1) as starting point we can implement the corresponding XPS workflow in the AiiDA automation and workflow management system. We leverage the modular system present in AiiDA to separate the steps defined previously into independent `WorkChains`, `CalcJobs`, and `CalcFunctions`. This allows us to create both fully-automated workflow for XPS calculations suitable for less-expert users, as well as a relatively comprehensive set of tools for power users to create more advanced workflows to suit their needs. The main, fully automated workflow is `XpsWorkChain`, which handle the complete calculation of XPS spectra, for all chosen elements. If requested, full geometry optimisation of the input structure is performed resorting to the `PwRelaxWorkChain`. Automatic calculation restarts are handled by relevant `WorkChains` (`PwRelaxWorkChain` and `PwBaseWorkChain`) based on the AiiDA `BaseRestartWorkChain` Class.

At lower level, the `get_xspectra_structures()` function is provided as a comprehensive tool in order to facilitate the generation of structures for calculation work. In general terms, this function takes an AiiDA `StructureData` node as input, analyses its symmetry, produces a supercell of the input structure of sufficient size to limit spurious interaction with cell replicas, and creates copies of the supercell for each inequivalent absorbing atom, with the relevant absorbing atom labeled using a special marker. For crystal symmetry analysis, we choose the `spglib` library [286] which has already been integrated with AiiDA-QE. Point group symmetry analysis is included in order to process molecular systems by interfacing `get_xspectra_structures()` with the `pymatgen` library.[287]

All created structures are returned as outputs alongside an AiiDA Dict node

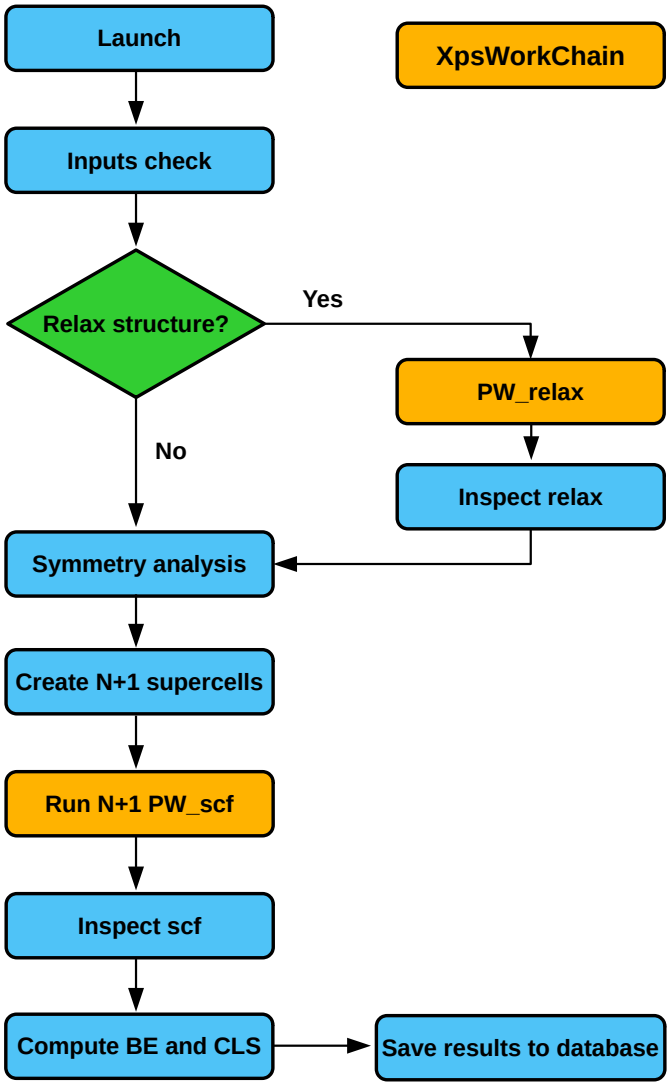

Figure 5.1: General structure of the XPS workflow for a system with N symmetrically non-equivalent absorbing atoms.

which reports the symmetry information of the input structure, data on inequivalent sites, and other metadata relevant to the analysis. The metadata are made readily available to the user in order to provide useful information on the symmetry properties of the chosen structure and is subsequently used by the main workchains to determine calculation inputs: for instance the number of calls to the `PwBaseWorkchain` for computing the SCF total energies needed for XPS.

The postprocessing steps are also fully implemented using the `CalcFunction` called `get_spectra_by_element()`. It uses a Voigt profile broadening scheme, and the incoming symmetry data to construct the final spectrum for each element according to the multiplicity of the non-equivalent atoms. Moreover, if requested and if core-level correction energies are provided for each absorbing element included in the workflow, it will also calculate the absolute binding energies using the ΔKS scheme, as described in subsection 2.2.2. The functionalities of `XpsWorkChain` and its `CalcFunctions` are summarized bellow and the process flow is presented in Figure 5.2, the source code is already integrated within AiiDA-QE and is available under the MIT licence.

- `get_xspectra_structures()`: Determines the symmetry of an incoming structure and returns a supercell for each non-equivalent absorbing atom site, including an unmarked supercell and the symmetry analysis metadata for the function call.

- `XpsWorkChain`: Computes a complete XPS spectrum for a set of core-states (defined in the inputs via the core-hole pseudopotentials). The input structure is optimized if requested using the `PwRelaxWorkChain`.

- `get_spectra_by_element()`: Calculates the CLS value for each absorbing atom site referred to the site with the lowest BE. Produces the final spectrum using a Voigt profile broadening scheme. If requested, and if core-level correction energies are provided for each absorbing element included in the workflow, will also calculate the absolute binding energies using the ΔKS scheme.

5.2.2 Workflow validation

To properly demonstrate the WorkChain system presented here, we will show the procedure for performing the calculation of the C 1s XPS spectrum on the Phenylacetylene molecule. This system was selected in order to showcase the capabilities of this implementation to treat properly systems in which the reference level for the BE is the vacuum. Nevertheless, the functionalities of the workflow for computing XPS in extended systems within a HPC regime will be exemplified in the next chapter. In this, we will detail key features available for the user, both for the

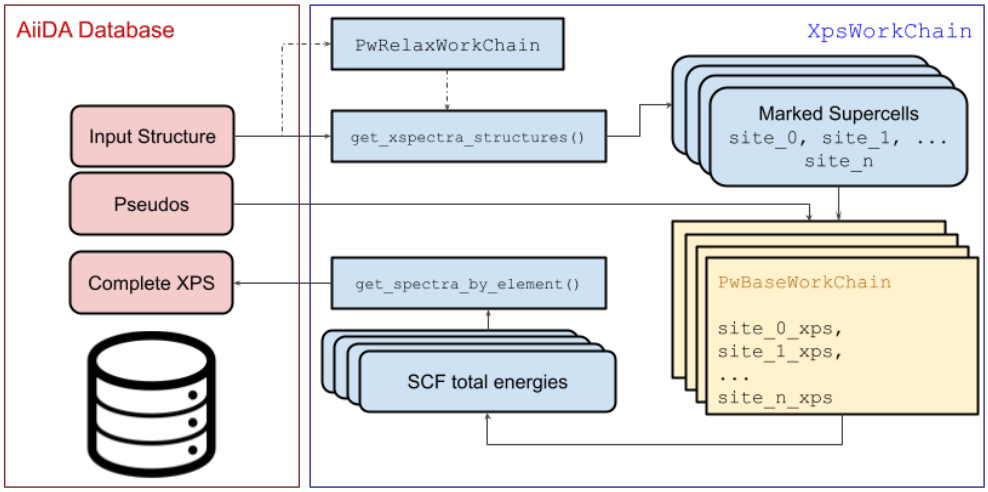

Figure 5.2: Process flow diagram for the `XpsWorkChain`.

case of full automation with minimal user input and also for producing more customised analysis with its sub-modules. In the case of molecules, the steps related to symmetry analysis and structure preparation are adjusted to satisfy the constraints imposed by finite dimensions. Figure 5.3 shows the information returned by `get_xspectra_structures()`, including the molecule point group and the list of non equivalent atoms with their index, label and multiplicity. The supercell utilized for all the SCF calculations is constructed during this step, adhering to the size requirements outlined in the Martyna-Tuckerman method for treating the system as isolated by removing the long range electrostatic interactions.

As a default, the `XpsWorkChain` will return the core-level shift for the set of non equivalent atoms. Moreover, it is possible to obtain absolute BE values and a final spectrum directly comparable to experiment (see Figure 5.4), as in this case, by setting the `calc_binding_energy` flag to `True` and populating the `correction_energies` dictionary in inputs. The procedure for obtaining the correction energies was described in subsection 2.2.2. The provenance graph for the calculation of the C 1s XPS in Phenylacetylene is presented in Figure 5.5.

```
'is_molecule_input': True,
'supercell_num_sites': 14,
'molecule_point_group': 'Cv2',
'equivalent_sites_data':    {
'site_6':   {'symbol': 'C',
    'site_index': 6,
    'multiplicity': 1,
    'equivalent_sites_list':[6]'
            },
                ...
                            }
'supercell_cell_matrix': [[15.1485, 0.0 , 0.0 ],
[0.0, 8.6601, 0.0], [0.0, 0.0, 8.0]],
'supercell_cell_lenghts': [15.1485, 8.6601, 8.0],
'absorbing_elements_list': ['C'],
```

Figure 5.3: Extract from IPython terminal output displaying part of the dictionary returned by `get_xspectra_structures()` for Phenylacetylene.

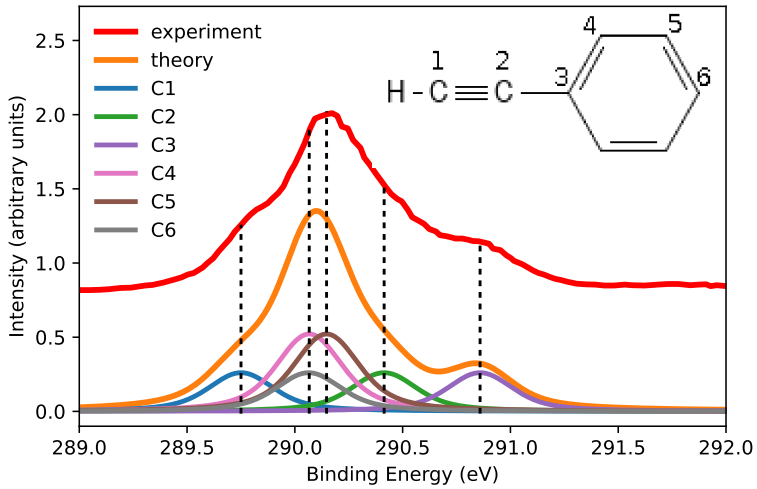

Figure 5.4: C 1s XPS spectrum of Phenylacetylene in gas phase, comparing experimental data from the literature (red curve, Ref. 10) to the spectrum computed using the FCH approach.

5.2.3 Integration with AiiDAlab

AiiDAlab [288] provides an excellent platform for researchers in computational materials science, offering an integrated space for reproducible and collaborative research workflows. By leveraging the AiiDA workflow engine alongside interactive computing and data analysis tools, AiiDAlab simplifies the research process from simulations to final analysis and publication. Its user-friendly interface enables

Figure 5.5: Provenance graph for the calculation of the C 1s XPS in Phenylacetylene. The first row on top contains all the inputs passed to the `XpsWorkChain` (red framed orange box), then subsequent calls to other workflow (orange boxes) launch the SCF calculations involved in the structural optimization (top left block) and BE calculations (bottom blocks). The outputs are returned to the workflow.

easy setup and execution of complex computational workflows, while its data management features ensure traceability and reproducibility of results. Additionally, the collaborative tools in AiiDAlab support seamless sharing of workflows and data, promoting interdisciplinary collaboration and advancing scientific discovery. AiiDAlab offers an accessible and efficient environment bringing cutting-edge materials research to a broader audience.

The workflow developed in this work was integrated by our collaborators into the QE App of AiiDAlab as a plugin, providing a user-friendly web interface for researchers to perform XPS calculations and visualize results. The plugin automates key steps such as pseudo-potential preparation, peak selection, and convoluting the results with a Voigt function to generate the final spectrum. It also includes result visualization tools, allowing users to upload experimental XPS data for direct comparison with the calculated spectra. This streamlined interface makes complex XPS analysis accessible and efficient.

5.3 Conclusions

We presented the design and implementation of a workflow dedicated to compute XPS. This tool offers capabilities and results with a level of theory comparable to others HPC implementations for XPS. As a surplus we added functionalities in order to treat successfully finite systems like molecules and clusters and to include final state effects in the CLS. The example presented demonstrates the potential of this Python-based approach for high-throughput calculations using the AiiDA framework. Its modular design allows users to fully customize workflows, including parameters like core-hole treatment and supercell dimensions, for tailored simulations.

Chapter 6

ML applications to core-level spectroscopies

Core-level spectroscopies are established strategies to assess the electronic structure and chemical environment of materials. While these techniques provide valuable insights into atomic and electronic structures, their interpretation in complex systems is not straightforward, emphasizing the importance of theoretical insights from ab-initio approaches. However, the time required for computing an X-ray spectrum scales linearly with the number of non-equivalent atoms, becoming prohibitively expensive for complex amorphous materials. This drawback can be addressed by employing a surrogate model that combines the precision of ab initio methods with computational efficiency.

In general, the application of ML approaches to core-level spectroscopies can be broadly grouped in two categories called *forward* and *backward* mapping [289]. The *forward* mapping is very similar to the standard ab initio approach in which the input is usually the atomic structure and the target quantity is the spectrum, represented by the BEs in XPS or the transition energies for XAS/XRS. In the case of the *backward* mapping the main task is to translate an (experimental) spectrum into a structure or environment. A topical review on the application of these approaches to XAS can be consulted in Ref. 289.

For the case of XPS considered here, there are only few models available. For the *forward* approach, a random forest model trained using the LMBTR descriptor is used to predict the C 1s spectrum for prototypical SEI [290]. Despite the accuracy showed by this model (0.05 eV in the mean squared error (MSE) of C 1s peak positions), it was trained for CLS computed within the ISA. This implies that a final alignment of the spectrum is always needed in order to compare with experimental data. In addition, further generalization to systems with poor screening should be done with care since the ISA neglects the core-hole relaxation effects in the final state. Another model based on KRR and SOAP descriptors was proposed to compute many-body corrections, within the GW approximation, to the DFT BEs [291] of CHO-containing molecules and materials. This model

presents an accuracy of the order of 0.05 eV for molecules and 0.1 eV for extended materials. The only limiting factor is the computational cost needed for the GW calculations.

In the context of *backwards* approaches, clustering and dimensionality reduction techniques are widely used to simplify to provide spectral interpretations. For example the clustering of SOAP descriptors was used to generate a basis of spectroscopic fingerprints in order to associate a spectral feature with a local environment [292, 293]. Another possibility is to consider machine-learned features, such as those extracted by neural networks (NN), as an alternative to handcrafted features based on physical or chemical intuition. In this sense, convolutional neural network (CNN) feature extractors, combined with regression or classification models, were utilized for automated detection and/or quantification of chemical elements in XPS data [294, 295].

In this work, we will adopt a methodology framed within the *forward* approach combining gaussian KRR with SOAP descriptors. A similar strategy was presented in Ref. 291 to include GW corrections on top of DFT. We will use theoretical results at the DFT level to avoid the computational cost of the GW method, as the accuracy improvements provided by GW in XPS are primarily significant for systems with larger bandgaps. Since we will focus on gapless or nearly gapless systems, DFT will be sufficient for our analysis.

6.1 ML model for predicting XPS

The key ingredient for the training of any ML model is data. In particular, the accuracy and generalization of the model are conditioned by the quality of both the inputs and the labels. Input quality is often measured in terms of diversity, while label quality depends primarily on the level of theory used. In this work, the accuracy of the labels is fixed, as they were generated using the `XpsWorkChain` workflow introduced in chapter 5. Two datasets from previous studies, relevant to graphite-based and silicon-based anodes in Li-ion batteries, were selected for the inputs.

6.1.1 Training datasets

In the following, we will provide a short description of the datasets used in this thesis. More details can be found in the original publications, see Ref. 296 for the Li_xC dataset and Ref. 297 in the case of the Li_xSi dataset.

6.1.1.1 Li_xC dataset

An evolutionary algorithm (EA), as implemented in USPEX [298, 299], was employed to explore the configuration space of Li–C systems, starting with a ReaxFF

model tuned to fit DFT results with van der Waals corrections. Initial configurations of 16 and 24 carbon atoms per unit cell were generated at pressures from 0 to 50 GPa. The EA population evolved through heredity, mutation, and random structure generation, with optimization occurring in multiple steps, including molecular dynamics (MD) and structural optimization. Only the most stable parents advanced to the next generation, ensuring diversity through thresholds on structural similarity. This process generated a variety of sp, sp^2, sp^3, and mixed sp^2/sp^3 carbon structures over up to 50 generations.

A hierarchical clustering approach, using single linkage and based on distance, was initially applied to all structures generated by the EA to identify unique polymorphs. In later iterations, clustering focused only on structures where the neural network potential (NNP) predictions differed from DFT ground-truth energy by more than 5 meV/atom, preventing over-sampling of well-described polymorphs. Structural similarity was measured using a fingerprint-based cosine distance [300, 301].

A representative structure from each cluster was selected, and a 0.5-ns classical NPT molecular dynamics (MD) simulation was performed using the LAMMPS package [302]. Snapshots were taken every 5 ps, yielding 100 configurations per cluster reaching a total of 20000 structures.

We utilized SOAP descriptors to construct structural representations of the C and Li environments, enabling us to visualize their diversity through PCA. Visualization of PCA was done using the interactive structure/property explorer for materials and molecules (Chemiscope) [303], while SOAP descriptors were constructed using the librascal code (lab-cosmo.github.io/librascal/). Figure 6.1 shows the projection of the local environment representation onto a 2D space defined by the PC_1 and the PC_3. In the left panel we can observe the results for the Li environments represented by SOAP descriptors with a cutoff radius r_{cut}=6 Å. This configuration clearly split the Li environments in two clusters, according to the dimentionality (e.g., 2D and 3D) of the host structure. The number of point in each cluster evidences an over representation of the 3D structures with respect to the 2D. This was confirmed by inspecting the amount of vacuum in each unit cell, leading to a demographic of ∼3000 2D structures and ∼17000 3D structures. A similar analysis was conducted for the C environments coded into SOAP descriptors with r_{cut}=2 Å. As shown in the right panel, this configuration leads to the formation of three clusters identified by the sp, sp^2 and sp^3 C hybridization. Moreover, a higher representation of sp^2 and sp (defective sites) C environments can be appreciated as compared to the sp^3 hybridization.

6.1.1.2 Li_xSi dataset

The initial dataset was created to provide a starting point for a deep potential (DP) model, employed to investigate the properties of crystalline and amorphous phases of Si for application as anode material in Si-based electrodes. This dataset

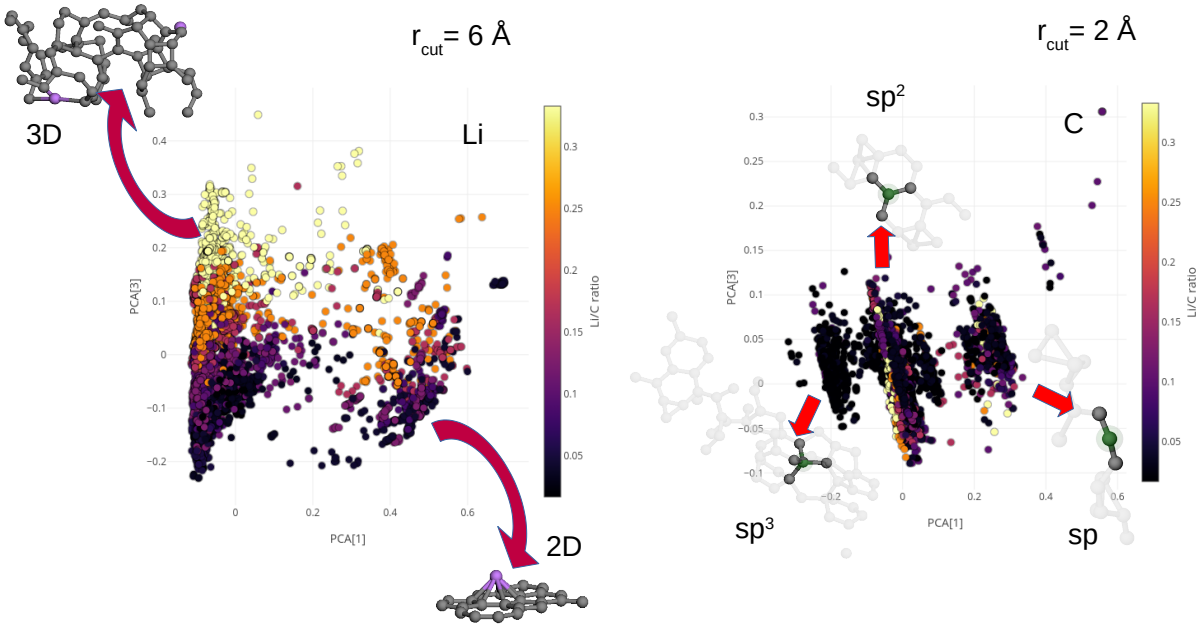

Figure 6.1: Visualization of the Li_xC dataset environments represented with SOAP descriptors using PCA. Li environments (left) are described with a larger cutoff distance (6 Å) splitting the dataset in two main clusters corresponding to 2D and 3D structures regardless of the Li concentration (colored vertical bar). C environments (right) are described with a smaller cutoff distance (2 Å) splitting the local environments according to *sp*, sp^2 and sp^3 C hybridization.

includes stable ground-state structures of Li, Si, and crystalline Li-Si (c-Li_xSi) phases from the Materials Project. The dataset incorporates structures like BCC, FCC, and HCP Li, diamond Si, I41/amd Si, and various c-Li_xSi phases ($0<x<4.5$). After generating supercells and optimizing them using DFT, perturbations were introduced to atomic coordinates and cell vectors, followed by ab initio MD simulations, resulting in 10625 configurations.

In the exploration stage, new configurations of Li, Si, and Li-Si systems were sampled using DeePMD simulations. Model deviation was used to screen configurations, improving the DP model over several iterations. A total of 28103 configurations were labeled for the explored dataset.

For the amorphous phases, nine a-Li_xSi phases were explored over 87 iterations. Additionally, low-lithium configurations (a-Li_xSi $0<x<1.33$) were created using DeePMD and grand canonical Monte Carlo simulations, leading to 34150 configurations in total.

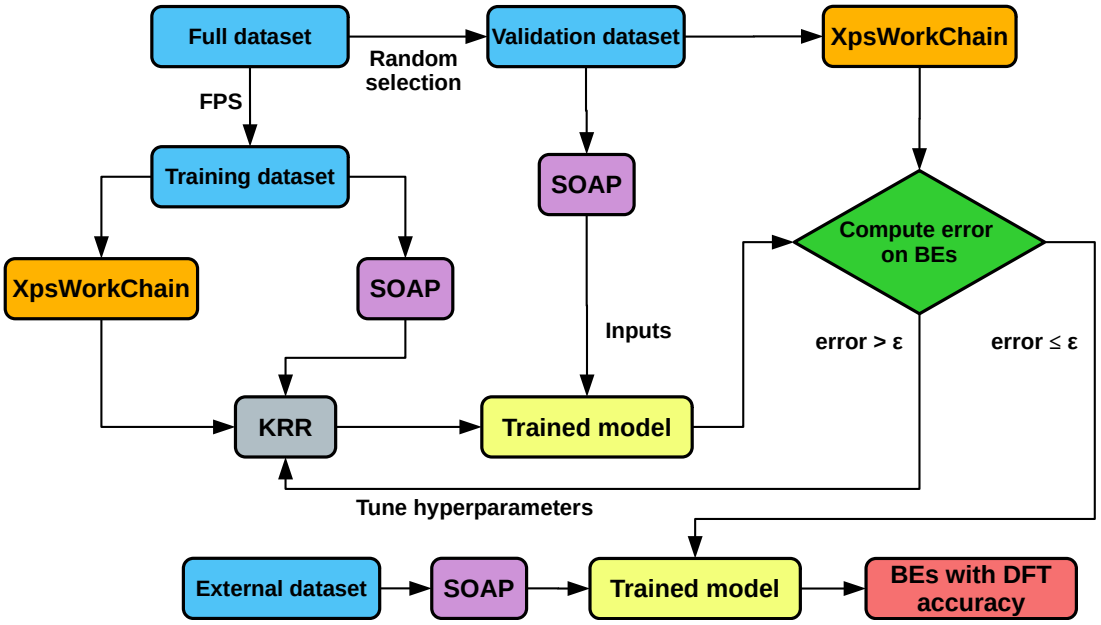

Figure 6.2: General scheme of the ML model used for the BEs predictions

6.1.2 Training and validation using DFT BEs

The systems within both datasets are expected to have either zero or small bandgap, according to the electronic properties of similar structures in Materials Project database. In this scenario, the theoretical approach adopted in the `XpsWorkChain` will provide BEs with an accuracy (ϵ) comparable to experimental results (see chapter 5 and subsection 2.2.2 for more details). Thus we can build a ML model to predict BEs by training the model on a database made up of theoretical BEs obtained from the Li_xC and Li_xSi datasets. We could expect results comparable with the experiments as long as the errors in the predictions are lower than some predefined threshold (δ). Formally, if $|BE_{DFT} - BE_{exp}| < \epsilon$ and $|BE_{ML} - BE_{DFT}| < \delta$, then, $|BE_{ML} - BE_{exp}| < \epsilon + \delta$. A similar hypothesis was used for the study of nuclear magnetic resonance (NMR) chemical shifts in molecular solids [304].

The general scheme of the training and validation process is depicted in Figure 6.2. We start with the full dataset of Li_xC (Li_xSi) containing ∼20000 (∼62000) structures. From this, we construct two disjoint subsets for training and validation. Since the training set should capture most of the diversity of the environments, we used the FPS algorithm for this task. In practice, the FPS algorihm was applied to the set of structure-averaged SOAP vectors constructed using r_{cut}=6Å , with radial scaling, and atomic density smoothing σ_a=0.3 Å. The resulting training set contains 2000 (245) structures, comprising ∼3500 Li and ∼45500 C (∼7800 Li and ∼7800 Si) atomic environments. The validation set was selected randomly from the remaining structures containing 480 (55) structures, with ∼890 Li and

∼10200 C (∼1900 Li and ∼1900 Si) atomic environments. In a next step the BEs corresponding to the structures in the training set are computed with the `XpsWorkChain` workflow and at the same time the local environments are coded in a SOAP representation. SOAP descriptors and BEs will be used as inputs and labels , respectively, for training the gaussian KRR model. The training is performed with the scikit-learn Python library [305] using a grid search in the space of the hyperparameters in order to find the optimal parameters minimizing the mean absolute error (MAE) in a 5-fold cross validation configuration. The validation step is done by computing the BEs of the validation set using both the ML model and the `XpsWorkChain` workflow. The model is evaluated on an out-of-sample system (external dataset) once the error threshold is met, in order to assess its generalization capabilities.

Figure 6.3 presents the training results for models predicting the Li 1s and C 1s BEs in the Li_xC dataset. The left panels show the prediction error of the BEs compared to those computed using ab initio DFT for the validation set. This illustrates the overall prediction accuracy for the C 1s (top) and Li 1s (bottom) BEs, represented by the MAE in eV, as a function of both the cutoff radius r_{cut} and the number of training structures included from the training set. The effect of the different r_{cut} values on the MAE is visible considering r_{cut}=5.5 Åand r_{cut}=3.5 Å, which produce the larger and smaller values of MAE, respectively. Irrespective of the value of r_{cut}, all the models show a decreasing error trend as we add more environments into the training process.

On the right panels we present the correlation plots for the Li 1s (top) and C 1s (bottom) BEs between the best model predictions and the ab initio DFT reference for both the training (red triangles) and the validation set (blue triangles for Li and grey triangles for C). We trained a total of 12 models by taking into account all the combinations of cutoff radii $r_{cut} = 2.5, 3.5, 4.5, 5.5$ Å, with radial scaling, and atomic density smoothing $\sigma_a = 0.3, 0.4, 0.5$ Å. The models with the best bias-variance trade-off, interpreted here as the one maximizing the coefficient of determination R^2 on the validation set and minimizing the difference between the R^2 in the validation and training sets, were selected as the best models. Such models were obtained by setting $r_{cut} = 3.5$ Å for Li 1s, $r_{cut} = 2.5$ Å for C 1s and $\sigma_a = 0.3$ Å in both cases. This models give an R^2 coefficient of 0.992 for Li 1s and 0.987 for C 1s.

The level of accuracy presented by these models within the Li_xC dataset could be leveraged to do some data analytics. For example in Figure 6.4 (weft) we show, using Chemiscope, the capabilities of the model for the identification of C sp^3/sp^2 hybridization harnessing the predicted BEs. Another application is presented on the right panel, where it could be appreciated that the presence of a Li atom (marked in green) only affects the nearest C atoms, in terms of BE, proving the locality of XPS.

In order to asses the generalization capabilities of the models trained on the Li_xC dataset, a group of external structures, generated with different approaches,

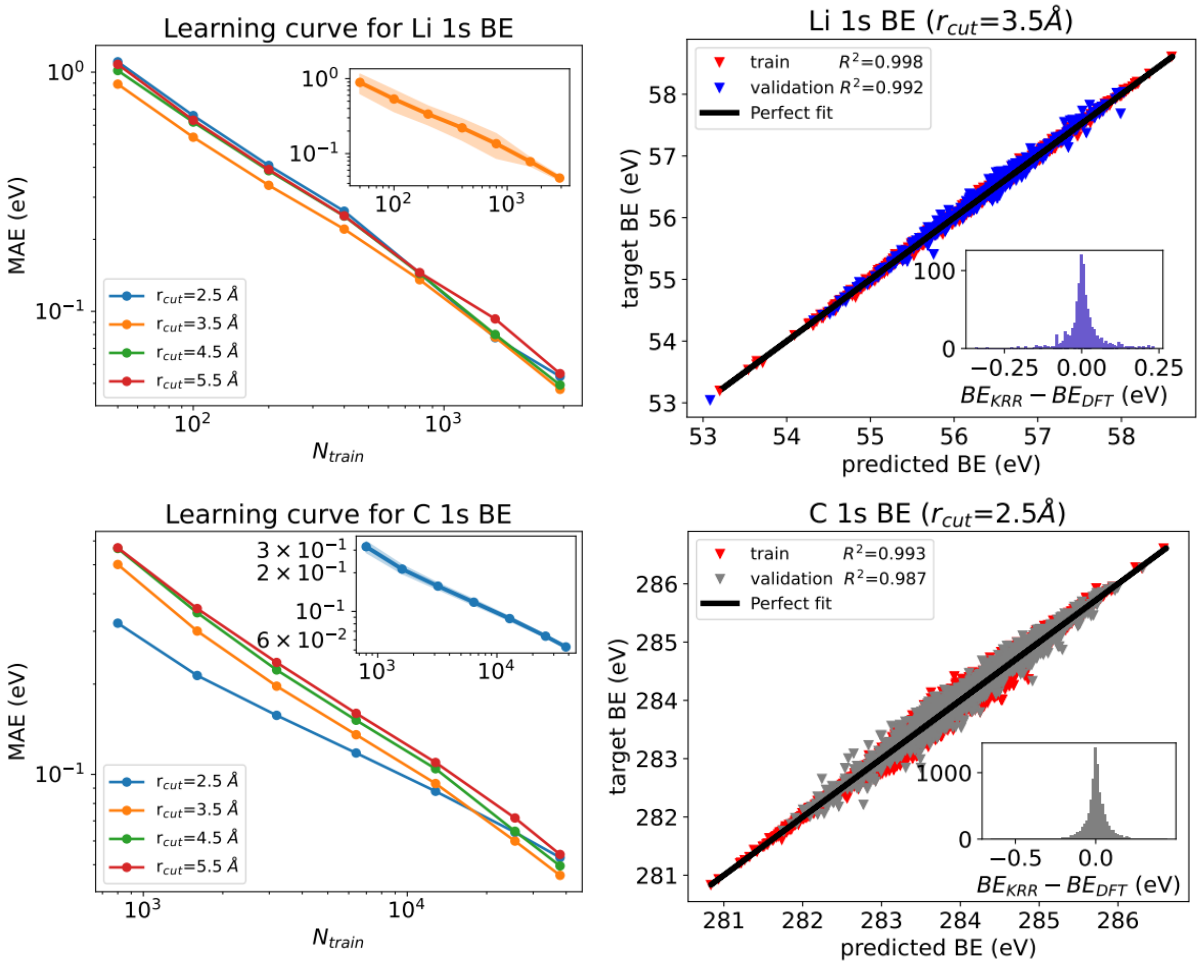

Figure 6.3: (Left) BE prediction error corresponding to the Li 1s (top) and C 1s (bottom) gaussian KRR models for the Li_xC validation set. The average MAE prediction error between BEs obtained with ML and ab initio DFT is shown for different local environment cutoff radii as a function of the training set size N_{train}. For the selected cutoff radius an inset was added containing the average BE error (dot and line) with its incertitude (shaded region) quantified as two times the standard deviation within the cross validation folds. (Right) Performance of the gaussian KRR models corresponding to the Li 1s (top) and C 1s (bottom) for the Li_xC train (red triangles) and validation (blue and grey triangles for Li and C respectively) sets. The coefficient of determination R^2 for the relation between ML predictions and ab initio DFT (target) is reported in the legend. The histogram for the difference between BEs obtained with ML and ab initio DFT is shown as an inset.

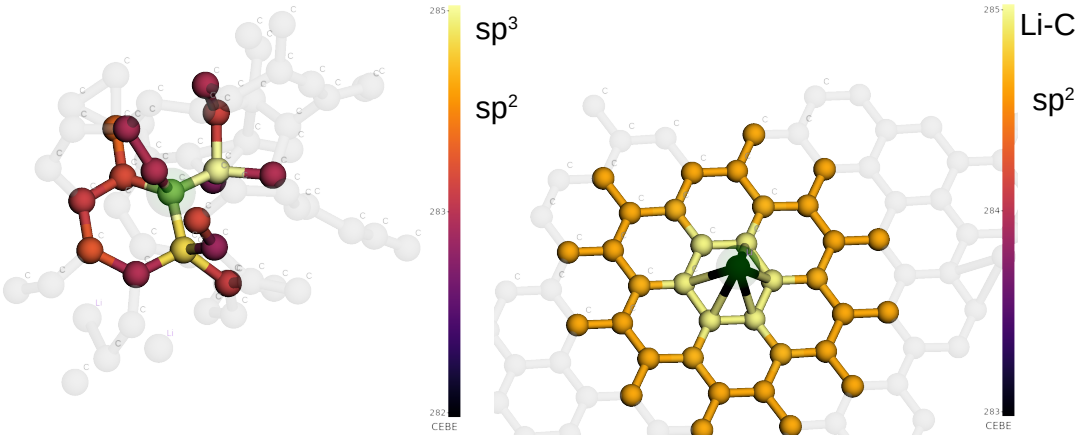

Figure 6.4: Examples of data analytics performed on the raw data generated by the ML model for the Li_xC dataset. (Left) C sp^3/sp^2 hybridization identification using the predicted BEs. The C atoms are colored according to the BE vertical scale. (Right) Probing the locality of XPS, only the nearest C atoms are influenced, in terms of BE, by the presence of a Li atom (marked in green).

was selected. This external dataset is composed by amorphous carbon (aC) from Ref. 306 and $Li_{12}C_{194}$, $Li_{28}C_{196}$ and $Li_{48}C_{216}$ from Ref. 307. The ML predictions are benchmarked against full and approximated ab initio DFT as described in Figure 6.5. The approximated DFT spectra were obtained by applying different clustering algorithm to the SOAP descriptors. This approach computes the spectrum using only the environments located at the center (centroids) of the clusters determined either by k-means (orange), CUR (green) or hierarchical clustering, in particular Agglomerative Clustering (AggC in maroon), where the remaining points within the cluster are approximated by the centroid. The number of clusters are determined by the CUR scores attributed to each SOAP vector. In general, the approximation of the spectra using clustering outperforms in terms accuracy the ML model trained on the Li_xC dataset, as could be observed in Figure 6.5 left (XPS for C 1s in aC) and right (XPS for C 1s in $Li_{12}C_{194}$) panels.

The quantification of the accuracy, expressed as the normalized XPS spectrum area difference (in %) are reported in Table 6.1, confirming the poor generalization of the models. These results suggest that new environments should be added to the Li_xC dataset in order to increase the flexibility of the model. For instance, there are no aC-like structures in the dataset, leading to the highest error among all the structures, quantified at 48.6 %. Nevertheless, generating approximate XPS spectra using clustering techniques offers the potential to significantly reduce the computational effort without sacrificing the accuracy much. It is worth noting that no optimizations were performed regarding the SOAP parameters or the metrics used in the clustering process, suggesting that there is room for improvement. We plan to explore these optimizations in the next future.

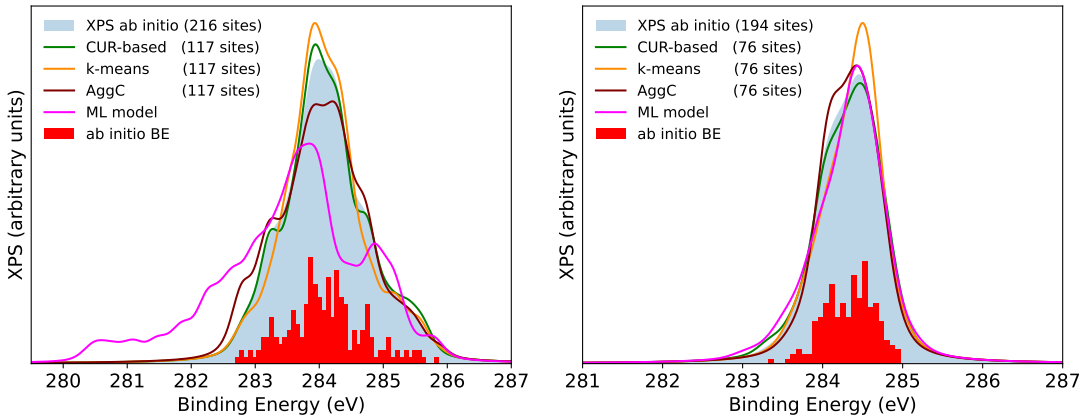

Figure 6.5: ML model predictions (magenta) on out of sample structures tested against full and approximated ab initio DFT benchmarks. The full ab initio results are represented by the BEs (red bars) and full XPS spectrum (blue area). The approximated ab initio computes the spectrum using the environments located at the center (centroids) of the clusters determined by k-means (orange), CUR (green) or hierarchical clustering (AggC in maroon), where the remaining points within the cluster are approximated by the centroid. (Left) Results for C 1s spectrum in amorphous C. (Right) Results for C 1s spectrum in $Li_{12}C_{194}$.

Figure 6.6 presents the training results for models predicting the Li 1s and Si 2p BEs in the Li_xSi dataset. The left panels show the prediction error of the BEs compared to those computed using ab initio DFT for the validation set, as a function of both the cutoff radius r_{cut} and the number of training structures included in the training set. Similar to what is observed in the models trained in the Li_xC dataset, r_{cut}=5.5 Å and r_{cut}=3.5 Å define the larger and smaller values of MAE, respectively. Once again, all models, regardless of the value of r_{cut}, exhibit a decreasing error trend as more environments are added during the training process. On the right panels, we present the correlation plots for Li 1s (top) and Si 2p (bottom) BEs between the best model predictions and the ab initio DFT reference for both the training (red triangles) and the validation set (blue triangles for Li and orange triangles for Si). We trained a total of 12 models by taking into account all the combinations of cutoff radii $r_{cut} = 2.5, 3.5, 4.5, 5.5$ Å, with radial scaling, and atomic density smoothing $\sigma_a = 0.3, 0.4, 0.5$ Å. The models with the best bias-variance trade-off, were selected as the best models. Such models were obtained by setting $r_{cut} = 3.5$ Å for Li 1s, $r_{cut} = 2.5$ Å for Si 2p and $\sigma_a = 0.3$ Å in both cases, generating an R^2 coefficient of 0.955 for Li 1s and 0.900 for Si 2p.

We follow the same procedure outlined earlier to assess the generalization capabilities of the models trained on the Li_xSi dataset. In this case, we tested the performance on a single out-of-sample structure, $Li_{96}Si_{128}$, from Ref. 308.

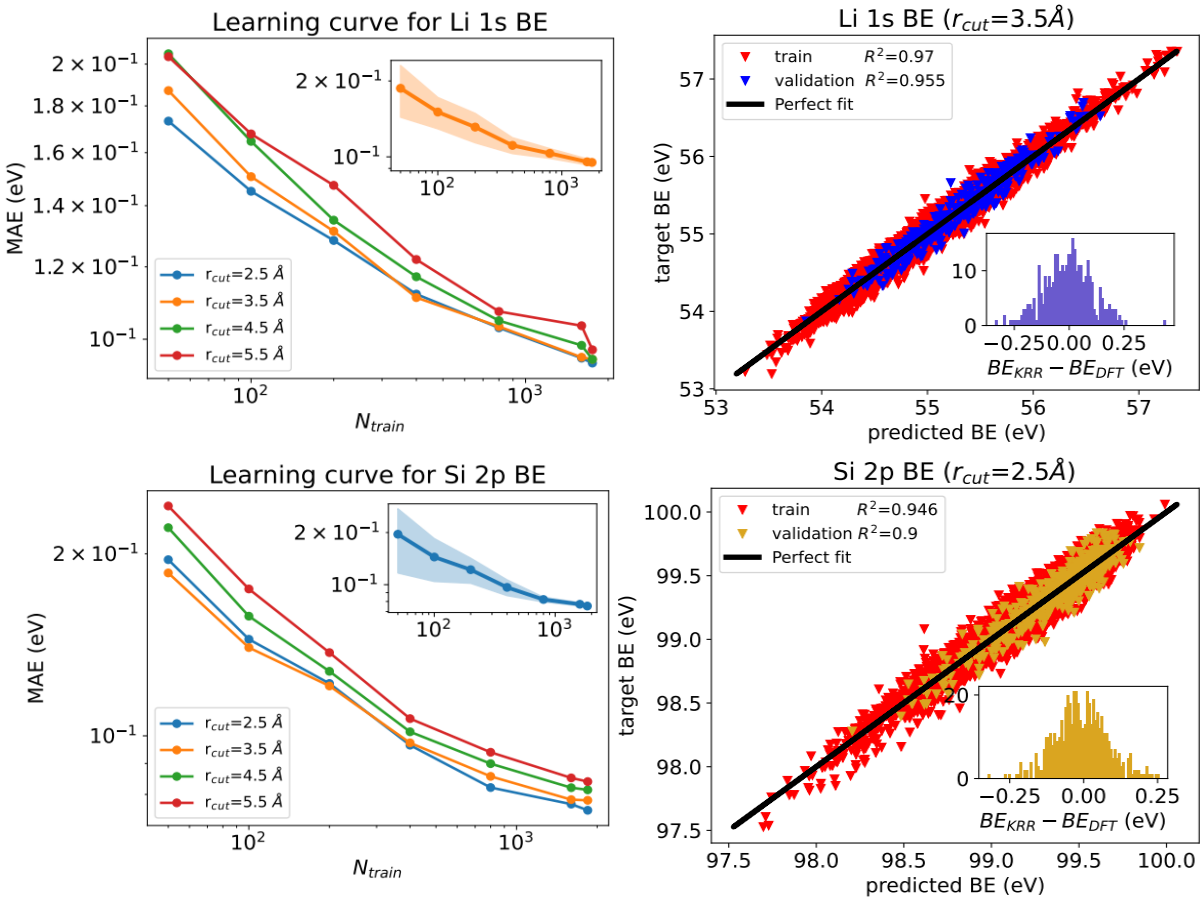

Figure 6.6: (Left) BE prediction error corresponding to the Li 1s (top) and Si 2p (bottom) gaussian KRR models for the Li_xSi validation set. The average MAE prediction error between BEs obtained with ML and ab initio DFT is shown for different local environment cutoff radii as a function of the training set size N_{train}. For the selected cutoff radius an inset was added containing the average BE error (dot and line) with its incertitude (shaded region) quantified as two times the standard deviation within the cross validation folds. (Right) Performance of the gaussian KRR models corresponding to the Li 1s (top) and Si 2p (bottom) for the Li_xSi train (red triangles) and validation (blue and orange triangles for Li and Si respectively) sets. The coefficient of determination R^2 for the relation between ML predictions and ab initio DFT (target) is reported in the legend. The histogram for the difference between BEs obtained with ML and ab initio DFT is shown as an inset.

Structure	error k-means (%)		error CUR (%)		error AggC (%)		error ML (%)	
	C 1s	Li 1s	C 1s	Li 1s	C 1s	Li 1s	C 1s	Li 1s
aC	10.6	-	4.7	-	10.8	-	48.6	-
$Li_{12}C_{194}$	12.2	0.9	5.2	1.3	6.6	1.3	10.2	46.0
$Li_{28}C_{196}$	3.2	2.5	5.9	6.0	15.4	7.5	13.8	56.1
$Li_{48}C_{216}$	8.8	8.1	12.3	24.9	13.6	13.2	26.6	27.2
	Si 2p	Li 1s	Si 2p	Li 1s	Si 2p	Li 1s	Si 2p	Li 1s
$Li_{96}Si_{128}$	4.8	2.6	4.3	6.0	4.0	12.9	1.9	5.9

Table 6.1: Absolute errors of the normalized XPS spectra constructed using clustering-approximated ab initio DFT and the ML model predictions for the structures in the external dataset. The errors (expressed as percentage) are referred to the full DFT calculation and computed as $\sum_i |y_i^{DFT} - y_i^{pred}|\Delta E_i$, where y_i reffers to the intensity of the XPS spectrum at the energy bin ΔE_i.

The ML predictions were benchmarked against both full and approximated ab initio DFT, as shown in Figure 6.7. Notably, the accuracy of the ML model matches or surpasses that achieved with clustering-based spectral approximations, as demonstrated in the left (XPS for Li 1s) and right (XPS for Si 2p) panels. The quantified accuracy, reported in Table 6.1, confirms the strong generalization performance of the trained models.

6.2 Stoichiometry map for Si-based anodes

One of the key limitations in the quantification process outlined in chapter 4 is the insufficient knowledge of the Li_xSi phases that form at various potentials. In the context of Si-based anodes, studies aimed at identifying the various Li_xSi phases are limited. From the experimental side, we found a study focused on an operando XPS investigation of the lithiation and delithiation processes in amorphous Si thin film electrodes [11]. However, we did not find a theoretical counterpart relating the XPS measurements with concentration of Li inside the Si matrix. The primary challenge lies in the formation of amorphous phases during the lithiation and delithiation of the electrode, which significantly increases the computational effort required for ab initio XPS simulations. For example, if we focus on some observable, A, its value in the case of an amorphous solid with no symmetry equivalence is analogous to an ensemble average of A over all the equally likely static configurations x_i. If we define A as the binding energy value at which the XPS spectrum is maximum for a given Li concentration x and static configuration x_i, $BE_{max}^{x_i}$, then for an amorphous Li_xSi phase we will have

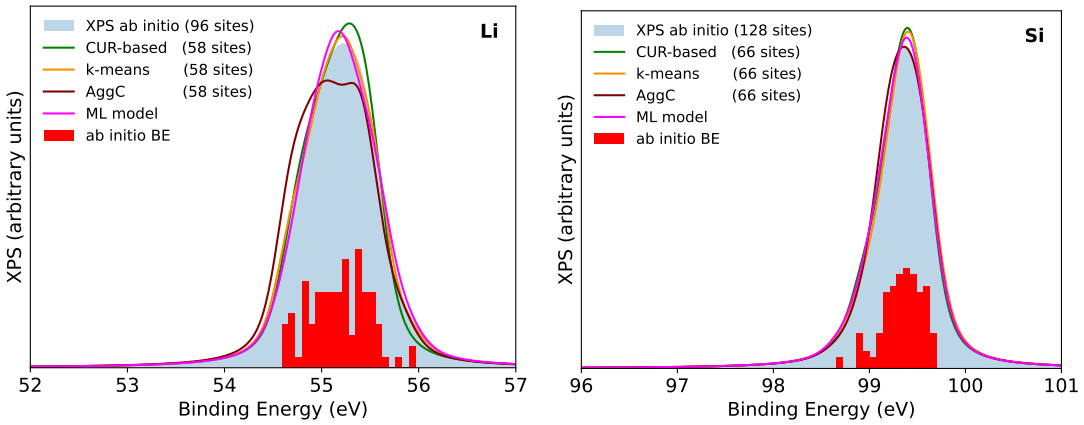

Figure 6.7: ML model predictions (magenta) on an out of sample structure tested against full and approximated ab initio DFT benchmarks. The full ab initio results are represented by the BEs (red bars) and full XPS spectrum (blue area). The approximated ab initio computes the spectrum using the environments located at the center (centroids) of the clusters determined by k-means (orange), CUR (green) or hierarchical clustering (AggC in maroon), where the remaining points within the cluster are approximated by the centroid. Results for Li 1s (left) and Si 2p (right) XPS spectrum in $Li_{96}Si_{128}$.

$$\mathrm{BE}^{x}_{max} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{BE}^{x_i}_{max}, \tag{6.1}$$

where N is the number of static configuration, or, in our case, structures, with the same Li concentration x. Determining BE^{x}_{max} requires computing N XPS spectra, which implies that generating a stoichiometry map with N_x Li concentrations would necessitate $N \times N_x$ XPS simulations. In this context, the ML model can be employed to efficiently compute all the XPS spectra for the structures in the Li_xSi dataset, significantly reducing the computational effort.

Figure 6.8 shows the energy position for the maximum of the Si 2p XPS spectrum against Li concentration. The red squares (green stars) represent BE^{x}_{max}, calculated from the ML predictions (ab initio DFT calculations) using all the (the training set) structures in the Li_xSi dataset with the same stoichiometry. The points are connected with a blue (orange) line and the uncertainty is represented by the blue area (error bar in the inset), corresponding to the standard deviation across all the structures with the same Li concentration. Interestingly, the BE^{x}_{max} predicted by the model shows a smooth variation with a small error in the $0 < x < 1.5$ region. This trend is less apparent for the ab initio results from the training set (green stars), suggesting that convergence has not yet been achieved with respect to the number of static configurations used.

We compared our results with operando XPS experiments presented in Fig-

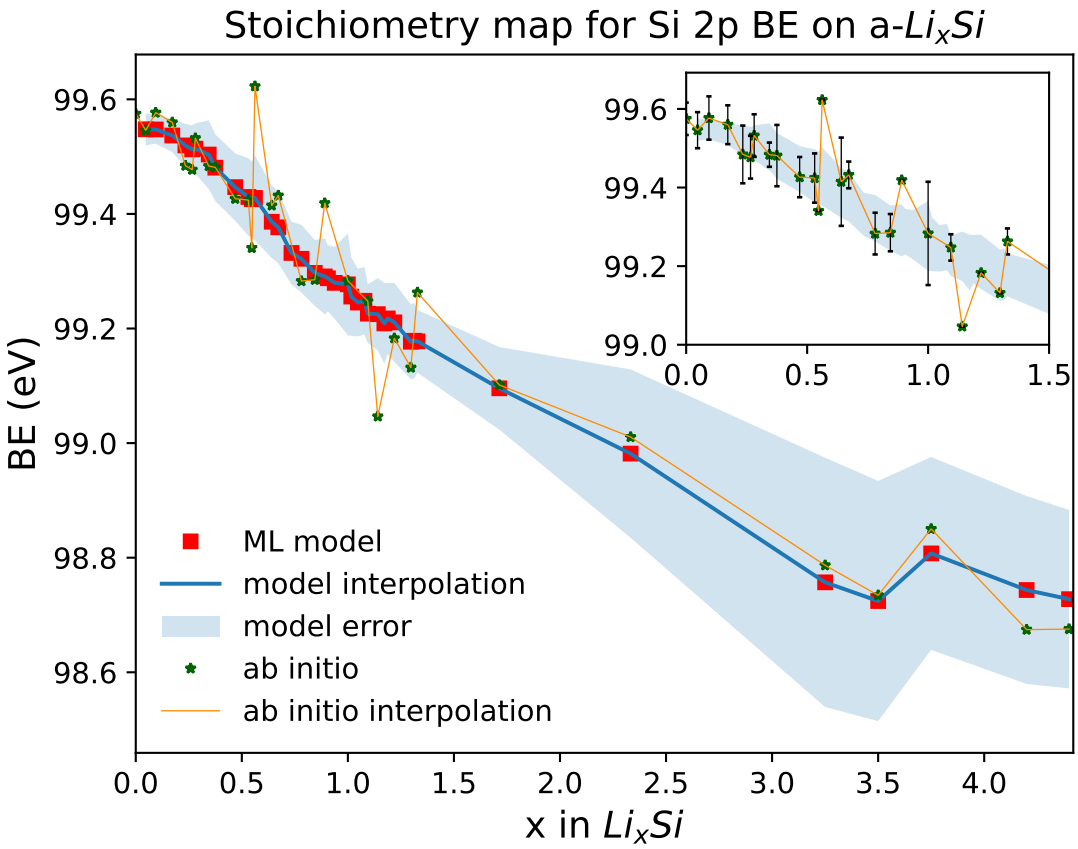

Figure 6.8: Energy position for the maximum of the Si 2p XPS spectrum against Li concentration for the Li_xSi dataset. The red squares (green stars) represent the average of the maximum obtained from the ML predictions (ab initio DFT calculations) using all the (the training set) structures in the Li_xSi dataset with the same stoichiometry. The points are connected with a blue (orange) line and the incertitude is represented by the blue area (error bar in the inset), corresponding to the standard deviation across all the structures with the same Li concentration.

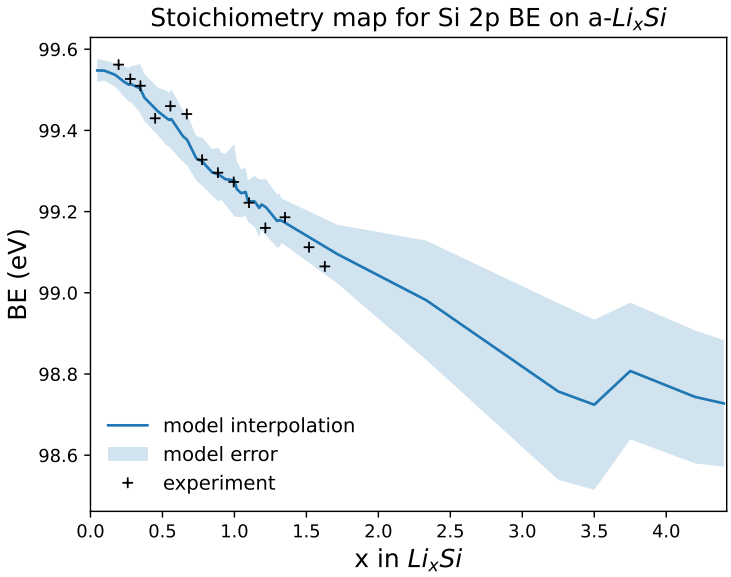

Figure 6.9: BE position for the maximum of the Si 2p XPS spectrum against Li concentration. ML model (line) and experiment (+) from Ref. 11.

ure 6.9. The experimental points (+) corresponds to the delithiation process in the region $0< x < 2$, rich in amorphous Li_xSi phases [11]. Since referencing and calibrating BE scales can be challenging in dynamic operando environments, especially with changes in sample charging, or shifts in electronic structure, the experimental points were shifted 2 eV upwards in order to coincide in BE with post-mortem experiments [309, 310]. The results obtained with the ML model reproduce nicely the trend measured in the experiment, validating the use of surrogate models in the construction of stoichiometry maps in amorphous systems.

6.3 Conclusions

In this Chapter, we presented our work on trained ML models, based on gaussian KRR and SOAP descriptors, for the prediction of XPS spectra. A comprehensive approach, integrating an automated AiiDA workflow for first-principles XPS simulation with sample sub-selection via FPS, is employed to generate the critical amount of data needed for the training process. The ML models were trained on two representative datasets of Li_xC and Li_xSi structures, achieving an accuracy of 0.1 eV, i.e., matching the typical experimental resolution of XPS. Generalization tests and benchmarks against approximated methods were performed. The results reveal poor generalization for the models trained on the Li_xC dataset. In contrast, the models trained on Li_xSi demonstrated greater flexibility, enabling the construction of a stoichiometry map with results comparable to those from operando XPS experiments.

Conclusions

In this Thesis, I explored multiple facets of materials characterization through core-level spectroscopies assisted with computational modeling, contributing insights into the analysis of hydrogenated graphene, silicon-based anodes, and machine learning applications in XPS prediction.

First, the simulations I performed on hydrogenated graphene provided a deeper understanding of its electronic and structural properties, particularly through the relationship between sp^2/sp^3 hybridization and core-level shifts. These findings are pivotal in interpreting experimental XPS data and validating structural transformations during hydrogenation, thus offering a framework for precise characterization of different hydrogenation states.

Second, the study of silicon nanoparticle anodes using XRS revealed unique chemical signatures tied to the SEI formation and its evolution during lithiation and delithiation. Through a multi-edge semi-quantitative approach, we identified key SEI components such as Li_2CO_3, $LiPF_6$, and LiF, and quantified irreversible lithium loss, correlating these findings with the observed first-cycle capacity loss in silicon anodes. These insights align with previous studies and underscore XRS as a powerful tool for bulk analysis of composite materials, complementing surface-limited XPS. Additionally, I demonstrated the practical application of ab initio DFT simulations in constructing reference datasets to assist experimental interpretation and seamlessly integrate them into data fitting and quantification processes.

To automate the construction of reference datasets, I developed a robust workflow for ab initio DFT XPS computation, providing a high-throughput, modular platform for simulating spectra with a level of accuracy comparable to existing implementations. This Python-based tool enables users to perform tailored simulations for finite and crystalline systems, proving highly adaptable for various material studies. The workflow is fully integrated into AiiDA-QE and AiiDAlab, making it easily accessible to a wider community of researchers.

Lastly, I trained surrogate models for computing the XPS spectrum on Li_xC and Li_xSi datasets, achieving a prediction accuracy consistent with experimental XPS resolution. For this purpose, I implemented a comprehensive approach, integrating the XPS workflow, for first-principles XPS simulation, with sample subselection via FPS, to generate the critical amount of data needed for training pro-

cess. I evaluated critically the generalization capabilities of the machine learning models against clustering-approximated approaches, and only those trained on the Li_xSi dataset showed enhanced flexibility. I applied the ML model trained for predicting Si 2p BEs in the creation of a stoichiometry map comparable to operando XPS experiments. These developments highlight the potential of ML methods for efficient spectral prediction, particularly in complex systems like amorphous Li_xSi.

In conclusion, this study advanced our understanding of material characterization by core-level spectroscopies, combining experimental and computational approaches. The contributions from this work open new possibilities for future research by leveraging the automation of ab initio core-level spectroscopy calculations with machine learning methods, enhancing data generation capabilities needed for spectroscopic analysis in materials science. Looking ahead, we plan to extend the methodology developed in this work to XAS and XRS, providing a scalable framework for the automated study of core-level spectroscopies.

List of publications and works in preparation

- Maria Grazia Betti, Dario Marchiani, Andrea Tonelli, Marco Sbroscia, Elena Blundo, Marta De Luca, Antonio Polimeni, Riccardo Frisenda, Carlo Mariani, Samuel Jeong, Yoshikazu Ito, Nicola Cavani, Roberto Biagi, Peter N.O. Gillespie, **Michael A. Hernandez Bertran**, Miki Bonacci, Elisa Molinari, Valentina De Renzi, Deborah Prezzi, "*Dielectric response and excitations of hydrogenated free-standing graphene*", Carbon Trends 12, 100274 (2023); doi: 10.1016/j.cartre.2023.100274;

- **Michael A. Hernandez Bertran**, Diana Zapata Dominguez, Christopher Berhaut, Alessandro Longo, Christoph Sahle, Chiara Cavallari, Ivan Marri, Nathalie Herlin-Boime, Elisa Molinari, Stéphanie Pouget, Deborah Prezzi, Sandrine Lyonnard, "*Understanding the irreversible lithium loss in silicon anodes using multi-edge X-ray scattering analysis*". Submitted to Chem. Mater.; arXiv: 10.48550/arXiv.2410.05794;

- Peter N. O. Gillespie, **Michael A. Hernandez Bertran**, Xing Wang, Giovanni Pizzi, Elisa Molinari, and Deborah Prezzi, "*Automated Workflows for Core-Level Spectroscopy Simulation*". *Preprint*;

- Xing Wang, Jusong Yu, Aliaksandr V. Yakutovich, Andres Ortega-Guerrero, Simon Adorf, Daniel Hollas, Miki Bonacci, Edan Bainglass, Timo Reents, Marnik Bercx, Christopher J. Sewell, Lorenzo Bastonero, Ifeanyi J. Onuorah, Pietro Bonfà, Peter N. O. Gillespie, **Michael A. Hernandez Bertran**, Deborah Prezzi, Iurii Timrov, Carlo A. Pignedoli, Nicola Marzari, and Giovanni Pizzi, "*Making atomistic materials calculations accessible with AiiDAlab*". *Preprint.*

Appendices

6.4 Multi-edge XRS data

6.5 Description of the C, O and F K-edges features

6.5.1 C K-edge features

The reference samples used for describing the C K-edge of the two electrode states in the main text are Li_2CO_3 and CMC+Super P. The first is known as one of the main components of SEI film. The second one is a mixture of binder and conductive additive used for preparing the electrode. For Li_2CO_3 the main features in the spectrum are attributed to the CO_3^{2-} ion. The peak positioned at 290.4 eV is associated with transitions from C=O 1s to antibonding π^* states. Additional broader features located at 297.8 and 300.7 eV are related with transitions from both C−O and C=O 1s to σ^* antibonding orbitals. Two peaks were identified for the CMC+Super P reference at 285.4 and 292.4 eV corresponding to transitions from C=C 1s to π^* and C–C 1s to σ^* respectively. Peaks associated with the carbonate group in the CMC+Super P spectra are suppressed as a consequence of the chemical bond between the CMC binder and the active particle [193, 194].

6.5.2 O K-edge features

The reference samples used for describing the O K-edge of the electrode states are Li_2CO_3 and the pristine electrode. The reference pristine has 50% of Si NPs blended with 25% CMC and 25% Super P deposited in Cu foil. It is expected that the pristine electrode signal contains features from both SiO_2 and CMC+Super P. The first is formed naturally as an external layer when the Si NPs are in contact with O, reaching up to 20% of the NP size at room temperature [311]. O K-edge spectrum for the pristine electrode resembles the one reported for SiO_2 films [8, 4] confirming the oxidation process on the surface of the NPs. This supports the findings for the C K-edge where the CMC+Super P spectra didn't show significant contribution from the C=O and C−O groups.

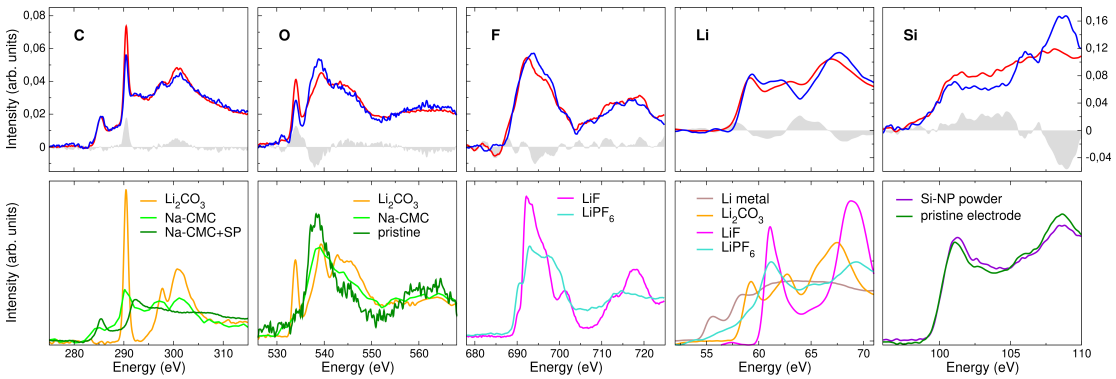

Figure 6.10: (Top panels) Experimental non-resonant inelastic X-ray scattering spectra of c-Si NPs electrodes for the C, O, F and Li K-edges and the Si $L_{2,3}$-edge in the lithiated (red) and delithiated (blue) SoC, as well as their difference (grey area) to highlight the main changes between the two SoC. (Bottom panels) XRS spectra for the reference compounds measured in the same conditions and set-up.

As the case of the C K-edge, the principal features for the O K-edge in Li_2CO_3 are related to environments containing ${CO_3}^{2-}$. A first peak located at 533.9 eV is associated with transitions from C=O 1s to π^* orbitals. Other features present at 539.3, 542.8 and 545.6 eV are related with transitions from C−O and C=O 1s to σ^* states. The pristine electrode spectrum has contributions from the principal characteristics of the SiO_2 O K-edge. On the white line two peaks located at 538.2 and 544.0 eV are identified, both attributed to transitions from O 1s to σ^* states in an environment with O 2p - Si 3sp hybridization [8, 312]. The feature at 544.0 eV depends on the structural order, ranging from near-negligible intensity in thermally grown SiO_2 films [4, 8] to a well-resolved peak in a more crystalline phase, like α−SiO_2 [313]. A last broader feature placed at 559 eV arise from transitions to σ^* states in environments containing O 2p states hybridized with Si 3d orbitals [312].

6.5.3 F K-edge features

In order to describe the F K-edge spectra, $LiPF_6$ and LiF were selected as references. Both compounds are the main fluorinated components of the SEI, the first is present in the electrolyte and the second appears after the first cycle as a decomposition product of the former [176]. The principal features of LiF (**a**-**f**) are originated by dipole-allowed transitions to final states with similar spatial symmetry [314, 315]. Dipole-forbidden transitions give raise to the excitonic peak $\mathbf{a}^*$ located at 690.2 eV contributing in a less extent to the features **e** and **f** [314, 116]. For $LiPF_6$ the main features in the spectrum are attributed to the ${PF_6}^-$ ion. According to molecular orbital description of octahedral coordinated molecules, peaks $\mathbf{a_1}$, $\mathbf{b_1}$ and ($\mathbf{c_1}$, $\mathbf{d_1}$) corresponds to transitions into F 2p - P 3s, F 2p - P 3p

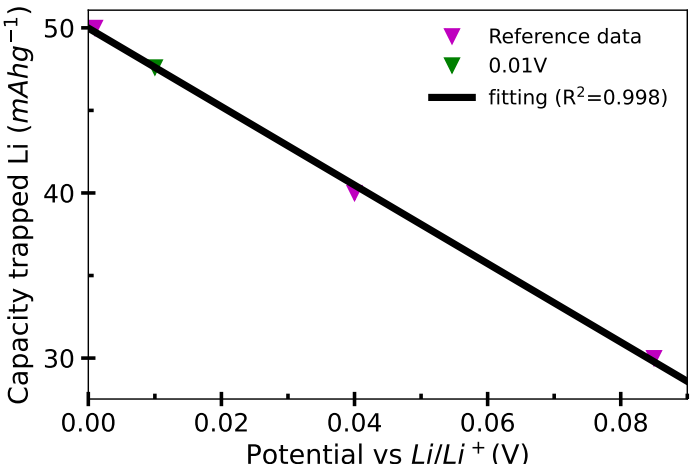

Figure 6.11: Linear fitting of the recovered capacity from Li trapped in Si NP (data taken from Ref. 12) vs. the final lithiation potential. The lithiation potential used in this work, 10 mV, implies that 17% of the total capacity loss corresponds to the Li trapped in Li-Si alloys.

and F $2p$ - P $3d$ hybridized orbitals respectively [216, 316]. The shoulder identified by $\mathbf{a_1^*}$ was reported previously for powder samples [217, 317, 318], however this feature is attenuated in $LiPF_6$-based electrolytes ($LiPF_6$ dissolved in EC:DMC) [217, 216]. This variation may be consequence of the different local environments and could serve as a fingerprint for distinguish between crystalline $LiPF_6$ (SEI) and [PF_6^-]-bearing solvation shell structures (electrolyte).

6.6 Estimate of the Li trapped in Si NPs

In the main text, we make use of recent titration-gas chromatography experiments [241] conducted on similar Si NPs with a Na-CMC binder and acetylene carbon black as the conductive additive to estimate the amount of Li trapped in Li–Si alloys. As a sanity check, we propose here an alternative approach based on galvanostatic measurements. We make use of the galvanostatic measurements presented in Ref. 12, where similar samples are held at decreasing low cutoff voltages (from 600 to 1 mV) and then fully delithiated at a cell potential of 2.0 V. The values of the recovered capacity for each lithiation potential, extracted from Figure 3a of Ref. 12, are reported in figure 6.11 (purple triangles). By using a linear fitting of these values, we interpolate the capacity loss corresponding to the lithiation potential used in this work, i.e. 10 mV, which corresponds to 17% of the total capacity loss. This agrees quite well with the value obtained from titration-gas chromatography [241], i.e. 19%.

Bibliography

[1] W.L. Jolly, K.D. Bomben, and C.J. Eyermann. Core-electron binding energies for gaseous atoms and molecules. Atomic Data and Nuclear Data Tables, 31(3):433–493, 1984.

[2] Canadian Light Source XAS Database (XASDB). sample id: idbzn2q4. 2024.

[3] E. de Clermont Gallerande, D. Cabaret, G. Lelong, C. Brouder, M.-B. Attaiaa, L. Paulatto, K. Gilmore, Ch J. Sahle, and G. Radtke. First-principles modeling of x-ray raman scattering spectra. Phys. Rev. B, 98:214104, 12 2018.

[4] S. Yu. Turishchev, E. V. Parinova, A. K. Pisliaruk, D. A. Koyuda, D. Yermukhamed, T. Ming, R. Ovsyannikov, D. Smirnov, A. Makarova, and V. Sivakov. Surface deep profile synchrotron studies of mechanically modified top-down silicon nanowires array using ultrasoft x-ray absorption near edge structure spectroscopy. Sci. Rep., 9(1):8066, 2019.

[5] M.B. Robin, I. Ishii, R. McLaren, and A.P. Hitchcock. Fluorination effects on the inner-shell spectra of unsaturated molecules. Journal of Electron Spectroscopy and Related Phenomena, 47:53–92, 1988.

[6] I. Ishii and A.P. Hitchcock. The oscillator strengths for c1s and o1s excitation of some saturated and unsaturated organic alcohols, acids and esters. Journal of Electron Spectroscopy and Related Phenomena, 46(1):55–84, 1988.

[7] AP Hitchcock. Gas phase core excitation database. http://unicorn.mcmaster. ca/corex/cedb-title. html, 2003.

[8] H. M. Tsai, S. C. Ray, C. W. Pao, J. W. Chiou, C. L. Huang, C. H. Du, W. F. Pong, M.-H. Tsai, A. Fukano, and H. Oyanagi. Enhancement of Si–O hybridization in low-temperature grown ultraviolet photo-oxidized SiO2 film observed by x-ray absorption and photoemission spectroscopy. J. Appl. Phys., 103(1):013704, 01 2008.

[9] L. Yang, B. Abeles, W. Eberhardt, H. Stasiewski, and D. Sondericker. Photoemission spectroscopy of heterojunctions of hydrogenated amorphous silicon with silicon oxide and nitride. Phys. Rev. B, 39:3801–3816, Feb 1989.

[10] V Carravetta, G Iucci, A Ferri, M.V Russo, S Stranges, M de Simone, and G Polzonetti. Synchrotron radiation photoemission study of some π-conjugated alkynes in the gas phase: Experiment and theory. Chemical Physics, 264(2):175–186, 2001.

[11] Raimu Endo, Tsuyoshi Ohnishi, Kazunori Takada, and Takuya Masuda. Electrochemical lithiation and delithiation in amorphous si thin film electrodes studied by operando x-ray photoelectron spectroscopy. The Journal of Physical Chemistry Letters, 13(31):7363–7370, 2022.

[12] Alison L. Michan, Michal Leskes, and Clare P. Grey. Voltage dependent solid electrolyte interphase formation in silicon electrodes: Monitoring the formation of organic decomposition products. Chem. Mater., 28(1):385–398, 2016.

[13] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. Annalen der Physik, 389(20):457–484, 1927.

[14] Ferenc Krausz. The birth of attosecond physics and its coming of age. Physica Scripta, 91(6):063011, may 2016.

[15] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. Mathematical Proceedings of the Cambridge Philosophical Society, 24(1):89–110, 1928.

[16] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. Zeitschrift für Physik, 61(1):126–148, Jan 1930.

[17] C. David Sherrill and Henry F. Schaefer. The configuration interaction method: Advances in highly correlated approaches. volume 34 of Advances in Quantum Chemistry, pages 143–269. Academic Press, 1999.

[18] F. Coester and H. Kümmel. Short-range correlations in nuclear wave functions. Nuclear Physics, 17:477–485, 1960.

[19] Jiří Čížek. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. The Journal of Chemical Physics, 45(11):4256–4266, 12 1966.

[20] M Peter Nightingale and Cyrus J Umrigar. Quantum Monte Carlo methods in physics and chemistry. Number 525. Springer Science & Business Media, 1998.

[21] L. H. Thomas. The calculation of atomic fields. Mathematical Proceedings of the Cambridge Philosophical Society, 23(5):542–548, 1927.

[22] Enrico Fermi. Statistical method to determine some properties of atoms. Rend. Accad. Naz. Lincei, 6(602-607):5, 1927.

[23] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. Phys. Rev., 136:B864–B871, Nov 1964.

[24] Mel Levy. Electron densities in search of hamiltonians. Phys. Rev. A, 26:1200–1208, Sep 1982.

[25] EH Lieb. Physics as Natural Philosophy ed A Shimony and H Feshbach. MIT Press, Cambridge, MA, 1982.

[26] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. Phys. Rev., 140:A1133–A1138, Nov 1965.

[27] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. Phys. Rev. Lett., 45:566–569, Aug 1980.

[28] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. Phys. Rev. B, 23:5048–5079, May 1981.

[29] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. Phys. Rev. Lett., 77:3865–3868, Oct 1996.

[30] Richard M Martin, Lucia Reining, and David M Ceperley. Interacting electrons. Cambridge University Press, 2016.

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

[32] N.W. Ashcroft and N.D. Mermin. Solid State Physics. Saunders College Publishing, Fort Worth, 1976.

[33] Enrico Fermi. Sopra lo spostamento per pressione delle righe elevate delle serie spettrali. Il Nuovo Cimento (1924-1942), 11(3):157–166, Mar 1934.

[34] James C. Phillips and Leonard Kleinman. New method for calculating wave functions in crystals and molecules. Phys. Rev., 116:287–294, Oct 1959.

[35] Leonard Kleinman and James C. Phillips. Crystal potential and energy bands of semiconductors. i. self-consistent calculations for diamond. Phys. Rev., 116:880–884, Nov 1959.

[36] 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. J. Phys.: Condens. Matter, 21(39):395502, sep 2009.

[37] P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo, A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, E Küçükbenli, M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V Nguyen, A Otero de-la Roza, L Paulatto, S Poncé, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with quantum espresso. J. Phys.: Condens. Matter, 29(46):465901, oct 2017.

[38] Paolo Giannozzi, Oscar Baseggio, Pietro Bonfà, Davide Brunato, Roberto Car, Ivan Carnimeo, Carlo Cavazzoni, Stefano de Gironcoli, Pietro Delugas, Fabrizio Ferrari Ruffino, Andrea Ferretti, Nicola Marzari, Iurii Timrov, Andrea Urru, and Stefano Baroni. Quantum ESPRESSO toward the exascale. J. Chem. Phys., 152(15):154105, 04 2020.

[39] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Phys. Rev. B, 41:7892–7895, Apr 1990.

[40] Kari Laasonen, Roberto Car, Changyol Lee, and David Vanderbilt. Implementation of ultrasoft pseudopotentials in ab initio molecular dynamics. Phys. Rev. B, 43:6796–6799, Mar 1991.

[41] Kari Laasonen, Alfredo Pasquarello, Roberto Car, Changyol Lee, and David Vanderbilt. Car-parrinello molecular dynamics with vanderbilt ultrasoft pseudopotentials. Phys. Rev. B, 47:10142–10153, Apr 1993.

[42] Kevin P Murphy. Machine learning: a probabilistic perspective. MIT press, 2012.

[43] Jonathan Schmidt, Mário RG Marques, Silvana Botti, and Miguel AL Marques. Recent advances and applications of machine learning in solid-state materials science. npj computational materials, 5(1):83, 2019.

[44] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11):559–572, 1901.

[45] Harold Hotelling. Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6):417, 1933.

[46] Michael W Mahoney and Petros Drineas. Cur matrix decompositions for improved data analysis. Proceedings of the National Academy of Sciences, 106(3):697–702, 2009.

[47] Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. Proceedings of the National Academy of Sciences, 108(32):13023–13028, 2011.

[48] J Macqueen. Some methods for classification and analysis of multivariate observations. In Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability/University of California Press, 1967.

[49] Leonard Kaufman. Partitioning around medoids (program pam). Finding groups in data, 344:68–125, 1990.

[50] Fionn Murtagh and Pedro Contreras. Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(1):86–97, 2012.

[51] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55–67, 1970.

[52] J Shawe-Taylor. Kernel methods for pattern analysis. Cambridge University Press google schola, 2:181–201, 2004.

[53] Felix Musil, Andrea Grisafi, Albert P Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Physics-inspired structural representations for molecules and materials. Chemical Reviews, 121(16):9759–9815, 2021.

[54] Albert P Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. Physical Review B—Condensed Matter and Materials Physics, 87(18):184115, 2013.

[55] Michael J Willatt, Félix Musil, and Michele Ceriotti. Atom-density representations for machine learning. The Journal of chemical physics, 150(15), 2019.

[56] D Nanda Gopala Krishna and John Philip. Review on surface-characterization applications of x-ray photoelectron spectroscopy (xps): Recent developments and challenges. Applied Surface Science Advances, 12:100332, 2022.

[57] Didem Ketenoglu. A general overview and comparative interpretation on element-specific x-ray spectroscopy techniques: Xps, xas, and xrs. X-Ray Spectrometry, 51(5-6):422–443, 2022.

[58] Marcus Fehse, Antonella Iadecola, Laura Simonelli, Alessandro Longo, and Lorenzo Stievano. The rise of x-ray spectroscopies for unveiling the functional mechanisms in batteries. Phys. Chem. Chem. Phys., 23:23445–23465, 2021.

[59] M. Weissbluth. Atoms and molecules. Academic Press, Inc, United States, 1978.

[60] M. Blume. Magnetic scattering of x rays (invited). Journal of Applied Physics, 57(8):3615–3618, 04 1985.

[61] G. Greczynski and L. Hultman. X-ray photoelectron spectroscopy: Towards reliable binding energy referencing. Progress in Materials Science, 107:100591, 2020.

[62] Benedikt P Klein, Samuel J Hall, and Reinhard J Maurer. The nuts and bolts of core-hole constrained ab initio simulation for k-shell x-ray photoemission and absorption spectra. Journal of Physics: Condensed Matter, 33(15):154005, feb 2021.

[63] T Koopmans. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. Physica, 1(1):104–113, 1934.

[64] A. R. Williams and N. D. Lang. Core-level binding-energy shifts in metals. Phys. Rev. Lett., 40:954–957, Apr 1978.

[65] Noèlia Pueyo Bellafont, Paul S. Bagus, and Francesc Illas. Prediction of core level binding energies in density functional theory: Rigorous definition of initial and final state contributions and implications on the physical meaning of Kohn-Sham energies. The Journal of Chemical Physics, 142(21):214102, 06 2015.

[66] Xuechen Zheng and Lan Cheng. Performance of delta-coupled-cluster methods for calculations of core-ionization energies of first-row elements. Journal of Chemical Theory and Computation, 15(9):4945–4955, 2019. PMID: 31365823.

[67] Junzi Liu, Devin Matthews, Sonia Coriani, and Lan Cheng. Benchmark calculations of k-edge ionization energies for first-row elements using scalar-relativistic core–valence-separated equation-of-motion coupled-cluster methods. Journal of Chemical Theory and Computation, 15(3):1642–1651, 2019. PMID: 30702889.

[68] J. J. Kas, F. D. Vila, J. J. Rehr, and S. A. Chambers. Real-time cumulant approach for charge-transfer satellites in x-ray photoemission spectra. Phys. Rev. B, 91:121112, Mar 2015.

[69] Ulf Ekström, Patrick Norman, Vincenzo Carravetta, and Hans Ågren. Polarization propagator for x-ray spectra. Phys. Rev. Lett., 97:143001, Oct 2006.

[70] Dorothea Golze, Jan Wilhelm, Michiel J. van Setten, and Patrick Rinke. Core-level binding energies from gw: An efficient full-frequency approach within a localized basis. Journal of Chemical Theory and Computation, 14(9):4856–4869, 2018. PMID: 30092140.

[71] P. S. Bagus. Self-consistent-field wave functions for hole states of some ne-like and ar-like ions. Phys. Rev., 139:A619–A634, Aug 1965.

[72] Noèlia Pueyo Bellafont, Paul S. Bagus, and Francesc Illas. Prediction of core level binding energies in density functional theory: Rigorous definition of initial and final state contributions and implications on the physical meaning of Kohn-Sham energies. The Journal of Chemical Physics, 142(21):214102, 06 2015.

[73] J. Matthias Kahk, Georg S. Michelitsch, Reinhard J. Maurer, Karsten Reuter, and Johannes Lischner. Core electron binding energies in solids from periodic all-electron δ-self-consistent-field calculations. J. Phys. Chem. Lett., 12(38):9353–9359, 2021.

[74] M.P. Ljungberg, J.J. Mortensen, and L.G.M. Pettersson. An implementation of core level spectroscopies in a real space projector augmented wave density functional theory code. J. Electron Spectros. Relat. Phenomena, 184(8):427–439, 2011.

[75] Michael Walter, Michael Moseler, and Lars Pastewka. Offset-corrected Δ-kohn-sham scheme for semiempirical prediction of absolute x-ray photoelectron energies in molecules and solids. Phys. Rev. B, 94:041112, 2016.

[76] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Computational materials science, 6(1):15–50, 1996.

[77] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical review B, 54(16):11169, 1996.

[78] Stewart J Clark, Matthew D Segall, Chris J Pickard, Phil J Hasnip, Matt IJ Probert, Keith Refson, and Mike C Payne. First principles methods using castep. Zeitschrift für kristallographie-crystalline materials, 220(5-6):567–570, 2005.

[79] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. Computer Physics Communications, 167(2):103–128, 2005.

[80] J. F. Janak. Proof that $\frac{\partial e}{\partial n_i} = \epsilon$ in density-functional theory. Phys. Rev. B, 18:7165–7168, Dec 1978.

[81] Subrata Jana and John M. Herbert. Slater transition methods for core-level electron binding energies. The Journal of Chemical Physics, 158(9):094111, 03 2023.

[82] David Prendergast and Giulia Galli. X-ray absorption spectra of water from first principles calculations. Phys. Rev. Lett., 96:215502, May 2006.

[83] Teruyasu Mizoguchi, Isao Tanaka, Shang-Peng Gao, and Chris J Pickard. First-principles calculation of spectral features, chemical shift and absolute threshold of elnes and xanes using a plane wave pseudopotential method. J. Phys. Condens. Matter, 21(10):104204, 2009.

[84] Glenn J. Martyna and Mark E. Tuckerman. A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. J. Chem. Phys., 110(6):2810–2821, 02 1999.

[85] Grant Bunker. Introduction to XAFS: a practical guide to X-ray absorption fine structure spectroscopy. Cambridge University Press, 2010.

[86] Wolfgang Grünert and Konstantin Klementiev. X-ray absorption spectroscopy principles and practical use in materials analysis. Physical Sciences Reviews, 5(4):20170181, 2020.

[87] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler. Chemistry with adf. Journal of Computational Chemistry, 22(9):931–967, 2001.

[88] Frank Neese. The orca program system. WIREs Computational Molecular Science, 2(1):73–78, 2012.

[89] John J. Rehr, Joshua J. Kas, Fernando D. Vila, Micah P. Prange, and Kevin Jorissen. Parameter-free calculations of x-ray spectra with feff9. Phys. Chem. Chem. Phys., 12:5503–5513, 2010.

[90] K. Schwarz, P. Blaha, and G.K.H. Madsen. Electronic structure calculations of solids using the wien2k package for material sciences. Computer Physics Communications, 147(1):71–76, 2002. Proceedings of the Europhysics Conference on Computational Physics Computational Modeling and Simulation of Complex Systems.

[91] Federica Frati, Myrtille O. J. Y. Hunault, and Frank M. F. de Groot. Oxygen k-edge x-ray absorption spectra. Chemical Reviews, 120(9):4056–4110, May 2020.

[92] Oana Bunău and Yves Joly. Time-dependent density functional theory applied to x-ray absorption spectroscopy. Phys. Rev. B, 85:155121, Apr 2012.

[93] O Bunău, Aline Y Ramos, and Yves Joly. The fdmnes code. 2021.

[94] J. Vinson, J. J. Rehr, J. J. Kas, and E. L. Shirley. Bethe-salpeter equation calculations of core excitation spectra. Phys. Rev. B, 83:115106, Mar 2011.

[95] J. van Elp and Arata Tanaka. Threshold electronic structure at the oxygen k edge of $3d$-transition-metal oxides: A configuration interaction approach. Phys. Rev. B, 60:5331–5339, Aug 1999.

[96] Sonia Coriani, Ove Christiansen, Thomas Fransson, and Patrick Norman. Coupled-cluster response theory for near-edge x-ray-absorption fine structure of atoms and molecules. Phys. Rev. A, 85:022507, Feb 2012.

[97] Marcel Nooijen and Rodney J. Bartlett. Description of core-excitation spectra by the open-shell electron-attachment equation-of-motion coupled cluster method. The Journal of Chemical Physics, 102(17):6735–6756, 05 1995.

[98] Mathieu Taillefumier, Delphine Cabaret, Anne-Marie Flank, and Francesco Mauri. X-ray absorption near-edge structure calculations with the pseudopotentials: Application to the k edge in diamond and α-quartz. Phys. Rev. B, 66:195107, Nov 2002.

[99] C. Gougoussis, M. Calandra, A. Seitsonen, Ch. Brouder, A. Shukla, and F. Mauri. Intrinsic charge transfer gap in nio from ni k-edge x-ray absorption spectroscopy. Phys. Rev. B, 79:045118, 1 2009.

[100] Christos Gougoussis, Matteo Calandra, Ari P. Seitsonen, and Francesco Mauri. First-principles calculations of x-ray absorption in a scheme based on ultrasoft pseudopotentials: From α-quartz to high-t_c compounds. Phys. Rev. B, 80:075102, 8 2009.

[101] Oana Bunău and Matteo Calandra. Projector augmented wave calculation of x-ray absorption spectra at the $l_{2,3}$ edges. Phys. Rev. B, 87:205105, 5 2013.

[102] P. E. Blöchl. Projector augmented-wave method. Phys. Rev. B, 50:17953–17979, Dec 1994.

[103] Cornelius Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. J. Res. Natl. Bur. Stand. B, 45:255–282, 1950.

[104] Cornelius Lanczos. Solution of Systems of Linear Equations by Minimized Iterations. J. Res. Natl. Bur. Stand. B, 49:33–53, 1952.

[105] M. Leetmaa, M.P. Ljungberg, A. Lyubartsev, A. Nilsson, and L.G.M. Pettersson. Theoretical approximations to x-ray absorption spectroscopy of liquid water and ice. Journal of Electron Spectroscopy and Related Phenomena, 177(2):135–157, 2010. Water and Hydrogen Bonds.

[106] Guido Fratesi, Valeria Lanzilotto, Luca Floreano, and Gian Paolo Brivio. Azimuthal dichroism in near-edge x-ray absorption fine structure spectra of planar molecules. The Journal of Physical Chemistry C, 117(13):6632–6638, 2013.

[107] Guido Fratesi, Valeria Lanzilotto, Luca Floreano, and Gian Paolo Brivio. Correction to "azimuthal dichroism in near-edge x-ray absorption fine structure spectra of planar molecules". The Journal of Physical Chemistry C, 118(4):2236–2236, 2014.

[108] L. Triguero, L. G. M. Pettersson, and H. Ågren. Calculations of near-edge x-ray-absorption spectra of gas-phase and chemisorbed molecules by means of density-functional and transition-potential theory. Phys. Rev. B, 58:8097–8110, Sep 1998.

[109] Benedikt P. Klein, Nadine J. van der Heijden, Stefan R. Kachel, Markus Franke, Claudio K. Krug, Katharina K. Greulich, Lukas Ruppenthal, Philipp Müller, Phil Rosenow, Shayan Parhizkar, François C. Bocquet, Martin Schmid, Wolfgang Hieringer, Reinhard J. Maurer, Ralf Tonner, Christian Kumpf, Ingmar Swart, and J. Michael Gottfried. Molecular topology and the surface chemical bond: Alternant versus nonalternant aromatic systems as functional structural elements. Phys. Rev. X, 9:011030, Feb 2019.

[110] Benedikt P. Klein, S. Elizabeth Harman, Lukas Ruppenthal, Griffin M. Ruehl, Samuel J. Hall, Spencer J. Carey, Jan Herritsch, Martin Schmid, Reinhard J. Maurer, Ralf Tonner, Charles T. Campbell, and J. Michael Gottfried. Enhanced bonding of pentagon–heptagon defects in graphene to

metal surfaces: Insights from the adsorption of azulene and naphthalene to pt(111). Chemistry of Materials, 32(3):1041–1053, 2020.

[111] M. W. Haverkort, A. Tanaka, L. H. Tjeng, and G. A. Sawatzky. Nonresonant inelastic x-ray scattering involving excitonic excitations: The examples of nio and coo. Phys. Rev. Lett., 99:257401, Dec 2007.

[112] Yejun Feng, G. T. Seidler, J. O. Cross, A. T. Macrander, and J. J. Rehr. Role of inversion symmetry and multipole effects in nonresonant x-ray raman scattering from icosahedral b_4C. Phys. Rev. B, 69:125402, Mar 2004.

[113] J. A. Soininen, A. L. Ankudinov, and J. J. Rehr. Inelastic scattering from core electrons: A multiple scattering approach. Phys. Rev. B, 72:045136, Jul 2005.

[114] A. Sakko, M. Hakala, J. A. Soininen, and K. Hämäläinen. Density functional study of x-ray raman scattering from aromatic hydrocarbons and polyfluorene. Phys. Rev. B, 76:205115, Nov 2007.

[115] Jussi Lehtola, Mikko Hakala, Arto Sakko, and Keijo Hämäläinen. Erkale—a flexible program package for x-ray properties of atoms and molecules. Journal of Computational Chemistry, 33(18):1572–1585, 2012.

[116] Yves Joly, Chiara Cavallari, Sergey A. Guda, and Christoph J. Sahle. Full-potential simulation of x-ray raman scattering spectroscopy. Journal of Chemical Theory and Computation, 13(5):2172–2177, 2017.

[117] Eric L. Shirley. Ab initio inclusion of electron-hole attraction: Application to x-ray absorption and resonant inelastic x-ray scattering. Phys. Rev. Lett., 80:794–797, Jan 1998.

[118] J A Soininen, K Hämäläinen, W A Caliebe, C-C Kao, and Eric L Shirley. Core-hole-electron interaction in x-ray raman scattering. Journal of Physics: Condensed Matter, 13(35):8039, aug 2001.

[119] K. Gilmore, John Vinson, E.L. Shirley, D. Prendergast, C.D. Pemmaraju, J.J. Kas, F.D. Vila, and J.J. Rehr. Efficient implementation of core-excitation bethe–salpeter equation calculations. Computer Physics Communications, 197:109–117, 2015.

[120] G. van der Laan. Nonresonant inelastic x-ray scattering from actinides and rare earths. Phys. Rev. B, 86:035138, Jul 2012.

[121] Simo Huotari, Edlira Suljoti, Christoph J Sahle, Stephanie Rädel, Giulio Monaco, and Frank M F de Groot. High-resolution nonresonant x-ray raman scattering study on rare earth phosphate nanoparticles. New Journal of Physics, 17(4):043041, apr 2015.

[122] T. Willers, F. Strigari, N. Hiraoka, Y. Q. Cai, M. W. Haverkort, K.-D. Tsuei, Y. F. Liao, S. Seiro, C. Geibel, F. Steglich, L. H. Tjeng, and A. Severing. Determining the in-plane orientation of the ground-state orbital of $cecu_2si_2$. Phys. Rev. Lett., 109:046401, Jul 2012.

[123] J.-P. Rueff, J. M. Ablett, F. Strigari, M. Deppe, M. W. Haverkort, L. H. Tjeng, and A. Severing. Absence of orbital rotation in superconducting $cecu_2ge_2$. Phys. Rev. B, 91:201108, May 2015.

[124] Jingquan Liu, Jianguo Tang, and J. Justin Gooding. Strategies for chemical modification of graphene and applications of chemically modified graphene. J. Mater. Chem., 22:12435–12452, 2012.

[125] Keith E. Whitener. Review article: Hydrogenated graphene: A user's guide. J. Vac. Sci. Technol. A, 36(5):05G401, 2018.

[126] Yuhuan Fei, Siyuan Fang, and Yun Hang Hu. Synthesis, properties and potential applications of hydrogenated graphene. Chem. Eng. J., 397:125408, 2020.

[127] Valentina Tozzini and Vittorio Pellegrini. Prospects for hydrogen storage in graphene. Phys. Chem. Chem. Phys., 15:80–89, 2013.

[128] Igor A. Baburin, Alexey Klechikov, Guillaume Mercier, Alexandr Talyzin, and Gotthard Seifert. Hydrogen adsorption by perforated graphene. International Journal of Hydrogen Energy, 40(20):6594–6599, 2015.

[129] Rupali Nagar, Bhaghavathi P. Vinayan, Sai Smruti Samantaray, and Sundara Ramaprabhu. Recent advances in hydrogen storage using catalytically and chemically modified graphene nanocomposites. J. Mater. Chem. A, 5:22897–22912, 2017.

[130] James R. Morse, David A. Zugell, Eric Patterson, Jeffrey W. Baldwin, and Heather D. Willauer. Hydrogenated graphene: Important material properties regarding its application for hydrogen storage. J. Pow. Sour., 494:229734, 2021.

[131] Xiuxiu Han, Xili Tong, Xingchen Liu, Ai Chen, Xiaodong Wen, Nianjun Yang, and Xiang-Yun Guo. Hydrogen evolution reaction on hybrid catalysts of vertical mos2 nanosheets and hydrogenated graphene. ACS Catal., 8(3):1828–1836, 2018.

[132] A. H. Reshak. Chairlike and boatlike graphane: Active photocatalytic water splitting solar-to-hydrogen energy conversion under uv irradiation. J. Phys. Chem. C, 122(15):8076–8081, 2018.

[133] Daniel C Elias, Rahul Raveendran Nair, TMG Mohiuddin, SV Morozov, P Blake, MP Halsall, Andrea Carlo Ferrari, DW Boukhvalov, MI Katsnelson, AK Geim, et al. Control of graphene's properties by reversible hydrogenation: evidence for graphane. Science, 323(5914):610–613, 2009.

[134] Prarthana Gowda, Dipti R. Mohapatra, and Abha Misra. Enhanced photoresponse in monolayer hydrogenated graphene photodetector. ACS Appl. Mater. Interfaces, 6(19):16763–16768, 2014.

[135] Jangyup Son, Soogil Lee, Sang Jin Kim, Byung Cheol Park, Han-Koo Lee, Sanghoon Kim, Jae Hoon Kim, Byung Hee Hong, and Jongill Hong. Hydrogenated monolayer graphene with reversible and tunable wide band gap and its field-effect transistor. Nature communications, 7(1):13261, 2016.

[136] Shaorui Li, Jiaheng Li, Yongchao Wang, Chenglin Yu, Yaoxin Li, Wenhui Duan, Yayu Wang, and Jinsong Zhang. Large transport gap modulation in graphene via electric-field-controlled reversible hydrogenation. Nat. Electron., 4:254–260, 2021.

[137] Shaorui Li, Chenglin Yu, Yongchao Wang, Ke Zhang, Kaili Jiang, Yayu Wang, and Jinsong Zhang. Tafel-kinetics-controlled high-speed switching in a electrochemical graphene field-effect transistor. ACS Appl. Mater. Interf., 14(42):47991–47998, 2022.

[138] Shu Min Tan, Zdeněk Sofer, and Martin Pumera. Biomarkers detection on hydrogenated graphene surfaces: Towards applications of graphane in biosensing. Electroanalysis, 25(3):703–705, 2013.

[139] Arghya Narayan Banerjee. Graphene and its derivatives as biomedical materials: future prospects and challenges. Interface Focus, 8(3):20170056, 2018.

[140] Mounia Chakik, Siziwe Bebe, and Ravi Prakash. Hydrogenated graphene based organic thin film transistor sensor for detection of chloride ions as corrosion precursors. Appl. Sci., 12(2), 2022.

[141] Pierluigi Cudazzo, Claudio Attaccalite, Ilya V Tokatly, and Angel Rubio. Strong charge-transfer excitonic effects and the bose-einstein exciton condensate¡? format?¿ in graphane. Physical review letters, 104(22):226804, 2010.

[142] Zeyu Jiang, Wenkai Lou, Yu Liu, Yuanchang Li, Haifeng Song, Kai Chang, Wenhui Duan, and Shengbai Zhang. Spin-triplet excitonic insulator: The case of semihydrogenated graphene. Phys. Rev. Lett., 124:166401, Apr 2020.

[143] Hiroki Katow, Ryosuke Akashi, Yoshiyuki Miyamoto, and Shinji Tsuneyuki. First-principles study of the optical dipole trap for two-dimensional excitons in graphane. Phys. Rev. Lett., 129:047401, Jul 2022.

[144] S Lebegue, Mattias Klintenberg, Olle Eriksson, and MI Katsnelson. Accurate electronic band gap of pure and functionalized graphane from gw calculations. Physical Review B—Condensed Matter and Materials Physics, 79(24):245117, 2009.

[145] Pierluigi Cudazzo, Lorenzo Sponza, Christine Giorgetti, Lucia Reining, Francesco Sottile, and Matteo Gatti. Exciton band structure in two-dimensional materials. Physical review letters, 116(6):066803, 2016.

[146] Zhiqiang Luo, Jingzhi Shang, Sanhua Lim, Dehui Li, Qihua Xiong, Zexiang Shen, Jianyi Lin, and Ting Yu. Modulating the electronic structures of graphene by controllable hydrogenation. Appl. Phys. Lett., 97(23):233111, 2010.

[147] Bernard R. Matis, James S. Burgess, Felipe A. Bulat, Adam L. Friedman, Brian H. Houston, and Jeffrey W. Baldwin. Surface doping and band gap tunability in hydrogenated graphene. ACS Nano, 6(1):17–22, 2012.

[148] Yang Yang, Yongjun Li, Zhong Huang, and Xiaoyu Huang. (c1.04h)n: A nearly perfect pure graphane. Carbon, 107:154–161, 2016.

[149] Maria Grazia Betti, Ernesto Placidi, Chiara Izzo, Elena Blundo, Antonio Polimeni, Marco Sbroscia, José Avila, Pavel Dudin, Kailong Hu, Yoshikazu Ito, et al. Gap opening in double-sided highly hydrogenated free-standing graphene. Nano Letters, 22(7):2971–2977, 2022.

[150] Marcel H. F. Sluiter and Yoshiyuki Kawazoe. Cluster expansion method for adsorption: Application to hydrogen chemisorption on graphene. Phys. Rev. B, 68:085410, Aug 2003.

[151] Jorge O Sofo, Ajay S Chaudhari, and Greg D Barber. Graphane: A two-dimensional hydrocarbon. Physical Review B—Condensed Matter and Materials Physics, 75(15):153401, 2007.

[152] Ricarda A. Schäfer, Jan M. Englert, Peter Wehrfritz, Walter Bauer, Frank Hauke, Thomas Seyller, and Andreas Hirsch. On the way to graphane—pronounced fluorescence of polyhydrogenated graphene. Angew. Chem. Int. Ed., 52(2):754–757, 2013.

[153] Volker Strauss, Ricarda A. Schäfer, Frank Hauke, Andreas Hirsch, and Dirk M. Guldi. Polyhydrogenated graphene: Excited state dynamics in photo- and electroactive two-dimensional domains. J. Am. Chem. Soc., 137(40):13079–13086, 2015.

[154] Ricarda A. Schäfer, Daniela Dasler, Udo Mundloch, Frank Hauke, and Andreas Hirsch. Basic insights into tunable graphene hydrogenation. J. Am. Chem. Soc., 138(5):1647–1652, 2016.

[155] Zhiqiang Luo, Ting Yu, Ki-jeong Kim, Zhenhua Ni, Yumeng You, Sanhua Lim, Zexiang Shen, Shanzhong Wang, and Jianyi Lin. Thickness-dependent reversible hydrogenation of graphene layers. Acs Nano, 3(7):1781–1788, 2009.

[156] Sunmin Ryu, Melinda Y Han, Janina Maultzsch, Tony F Heinz, Philip Kim, Michael L Steigerwald, and Louis E Brus. Reversible basal plane hydrogenation of graphene. Nano letters, 8(12):4597–4602, 2008.

[157] Keith E Whitener Jr, Woo K Lee, Paul M Campbell, Jeremy T Robinson, and Paul E Sheehan. Chemical hydrogenation of single-layer graphene enables completely reversible removal of electrical conductivity. Carbon, 72:348–353, 2014.

[158] Richard Balog, Mie Andersen, Bjarke Jørgensen, Zeljko Sljivancanin, Bjørk Hammer, Alessandro Baraldi, Rosanna Larciprete, Philip Hofmann, Liv Hornekær, and Silvano Lizzit. Controlling hydrogenation of graphene on ir (111). ACS nano, 7(5):3823–3832, 2013.

[159] Srivats Rajasekaran, Frank Abild-Pedersen, Hirohito Ogasawara, Anders Nilsson, and Sarp Kaya. Interlayer carbon bond formation induced by hydrogen adsorption¡? format?¿ in few-layer supported graphene. Physical review letters, 111(8):085503, 2013.

[160] Daniel Lizzit, Mario I. Trioni, Luca Bignardi, Paolo Lacovig, Silvano Lizzit, Rocco Martinazzo, and Rosanna Larciprete. Dual-route hydrogenation of the graphene/ni interface. ACS Nano, 13(2):1828–1838, 2019.

[161] Iolanda Di Bernardo, Giulia Avvisati, Carlo Mariani, Nunzio Motta, Chaoyu Chen, Jose Avila, Maria Carmen Asensio, Stefano Lupi, Yoshikazu Ito, Mingwei Chen, et al. Two-dimensional hallmark of highly interconnected three-dimensional nanoporous graphene. ACS omega, 2(7):3691–3697, 2017.

[162] Iolanda Di Bernardo, Giulia Avvisati, Chaoyu Chen, José Avila, Maria Carmen Asensio, Kailong Hu, Yoshikazu Ito, Peter Hines, Josh Lipton-Duffin, Llew Rintoul, et al. Topology and doping effects in three-dimensional nanoporous graphene. Carbon, 131:258–265, 2018.

[163] Maria Grazia Betti, Dario Marchiani, Andrea Tonelli, Marco Sbroscia, Elena Blundo, Marta De Luca, Antonio Polimeni, Riccardo Frisenda, Carlo Mariani, Samuel Jeong, Yoshikazu Ito, Nicola Cavani, Roberto Biagi, Peter N.O. Gillespie, Michael A. Hernandez Bertran, Miki Bonacci, Elisa Molinari,

Valentina De Renzi, and Deborah Prezzi. Dielectric response and excitations of hydrogenated free-standing graphene. Carbon Trends, 12:100274, 2023.

[164] J Zhou, Q Wang, Q Sun, XS Chen, Y Kawazoe, and P Jena. Ferromagnetism in semihydrogenated graphene sheet. Nano letters, 9(11):3867–3870, 2009.

[165] Lennart Bengtsson. Dipole correction for surface supercell calculations. Phys. Rev. B, 59:12301–12304, May 1999.

[166] Mahmoud Mohamed Saad Abdelnabi, Chiara Izzo, Elena Blundo, Maria Grazia Betti, Marco Sbroscia, Giulia Di Bella, Gianluca Cavoto, Antonio Polimeni, Isabel García-Cortés, Isabel Rucandio, et al. Deuterium adsorption on free-standing graphene. Nanomaterials, 11(1):130, 2021.

[167] Mahmoud Mohamed Saad Abdelnabi, Elena Blundo, Maria Grazia Betti, Gianluca Cavoto, Ernesto Placidi, Antonio Polimeni, Alessandro Ruocco, Kailong Hu, Yoshikazu Ito, and Carlo Mariani. Towards free-standing graphane: atomic hydrogen and deuterium bonding to nano-porous graphene. Nanotechnology, 32(3):035707, 2020.

[168] Giulio D'Acunto, Francesca Ripanti, Paolo Postorino, Maria Grazia Betti, Mattia Scardamaglia, Carla Bittencourt, and Carlo Mariani. Channelling and induced defects at ion-bombarded aligned multiwall carbon nanotubes. Carbon, 139:768–775, 2018.

[169] William Yourey. Silicon negative electrodes—what can be achieved for commercial cell energy densities. Batteries, 9(12), 2023.

[170] Paul Gasper, Aron Saxon, Ying Shi, Elizabeth Endler, Kandler Smith, and Foram M. Thakkar. Degradation and modeling of large-format commercial lithium-ion cells as a function of chemistry, design, and aging conditions. Journal of Energy Storage, 73:109042, 2023.

[171] Hui Wu, Gerentt Chan, Jang Wook Choi, Ill Ryu, Yan Yao, Matthew T. McDowell, Seok Woo Lee, Ariel Jackson, Yuan Yang, Liangbing Hu, and Yi Cui. Stable cycling of double-walled silicon nanotube battery anodes through solid–electrolyte interphase control. Nature Nanotechnology, 7(5):310–315, May 2012.

[172] Gebrekidan Gebresilassie Eshetu, Heng Zhang, Xabier Judez, Henry Adenusi, Michel Armand, Stefano Passerini, and Egbert Figgemeier. Production of high-energy li-ion batteries comprising silicon-containing anodes and insertion-type cathodes. Nature Communications, 12(1):5459, Sep 2021.

[173] Zhenzhen Yang, Stephen E. Trask, Xianyang Wu, and Brian J. Ingram. Effect of si content on extreme fast charging behavior in silicon–graphite composite anodes. Batteries, 9(2), 2023.

[174] Sujong Chae, Seong-Hyeon Choi, Namhyung Kim, Jaekyung Sung, and Jaephil Cho. Integration of graphite and silicon anodes for the commercialization of high-energy lithium-ion batteries. Angewandte Chemie International Edition, 59(1):110–135, 2020.

[175] Zhenda Lu, Nian Liu, Hyun-Wook Lee, Jie Zhao, Weiyang Li, Yuzhang Li, and Yi Cui. Nonfilling carbon coating of porous silicon micrometer-sized particles for high-performance lithium battery anodes. ACS Nano, 9(3):2540–2547, 2015. PMID: 25738223.

[176] Alison L. Michan, Giorgio Divitini, Andrew J. Pell, Michal Leskes, Caterina Ducati, and Clare P. Grey. Solid electrolyte interphase growth and capacity loss in silicon electrodes. J. Am. Chem. Soc., 138(25):7918–7931, 2016.

[177] Yang He, Lin Jiang, Tianwu Chen, Yaobin Xu, Haiping Jia, Ran Yi, Dingchuan Xue, Miao Song, Arda Genc, Cedric Bouchet-Marquis, Lee Pullan, Ted Tessner, Jinkyoung Yoo, Xiaolin Li, Ji-Guang Zhang, Sulin Zhang, and Chongmin Wang. Progressive growth of the solid–electrolyte interphase towards the si anode interior causes capacity fading. Nature Nanotechnology, 16(10):1113–1120, Oct 2021.

[178] Yanli Chu, Yanbin Shen, Feng Guo, Xuan Zhao, Qingyu Dong, Qingyong Zhang, Wei Li, Hui Chen, Zhaojun Luo, and Liwei Chen. Advanced characterizations of solid electrolyte interphases in lithium-ion batteries. Electrochemical Energy Reviews, 3(1):187–219, Mar 2020.

[179] Jijian Xu. CEI and SEI Formation in Li-Ion Batteries, pages 307–324. Springer Nature Switzerland, Cham, 2024.

[180] E. Peled and S. Menkin. Review—sei: Past, present and future. Journal of The Electrochemical Society, 164(7):A1703, jun 2017.

[181] Ujwal Shreenag Meda, Libin Lal, Sushantha M, and Paridhi Garg. Solid electrolyte interphase (sei), a boon or a bane for lithium batteries: A review on the recent advances. Journal of Energy Storage, 47:103564, 2022.

[182] Ken Tasaki, Alex Goldberg, Jian-Jie Lian, Merry Walker, Adam Timmons, and Stephen J. Harris. Solubility of lithium salts formed on the lithium-ion battery negative electrode surface in organic solvents. J. Electrochem. Soc., 156(12):A1019, oct 2009.

[183] Kjell Schroder, Judith Alvarado, Thomas A. Yersak, Juchuan Li, Nancy Dudney, Lauren J. Webb, Ying Shirley Meng, and Keith J. Stevenson. The effect of fluoroethylene carbonate as an additive on the solid electrolyte interphase on silicon lithium-ion electrodes. Chem. Mater., 27(16):5531–5542, 2015.

[184] Oluwadamilola O. Taiwo, Melanie Loveridge, Shane D. Beattie, Donal P. Finegan, Rohit Bhagat, Daniel J.L. Brett, and Paul R. Shearing. Investigation of cycling-induced microstructural degradation in silicon-based electrodes in lithium-ion batteries using x-ray nanotomography. Electrochimica Acta, 253:85–92, 2017.

[185] Simon Müller, Manuel Lippuner, Mariana Verezhak, Vincent De Andrade, Francesco De Carlo, and Vanessa Wood. Multimodal nanoscale tomographic imaging for battery electrodes. Advanced Energy Materials, 10(28):1904119, 2020.

[186] Simon Müller, Patrick Pietsch, Ben-Elias Brandt, Paul Baade, Vincent De Andrade, Francesco De Carlo, and Vanessa Wood. Quantification and modeling of mechanical degradation in lithium-ion batteries based on nanoscale imaging. Nature communications, 9(1):2340, 2018.

[187] Praveen Kumar, Christopher L. Berhaut, Diana Zapata Dominguez, Eric De Vito, Samuel Tardif, Stéphanie Pouget, Sandrine Lyonnard, and Pierre-Henri Jouneau. Nano-architectured composite anode enabling long-term cycling stability for high-capacity lithium-ion batteries. Small, 16(11):1906812, 2020.

[188] Nian Liu, Hui Wu, Matthew T. McDowell, Yan Yao, Chongmin Wang, and Yi Cui. A yolk-shell design for stabilized and scalable li-ion battery alloy anodes. Nano Letters, 12(6):3315–3321, 2012. PMID: 22551164.

[189] Hansu Kim, Eung-Ju Lee, and Yang-Kook Sun. Recent advances in the si-based nanocomposite materials as high capacity anode materials for lithium ion batteries. Materials Today, 17(6):285–297, 2014.

[190] Sunghun Choi, Tae woo Kwon, Ali Coskun, and Jang Wook Choi. Highly elastic binders integrating polyrotaxanes for silicon microparticle anodes in lithium ion batteries. Science, 357(6348):279–283, 2017.

[191] Nian Liu, Zhenda Lu, Jie Zhao, Matthew T. McDowell, Hyun-Wook Lee, Wenting Zhao, and Yi Cui. A pomegranate-inspired nanoscale design for large-volume-change lithium battery anodes. Nature Nanotechnology, 9(3):187–192, Mar 2014.

[192] Weili An, Biao Gao, Shixiong Mei, Ben Xiang, Jijiang Fu, Lei Wang, Qiaobao Zhang, Paul K. Chu, and Kaifu Huo. Scalable synthesis of ant-nest-like bulk porous silicon for high-performance lithium-ion battery anodes. Nature Communications, 10(1):1447, Mar 2019.

[193] N. S. Hochgatterer, M. R. Schweiger, S. Koller, P. R. Raimann, T. Wöhrle, C. Wurm, and M. Winter. Silicon/graphite composite electrodes for high-capacity anodes: Influence of binder chemistry on cycling stability. Electrochem. Solid-State Lett., 11(5):A76, mar 2008.

[194] U. S. Vogl, P. K. Das, A. Z. Weber, M. Winter, R. Kostecki, and S. F. Lux. Mechanism of interactions between cmc binder and si single crystal facets. Langmuir, 30(34):10299–10307, 2014.

[195] Mengyun Nie, Daniel P. Abraham, Yanjing Chen, Arijit Bose, and Brett L. Lucht. Silicon solid electrolyte interphase (sei) of lithium ion battery characterized by microscopy and spectroscopy. J. Phys. Chem. C, 117(26):13403–13412, 2013.

[196] Julibeth M. Martínez de la Hoz and Perla B. Balbuena. Reduction mechanisms of additives on si anodes of li-ion batteries. Phys. Chem. Chem. Phys., 16:17091–17098, 2014.

[197] Lin Ma, Jian Xia, Xin Xia, and J. R. Dahn. The impact of vinylene carbonate, fluoroethylene carbonate and vinyl ethylene carbonate electrolyte additives on electrode/electrolyte reactivity studied using accelerating rate calorimetry. J. Electrochem. Soc., 161(10):A1495, jul 2014.

[198] Yanting Jin, Nis-Julian H. Kneusels, Lauren E. Marbella, Elizabeth Castillo-Martínez, Pieter C. M. M. Magusin, Robert S. Weatherup, Erlendur Jónsson, Tao Liu, Subhradip Paul, and Clare P. Grey. Understanding fluoroethylene carbonate and vinylene carbonate based electrolytes for si anodes in lithium ion batteries with nmr spectroscopy. J. Am. Chem. Soc., 140(31):9854–9867, 2018.

[199] Qi Li, Xiangsi Liu, Xiang Han, Yuxuan Xiang, Guiming Zhong, Jian Wang, Bizhu Zheng, Jigang Zhou, and Yong Yang. Identification of the solid electrolyte interface on the si/c composite anode with fec as the additive. ACS Appl. Mater. & Interfaces, 11(15):14066–14075, 2019.

[200] Fei Wang, Bo Wang, Jingxuan Li, Bin Wang, Yu Zhou, Dianlong Wang, Huakun Liu, and Shixue Dou. Prelithiation: a crucial strategy for boosting the practical application of next-generation lithium ion battery. ACS nano, 15(2):2197–2218, 2021.

[201] Diana Zapata Dominguez, Christopher L. Berhaut, Praveen Kumar, Pierre-Henri Jouneau, Antoine Desrues, Nathalie Herlin-Boime, Nathalie Boudet, Nils Blanc, Gilbert A. Chahine, Cédric Haon, Samuel Tardif, Sandrine Lyonnard, and Stéphanie Pouget. Influence of the ge content on the lithiation process of crystalline si1xgex nanoparticle-based anodes for li-ion batteries. J. Mater. Chem. A, 11:19025–19035, 2023.

[202] Dominik Petz, Martin J Mühlbauer, Volodymyr Baran, Alexander Schökel, Vladislav Kochetov, Michael Hofmann, Vadim Dyadkin, Peter Staron, Gavin Vaughan, Ulrich Lienert, et al. Lithium distribution and transfer in high-power 18650-type li-ion cells at multiple length scales. Energy Storage Materials, 41:546–553, 2021.

[203] Bin Hu, Sisi Jiang, Ilya A. Shkrob, Jingjing Zhang, Stephen E. Trask, Bryant J. Polzin, Andrew Jansen, Wei Chen, Chen Liao, Zhengcheng Zhang, and Lu Zhang. Understanding of pre-lithiation of poly(acrylic acid) binder: Striking the balances between the cycling performance and slurry stability for silicon-graphite composite electrodes in li-ion batteries. J. Power Sources, 416:125–131, 2019.

[204] Alison L. Michan, Bharathy. S. Parimalam, Michal Leskes, Rachel N. Kerber, Taeho Yoon, Clare P. Grey, and Brett L. Lucht. Fluoroethylene carbonate and vinylene carbonate reduction: Understanding lithium-ion battery electrolyte additives and solid electrolyte interphase formation. Chem. Mater., 28(22):8149–8159, 2016.

[205] Chao Xu, Fredrik Lindgren, Bertrand Philippe, Mihaela Gorgoi, Fredrik Björefors, Kristina Edström, and Torbjörn Gustafsson. Improved performance of the silicon anode for li-ion batteries: Understanding the surface modification mechanism of fluoroethylene carbonate as an effective electrolyte additive. Chem. Mater., 27(7):2591–2599, 2015.

[206] Chao Xu, Bing Sun, Torbjörn Gustafsson, Kristina Edström, Daniel Brandell, and Maria Hahlin. Interface layer formation in solid polymer electrolyte lithium batteries: an xps study. J. Mater. Chem. A, 2:7256–7264, 2014.

[207] Bertrand Philippe, Rémi Dedryvère, Joachim Allouche, Fredrik Lindgren, Mihaela Gorgoi, Håkan Rensmo, Danielle Gonbeau, and Kristina Edström. Nanosilicon electrodes for lithium-ion batteries: Interfacial mechanisms studied by hard and soft x-ray photoelectron spectroscopy. Chem. Mater., 24(6):1107–1115, 2012.

[208] Bertrand Philippe, Rémi Dedryvère, Mihaela Gorgoi, Håkan Rensmo, Danielle Gonbeau, and Kristina Edström. Role of the $LiPF_6$ salt for the

long-term stability of silicon electrodes in li-ion batteries – a photoelectron spectroscopy study. Chem. Mater., 25(3):394–404, 2013.

[209] Guiomar Hernández, Andrew J. Naylor, Yu-Chuan Chien, Daniel Brandell, Jonas Mindemark, and Kristina Edström. Elimination of fluorination: The influence of fluorine-free electrolytes on the performance of lini1/3mn1/3co1/3o2/silicon–graphite li-ion battery cells. ACS Sustainable Chem. Eng., 8(27):10041–10052, 2020.

[210] Jagjit Nanda, Guang Yang, Tingzheng Hou, Dmitry N. Voylov, Xin Li, Rose E. Ruther, Michael Naguib, Kristin Persson, Gabriel M. Veith, and Alexei P. Sokolov. Unraveling the nanoscale heterogeneity of solid electrolyte interphase using tip-enhanced raman spectroscopy. Joule, 3(8):2001–2019, 2019.

[211] Rose E. Ruther, Kevin A. Hays, Seong Jin An, Jianlin Li, David L. Wood, and Jagjit Nanda. Chemical evolution in silicon–graphite composite anodes investigated by vibrational spectroscopy. ACS Applied Materials & Interfaces, 10(22):18641–18649, 2018.

[212] Maxime Boniface, Lucille Quazuguel, Julien Danet, Dominique Guyomard, Philippe Moreau, and Pascale Bayle-Guillemaud. Nanoscale chemical evolution of silicon negative electrodes characterized by low-loss stem-eels. Nano Lett., 16(12):7381–7388, 2016.

[213] Magali Gauthier, Thomas J. Carney, Alexis Grimaud, Livia Giordano, Nir Pour, Hao-Hsun Chang, David P. Fenning, Simon F. Lux, Odysseas Paschos, Christoph Bauer, Filippo Maglia, Saskia Lupart, Peter Lamp, and Yang Shao-Horn. Electrode–electrolyte interface in li-ion batteries: Current understanding and new insights. J. Phys. Chem. Lett, 6(22):4653–4672, 2015.

[214] Antoine Desrues, Eric De Vito, Florent Boismain, John P Alper, Cédric Haon, Nathalie Herlin-Boime, and Sylvain Franger. Electrochemical and x-ray photoelectron spectroscopic study of early sei formation and evolution on si and si@ c nanoparticle-based electrodes. Materials, 15(22):7990, 2022.

[215] Chuntian Cao, Badri Shyam, Jiajun Wang, Michael F Toney, and Hans-Georg Steinruck. Shedding x-ray light on the interfacial electrochemistry of silicon anodes for li-ion batteries. Accounts of Chemical Research, 52(9):2673–2683, 2019.

[216] Jack E. N. Swallow, Michael W. Fraser, Nis-Julian H. Kneusels, Jodie F. Charlton, Christopher G. Sole, Conor M. E. Phelan, Erik Björklund, Peter Bencok, Carlos Escudero, Virginia Pérez-Dieste, Clare P. Grey, Rebecca J. Nicholls, and Robert S. Weatherup. Revealing solid electrolyte

interphase formation through interface-sensitive operando x-ray absorption spectroscopy. Nat. Commun., 13(1):6070, Oct 2022.

[217] M. Schellenberger, R. Golnak, W.G. Quevedo Garzon, S. Risse, and R. Seidel. Accessing the solid electrolyte interphase on silicon anodes for lithium-ion batteries in-situ through transmission soft x-ray absorption spectroscopy. Materials Today Advances, 14:100215, 2022.

[218] Patrick Pietsch, Michael Hess, Wolfgang Ludwig, Jens Eller, and Vanessa Wood. Combining operando synchrotron x-ray tomographic microscopy and scanning x-ray diffraction to study lithium ion batteries. Scientific reports, 6(1):27994, 2016.

[219] Jeremy IG Dawkins, Isaac Martens, Andrew Danis, Isabelle Beaulieu, Danny Chhin, Marta Mirolo, Jakub Drnec, Steen B Schougaard, and Janine Mauzeroll. Mapping the total lithium inventory of li-ion batteries. Joule, 7(12):2783–2797, 2023.

[220] Didem Ketenoglu, Georg Spiekermann, Manuel Harder, Erdinc Oz, Cevriye Koz, Mehmet C. Yagci, Eda Yilmaz, Zhong Yin, Christoph J. Sahle, Blanka Detlefs, and Hasan Yavaş. X-ray raman spectroscopy of lithium-ion battery electrolyte solutions in a flow cell. J. Synchrotron Rad., 25(2):537–542, 2018.

[221] G. E. Stutz, M. Otero, S. A. Ceppi, C. B. Robledo, G. Luque, E. Leiva, and D. E. Barraco Díaz. Intercalation stage dependence of core electronic excitations in Li-intercalated graphite from inelastic X-ray scattering. Appl. Phys. Lett., 110(25):253901, 06 2017.

[222] Takamasa Nonaka, Hiroyuki Kawaura, Yoshinari Makimura, Yusaku F. Nishimura, and Kazuhiko Dohmae. In situ x-ray raman scattering spectroscopy of a graphite electrode for lithium-ion batteries. J. Power Sources, 419:203–207, 2019.

[223] Ulrike Boesenberg, Dimosthenis Sokaras, Dennis Nordlund, Tsu-Chien Weng, Evgeny Gorelov, Thomas J. Richardson, Robert Kostecki, and Jordi Cabana. Electronic structure changes upon lithium intercalation into graphite – insights from ex situ and operando x-ray raman spectroscopy. Carbon, 143:371–377, 2019.

[224] Tristan de Boer, Jacob G. Lapping, Jeffrey A. Read, Timothy T. Fister, Mahalingam Balasubramanian, Jordi Cabana, and Alexander Moewes. Direct evidence of charge transfer upon anion intercalation in graphite cathodes through new electronic states: An experimental and theoretical study of hexafluorophosphate. Chem. Mater., 32(5):2036–2043, 2020.

[225] Ava Rajh, Iztok Arčon, Klemen Bučar, Matjaž Žitnik, Marko Petric, Alen Vizintin, Jan Bitenc, Urban Košir, Robert Dominko, Hlynur Gretarsson, Martin Sundermann, and Matjaž Kavčič. Characterization of electrochemical processes in metal–organic batteries by x-ray raman spectroscopy. J. Phys. Chem. C, 126(12):5435–5442, 2022.

[226] Laura M. de Kort, Masoud Lazemi, Alessandro Longo, Valerio Gulino, Henrik P. Rodenburg, Didier Blanchard, Christoph Sahle, Martin Sundermann, Hlynur Gretarsson, Ad M. J. van der Eerden, Hebatalla Elnaggar, Frank M. F. de Groot, and Peter Ngene. Deciphering the origin of interface-induced high li and na ion conductivity in nanocomposite solid electrolytes using x-ray raman spectroscopy. Adv. Energy Mater., 14(9):2303381, 2024.

[227] Jazer Jose H. Togonon, Antonella Iadecola, Romain Wernert, Kriti Choudhary, Mauro Rovezzi, Jean-Noël Chotard, Lorenzo Stievano, Alessandro Longo, and Laurence Croguennec. Disentangling the ligand and electronic structure in kvpo4f1-xox positive electrode materials by valence-to-core x-ray emission spectroscopy. Energy Storage Mater., 69:103406, 2024.

[228] Artur Braun, Dennis Nordlund, Seung-Wan Song, Tzu-Wen Huang, Dimosthenis Sokaras, Xiasong Liu, Wanli Yang, Tsu-Chien Weng, and Zhi Liu. Hard x-rays in–soft x-rays out: An operando piggyback view deep into a charging lithium ion battery with x-ray raman spectroscopy. J. Electron Spectros. Relat. Phenomena, 200:257–263, 2015. Special Anniversary Issue: Volume 200.

[229] Xiaosong Liu, Jun Liu, Ruimin Qiao, Yan Yu, Hong Li, Liumin Suo, Yongsheng Hu, Yi-De Chuang, Guojiun Shu, Fangcheng Chou, Tsu-Chien Weng, Dennis Nordlund, Dimosthenis Sokaras, Yung Jui Wang, Hsin Lin, Bernardo Barbiellini, Arun Bansil, Xiangyun Song, Zhi Liu, Shishen Yan, Gao Liu, Shan Qiao, Thomas J. Richardson, David Prendergast, Zahid Hussain, Frank M. F. de Groot, and Wanli Yang. Phase transformation and lithiation effect on electronic structure of $LixFePO_4$: An in-depth study by soft x-ray and simulations. J. Am. Chem. Soc., 134(33):13708–13715, 2012.

[230] Antoine Desrues, John P. Alper, Florent Boismain, Diana Zapata Dominguez, Christopher Berhaut, Pierre-Eugène Coulon, Adrien Soloy, Frédéric Grisch, Samuel Tardif, Stéphanie Pouget, Sandrine Lyonnard, Cédric Haon, and Nathalie Herlin-Boime. Best performing sige/si core-shell nanoparticles synthesized in one step for high capacity anodes. Batter. Supercaps, 2(12):970–978, 2019.

[231] M. N. Obrovac and Leif Christensen. Structural changes in silicon anodes during lithium insertion/extraction. Electrochemical and Solid-State Letters, 7(5):A93, mar 2004.

[232] M. N. Obrovac and L. J. Krause. Reversible cycling of crystalline silicon powder. J. Electrochem. Soc., 154(2):A103, dec 2006.

[233] Sergio Pinilla, Sang-Hoon Park, Kenneth Fontanez, Francisco Márquez, Valeria Nicolosi, and Carmen Morant. 0d-1d hybrid silicon nanocomposite as lithium-ion batteries anodes. Nanomaterials, 10(3), 2020.

[234] J.-S. Bridel, T. Azaïs, M. Morcrette, J.-M. Tarascon, and D. Larcher. Key parameters governing the reversibility of si/carbon/cmc electrodes for li-ion batteries. Chem. Mater., 22(3):1229–1241, 2010.

[235] Etienne Radvanyi, Kristof Van Havenbergh, Willy Porcher, Séverine Jouanneau, Jean-Sébastien Bridel, Stijn Put, and Sylvain Franger. Study and modeling of the solid electrolyte interphase behavior on nano-silicon anodes by electrochemical impedance spectroscopy. Electrochim. Acta, 137:751–757, 2014.

[236] Candace K Chan, Riccardo Ruffo, Seung Sae Hong, and Yi Cui. Surface chemistry and morphology of the solid electrolyte interphase on silicon nanowire lithium-ion battery anodes. Journal of Power Sources, 189(2):1132–1140, 2009.

[237] Xing-Rui Liu, Xin Deng, Ran-Ran Liu, Hui-Juan Yan, Yu-Guo Guo, Dong Wang, and Li-Jun Wan. Single nanowire electrode electrochemistry of silicon anode by in situ atomic force microscopy: Solid electrolyte interphase growth and mechanical properties. ACS Applied Materials & Interfaces, 6(22):20317–20323, 2014. PMID: 25380518.

[238] K. Ogata, E. Salager, C. J. Kerr, A. E. Fraser, C. Ducati, A. J. Morris, S. Hofmann, and C. P. Grey. Revealing lithium–silicide phase transformations in nano-structured silicon-based lithium ion batteries via in situ nmr spectroscopy. Nat. Commun., 5(1):3217, Feb 2014.

[239] Wei-Ren Liu, Zheng-Zao Guo, Wen-Shiue Young, Deng-Tswen Shieh, Hung-Chun Wu, Mo-Hua Yang, and Nae-Lih Wu. Effect of electrode structure on performance of si anode in li-ion batteries: Si particle size and conductive additive. Journal of Power Sources, 140(1):139–144, 2005.

[240] Xiuyun Zhao and Vesa-Pekka Lehto. Challenges and prospects of nanosized silicon anodes in lithium-ion batteries. Nanotechnology, 32(4):042002, 2020.

[241] Bhagath Sreenarayanan, Darren H.S. Tan, Shuang Bai, Weikang Li, Wurigumula Bao, and Ying Shirley Meng. Quantification of lithium inventory loss in micro silicon anode via titration-gas chromatography. J. Power Sources, 531:231327, 2022.

[242] Doron Aurbach. Review of selected electrode–solution interactions which determine the performance of li and li ion batteries. J. Power Sources, 89(2):206–218, 2000.

[243] Samuel Tardif, Ekaterina Pavlenko, Lucille Quazuguel, Maxime Boniface, Manuel Maréchal, Jean-Sébastien Micha, Laurent Gonon, Vincent Mareau, Gérard Gebel, Pascale Bayle-Guillemaud, François Rieutord, and Sandrine Lyonnard. Operando raman spectroscopy and synchrotron x-ray diffraction of lithiation/delithiation in silicon nanoparticle anodes. ACS Nano, 11(11):11306–11316, 2017.

[244] Kjell W. Schroder, Hugo Celio, Lauren J. Webb, and Keith J. Stevenson. Examining solid electrolyte interphase formation on crystalline silicon electrodes: Influence of electrochemical preparation and ambient exposure conditions. J. Phys. Chem C, 116(37):19737–19747, 2012.

[245] Ch. J. Sahle, A. Mirone, J. Niskanen, J. Inkinen, M. Krisch, and S. Huotari. Planning, performing and analyzing X-ray Raman scattering experiments. J. Synchrotron Rad., 22(2):400–409, Mar 2015.

[246] S. Huotari, Ch. J. Sahle, Ch. Henriquet, A. Al-Zein, K. Martel, L. Simonelli, R. Verbeni, H. Gonzalez, M.-C. Lagier, C. Ponchut, M. Moretti Sala, M. Krisch, and G. Monaco. A large-solid-angle X-ray Raman scattering spectrometer at ID20 of the European Synchrotron Radiation Facility. J. Synchrotron Rad., 24(2):521–530, Mar 2017.

[247] Anubhav Jain, Joseph Montoya, Shyam Dwaraknath, Nils E. R. Zimmermann, John Dagdelen, Matthew Horton, Patrick Huck, Donny Winston, Shreyas Cholia, Shyue Ping Ong, and Kristin Persson. The Materials Project: Accelerating Materials Design Through Theory-Driven Data and Tools, pages 1751–1784. Springer International Publishing, Cham, 2020.

[248] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. PubChem 2023 update. Nucleic Acids Res., 51(D1):D1373–D1380, 10 2022.

[249] Gianluca Prandini, Antimo Marrazzo, Ivano E Castelli, Nicolas Mounet, and Nicola Marzari. Precision and efficiency in solid-state pseudopotential calculations. npj Comput. Mater., 4(1):72, 2018. http://materialscloud.org/sssp.

[250] Alice H. England, Andrew M. Duffin, Craig P. Schwartz, Janel S. Uejio, David Prendergast, and Richard J. Saykally. On the hydration and hydrolysis of carbon dioxide. Chem. Phys. Lett., 514(4):187–195, 2011.

[251] Haoyue Guo, Matthew R. Carbone, Chuntian Cao, Jianzhou Qu, Yonghua Du, Seong-Min Bak, Conan Weiland, Feng Wang, Shinjae Yoo, Nongnuch Artrith, Alexander Urban, and Deyu Lu. Simulated sulfur k-edge x-ray absorption spectroscopy database of lithium thiophosphate solid electrolytes. Sci. Data, 10(1):349, Jun 2023.

[252] Charles Daniel Wagner. Nist x-ray photoelectron spectroscopy database. NIST Standarad Reference Database 20, 2000.

[253] Giannantonio Cibin, Diego Gianolio, Stephen A. Parry, Tom Schoonjans, Oliver Moore, Rachael Draper, Laura A. Miller, Alexander Thoma, Claire L. Doswell, and Abigail Graham. An open access, integrated xas data repository at diamond light source. Radiation Physics and Chemistry, 175:108479, 2020. 17th International Conference on X-ray Absorption Fine Structure - XAFS2018.

[254] Kiran Mathew, Chen Zheng, Donald Winston, Chi Chen, Alan Dozier, John J. Rehr, Shyue Ping Ong, and Kristin A. Persson. Data descriptor: High-throughput computational x-ray absorption spectroscopy. Scientific Data, 5:1–8, 2018.

[255] Yiming Chen, Chi Chen, Chen Zheng, Shyam Dwaraknath, Matthew K. Horton, Jordi Cabana, John Rehr, John Vinson, Alan Dozier, Joshua J. Kas, Kristin A. Persson, and Shyue Ping Ong. Database of ab initio l-edge x-ray absorption near edge structure. Scientific Data, 8:153, 12 2021.

[256] Grzegorz Greczynski and Lars Hultman. A step-by-step guide to perform x-ray photoelectron spectroscopy. Journal of Applied Physics, 132(1):011101, 07 2022.

[257] Geoffroy Hautier, Anubhav Jain, and Shyue Ping Ong. From the computer to the laboratory: materials discovery and design using first-principles calculations. Journal of Materials Science, 47:7317–7340, 2012.

[258] Kirstin Alberi, Marco Buongiorno Nardelli, Andriy Zakutayev, Lubos Mitas, Stefano Curtarolo, Anubhav Jain, Marco Fornari, Nicola Marzari, Ichiro Takeuchi, Martin L Green, et al. The 2019 materials by design roadmap. Journal of Physics D: Applied Physics, 52(1):013001, 2018.

[259] Stefano Curtarolo, Wahyu Setyawan, Gus L.W. Hart, Michal Jahnatek, Roman V. Chepulskii, Richard H. Taylor, Shidong Wang, Junkai Xue, Kesong Yang, Ohad Levy, Michael J. Mehl, Harold T. Stokes, Denis O. Demchenko, and Dane Morgan. Aflow: An automatic framework for high-throughput materials discovery. Computational Materials Science, 58:218–226, 2012.

[260] 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.

[261] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, Eric D Hermes, Paul C Jennings, Peter Bjerre Jensen, James Kermode, John R Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson, Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob Schiøtz, Ole Schütt, Mikkel Strange, Kristian S Thygesen, Tejs Vegge, Lasse Vilhelmsen, Michael Walter, Zhenhua Zeng, and Karsten W Jacobsen. The atomic simulation environment—a python library for working with atoms. Journal of Physics: Condensed Matter, 29(27):273002, jun 2017.

[262] Kiran Mathew, Joseph H. Montoya, Alireza Faghaninia, Shyam Dwarakanath, Muratahan Aykol, Hanmei Tang, Iek heng Chu, Tess Smidt, Brandon Bocklund, Matthew Horton, John Dagdelen, Brandon Wood, Zi-Kui Liu, Jeffrey Neaton, Shyue Ping Ong, Kristin Persson, and Anubhav Jain. Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. Computational Materials Science, 139:140–152, 2017.

[263] Tam Mayeshiba, Henry Wu, Thomas Angsten, Amy Kaczmarowski, Zhewen Song, Glen Jenness, Wei Xie, and Dane Morgan. The materials simulation toolkit (mast) for atomistic modeling of defects and diffusion. Computational Materials Science, 126:90–102, 2017.

[264] Xiaoyu Yang, Zongguo Wang, Xushan Zhao, Jianlong Song, Mingming Zhang, and Haidong Liu. Matcloud: A high-throughput computational infrastructure for integrated management of materials simulation, data and resources. Computational Materials Science, 146:319–333, 2018.

[265] Kiran Mathew, Arunima K Singh, Joshua J Gabriel, Kamal Choudhary, Susan B Sinnott, Albert V Davydov, Francesca Tavazza, and Richard G Hennig. Mpinterfaces: A materials project based python tool for high-throughput computational screening of interfacial systems. Computational Materials Science, 122:183–190, 2016.

[266] Dassault Systèmes Biovia. Materials studio. R2 (Dassault Systèmes BIOVIA, San Diego, 2017.

[267] Inc. Materials Design. Medea. Angel Fire, NM, USA, 2016.

[268] Søren Smidstrup, Troels Markussen, Pieter Vancraeyveld, Jess Wellendorff, Julian Schneider, Tue Gunst, Brecht Verstichel, Daniele Stradi, Petr A Khomyakov, Ulrik G Vej-Hansen, Maeng-Eun Lee, Samuel T Chill, Filip Rasmussen, Gabriele Penazzi, Fabiano Corsetti, Ari Ojanperä, Kristian Jensen, Mattias L N Palsgaard, Umberto Martinez, Anders Blom, Mads Brandbyge, and Kurt Stokbro. Quantumatk: an integrated platform of electronic and atomic-scale modelling tools. Journal of Physics: Condensed Matter, 32(1):015901, oct 2019.

[269] Celso RC Rêgo, Jörg Schaarschmidt, Tobias Schlöder, Montserrat Penaloza-Amion, Saientan Bag, Tobias Neumann, Timo Strunk, and Wolfgang Wenzel. Simstack: an intuitive workflow framework. Frontiers in Materials, 9:877597, 2022.

[270] Scott Kirklin, James E Saal, Bryce Meredig, Alex Thompson, Jeff W Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. npj Computational Materials, 1(1):1–15, 2015.

[271] Claudia Draxl and Matthias Scheffler. The nomad laboratory: from data sharing to artificial intelligence. Journal of Physics: Materials, 2(3):036001, 2019.

[272] Alec Belsky, Mariette Hellenbrandt, Vicky Lynn Karen, and Peter Luksch. New developments in the inorganic crystal structure database (icsd): accessibility in support of materials research and design. Acta Crystallographica Section B: Structural Science, 58(3):364–369, 2002.

[273] Leopold Talirz, Snehal Kumbhar, Elsa Passaro, Aliaksandr V Yakutovich, Valeria Granata, Fernando Gargiulo, Marco Borelli, Martin Uhrin, Sebastiaan P Huber, Spyros Zoupanos, et al. Materials cloud, a platform for open computational science. Scientific data, 7(1):299, 2020.

[274] Kamal Choudhary, Kevin F. Garrity, Andrew C. E. Reid, Brian DeCost, Adam J. Biacchi, Angela R. Hight Walker, Zachary Trautt, Jason Hattrick-Simpers, A. Gilad Kusne, Andrea Centrone, Albert Davydov, Jie Jiang, Ruth Pachter, Gowoon Cheon, Evan Reed, Ankit Agrawal, Xiaofeng Qian, Vinit Sharma, Houlong Zhuang, Sergei V. Kalinin, Bobby G. Sumpter, Ghanshyam Pilania, Pinar Acar, Subhasish Mandal, Kristjan Haule, David

Vanderbilt, Karin Rabe, and Francesca Tavazza. The joint automated repository for various integrated simulations (jarvis) for data-driven materials design. npj Computational Materials, 6(1):173, Nov 2020.

[275] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 1(1):011002, 07 2013.

[276] National Science and Technology Council (US). Materials genome initiative for global competitiveness. Executive Office of the President, National Science and Technology Council, 2011.

[277] Jens Broeder, Daniel Wortmann, and Stefan Blügel. Using the aiida-fleur package for all-electron ab initio electronic structure data generation and processing in materials science. Schriften des Forschungszentrums Jülich IAS Series Band/Volume 40, 18:19, 2018.

[278] Daniel Wortmann, Gregor Michalicek, Nadjib Baadji, Markus Betzinger, Gustav Bihlmayer, Jens Bröder, Tobias Burnus, Jussi Enkovaara, Frank Freimuth, Christoph Friedrich, Christian-Roman Gerhorst, Sabastian Granberg Cauchi, Uliana Grytsiuk, Andrea Hanke, Jan-Philipp Hanke, Marcus Heide, Stefan Heinze, Robin Hilgers, Henning Janssen, Daniel Aaaron Klüppelberg, Roman Kovacik, Philipp Kurz, Marjana Lezaic, Georg K. H. Madsen, Yuriy Mokrousov, Alexander Neukirchen, Matthias Redies, Stefan Rost, Martin Schlipf, Arno Schindlmayr, Miriam Winkelmann, and Stefan Blügel. FLEUR. Zenodo, August 2024.

[279] 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.

[280] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. Scientific data, 3(1):1–9, 2016.

[281] Giovanni Pizzi, Andrea Cepellotti, Riccardo Sabatini, Nicola Marzari, and Boris Kozinsky. Aiida: automated interactive infrastructure and database for computational science. Computational Materials Science, 111:218–230, 2016.

[282] Luc Moreau, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Beth Plale, Yogesh Simmhan, Eric Stephan, and Jan Van den Bussche. The open provenance model core specification (v1.1). Future Gener. Comput. Syst., 27(6):743–756, jun 2011.

[283] Sydney R Hall, Frank H Allen, and I David Brown. The crystallographic information file (cif): a new standard archive file for crystallography. Foundations of Crystallography, 47(6):655–685, 1991.

[284] Anton Kokalj. Xcrysden—a new program for displaying crystalline structures and electron densities. Journal of Molecular Graphics and Modelling, 17(3-4):176–179, 1999.

[285] Toma Susi, Duncan J. Mowbray, Mathias P. Ljungberg, and Paola Ayala. Calculation of the graphene c 1s core level binding energy. Phys. Rev. B, 91:081401, 2015.

[286] Atsushi Togo and Isao Tanaka. Spglib: a software library for crystal symmetry search. arXiv preprint arXiv:1808.01590, 5, 2018.

[287] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. Computational Materials Science, 68:314–319, 2013.

[288] Aliaksandr V. Yakutovich, Kristjan Eimre, Ole Schütt, Leopold Talirz, Carl S. Adorf, Casper W. Andersen, Edward Ditler, Dou Du, Daniele Passerone, Berend Smit, Nicola Marzari, Giovanni Pizzi, and Carlo A. Pignedoli. Aiidalab – an ecosystem for developing, executing, and sharing scientific workflows. Computational Materials Science, 188:110165, 2021.

[289] Thomas Penfold, Luke Watson, Clelia Middleton, Tudur David, Sneha Verma, Thomas Pope, Julia Kaczmarek, and Conor Douglas Rankine. Machine-learning strategies for the accurate and efficient analysis of x-ray spectroscopy. Machine Learning: Science and Technology, 2024.

[290] Qintao Sun, Yan Xiang, Yue Liu, Liang Xu, Tianle Leng, Yifan Ye, Alessandro Fortunelli, William A Goddard III, and Tao Cheng. Machine learning predicts the x-ray photoelectron spectroscopy of the solid electrolyte interface of lithium metal battery. The Journal of Physical Chemistry Letters, 13(34):8047–8054, 2022.

[291] Dorothea Golze, Markus Hirvensalo, Patricia Hernández-León, Anja Aarva, Jarkko Etula, Toma Susi, Patrick Rinke, Tomi Laurila, and Miguel A

Caro. Accurate computational prediction of core-electron binding energies in carbon-based materials: A machine-learning model combining density-functional theory and gw. Chemistry of Materials, 34(14):6240–6254, 2022.

[292] Anja Aarva, Volker L Deringer, Sami Sainio, Tomi Laurila, and Miguel A Caro. Understanding x-ray spectroscopy of carbonaceous materials by combining experiments, density functional theory, and machine learning. part i: Fingerprint spectra. Chemistry of Materials, 31(22):9243–9255, 2019.

[293] Anja Aarva, Sami Sainio, Volker L Deringer, Miguel A Caro, and Tomi Laurila. X-ray spectroscopy fingerprints of pristine and functionalized graphene. The Journal of Physical Chemistry C, 125(33):18234–18246, 2021.

[294] Giovanni Drera, Chahan M Kropf, and Luigi Sangaletti. Deep neural network for x-ray photoelectron spectroscopy data analysis. Machine Learning: Science and Technology, 1(1):015008, 2020.

[295] Lukas Pielsticker, Rachel L Nicholls, Serena DeBeer, and Mark Greiner. Convolutional neural network framework for the automated analysis of transition metal x-ray photoelectron spectra. Analytica Chimica Acta, 1271:341433, 2023.

[296] Yusuf Shaidu, Emine Küçükbenli, Ruggero Lot, Franco Pellegrini, Efthimios Kaxiras, and Stefano de Gironcoli. A systematic approach to generating accurate neural network potentials: The case of carbon. npj Computational Materials, 7(1):52, 2021.

[297] Fangjia Fu, Xiaoxu Wang, Linfeng Zhang, Yifang Yang, Jianhui Chen, Bo Xu, Chuying Ouyang, Shenzhen Xu, Fu-Zhi Dai, and Weinan E. Unraveling the atomic-scale mechanism of phase transformations and structural evolutions during (de) lithiation in si anodes. Advanced Functional Materials, 33(37):2303936, 2023.

[298] Colin W Glass, Artem R Oganov, and Nikolaus Hansen. Uspex—evolutionary crystal structure prediction. Computer physics communications, 175(11-12):713–720, 2006.

[299] Artem R Oganov and Colin W Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. The Journal of chemical physics, 124(24), 2006.

[300] Artem R Oganov and Mario Valle. How to quantify energy landscapes of solids. The Journal of chemical physics, 130(10), 2009.

[301] Mario Valle and Artem R Oganov. Crystal fingerprint space–a novel paradigm for studying crystal-structure sets. Acta Crystallographica Section A: Foundations of Crystallography, 66(5):507–517, 2010.

[302] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. Journal of computational physics, 117(1):1–19, 1995.

[303] Guillaume Fraux, Rose K. Cersonsky, and Michele Ceriotti. Chemiscope: interactive structure-property explorer for materials and molecules. Journal of Open Source Software, 5(51):2117, 2020.

[304] Federico M Paruzzo, Albert Hofstetter, Félix Musil, Sandip De, Michele Ceriotti, and Lyndon Emsley. Chemical shifts in molecular solids by machine learning. Nature communications, 9(1):4501, 2018.

[305] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[306] Volker L Deringer and Gábor Csányi. Machine learning based interatomic potential for amorphous carbon. Physical Review B, 95(9):094203, 2017.

[307] Jian-Xing Huang, Gábor Csányi, Jin-Bao Zhao, Jun Cheng, and Volker L Deringer. First-principles study of alkali-metal intercalation in disordered carbon anode materials. Journal of Materials Chemistry A, 7(32):19070–19080, 2019.

[308] Michael S Chen, Tobias Morawietz, Hideki Mori, Thomas E Markland, and Nongnuch Artrith. Aenet–lammps and aenet–tinker: Interfaces for accurate and efficient molecular dynamics simulations with machine learning potentials. The Journal of Chemical Physics, 155(7), 2021.

[309] Etienne Radvanyi, Eric De Vito, Willy Porcher, and Séverine Jouanneau Si Larbi. An xps/aes comparative study of the surface behaviour of nano-silicon anodes for li-ion batteries. Journal of Analytical Atomic Spectrometry, 29(6):1120–1131, 2014.

[310] Giulio Ferraresi, Lukas Czornomaz, Claire Villevieille, Petr Novák, and Mario El Kazzi. Elucidating the surface reactions of an amorphous si thin film as a model electrode for li-ion batteries. ACS applied materials & interfaces, 8(43):29791–29798, 2016.

[311] Linghong Zhang, Yuzi Liu, Baris Key, Stephen E. Trask, Zhenzhen Yang, and Wenquan Lu. Silicon nanoparticles: Stability in aqueous slurries and the optimization of the oxide layer thickness for optimal electrochemical performance. ACS Appl. Mater. Interfaces, 9(38):32727–32736, 2017.

[312] Z Y Wu, F Jollet, and F Seifert. Electronic structure analysis of via x-ray absorption near-edge structure at the si k, and o k edges. J. Phys.: Condens. Matter, 10(36):8083, 1998.

[313] Jung-Fu Lin, Hiroshi Fukui, David Prendergast, Takuo Okuchi, Yong Q. Cai, Nozomu Hiraoka, Choong-Shik Yoo, Andrea Trave, Peter Eng, Michael Y. Hu, and Paul Chow. Electronic bonding transition in compressed sio_2 glass. Phys. Rev. B, 75:012201, 2007.

[314] K. Hämäläinen, S. Galambosi, J. A. Soininen, Eric L. Shirley, J.-P. Rueff, and A. Shukla. Momentum dependence of fluorine k-edge core exciton in lif. Phys. Rev. B, 65:155111, Mar 2002.

[315] E. Hudson, E. Moler, Y. Zheng, S. Kellar, P. Heimann, Z. Hussain, and D. A. Shirley. Near-edge sodium and fluorine k-shell photoabsorption of alkali halides. Phys. Rev. B, 49:3701–3708, Feb 1994.

[316] A. S. Vinogradov, S. I. Fedoseenko, S. A. Krasnikov, A. B. Preobrajenski, V. N. Sivkov, D. V. Vyalikh, S. L. Molodtsov, V. K. Adamchuk, C. Laubschat, and G. Kaindl. Low-lying unoccupied electronic states in $3d$ transition-metal fluorides probed by nexafs at the F $1s$ threshold. Phys. Rev. B, 71:045127, Jan 2005.

[317] Ruimin Qiao, Ivan T. Lucas, Altaf Karim, Jaroslaw Syzdek, Xiaosong Liu, Wei Chen, Kristin Persson, Robert Kostecki, and Wanli Yang. Distinct solid-electrolyte-interphases on sn (100) and (001) electrodes studied by soft x-ray spectroscopy. Advanced Materials Interfaces, 1(3):1300115, 2014.

[318] Ivana Hasa, Atetegeb M. Haregewoin, Liang Zhang, Wan-Yu Tsai, Jinghua Guo, Gabriel M. Veith, Philip N. Ross, and Robert Kostecki. Electrochemical reactivity and passivation of silicon thin-film electrodes in organic carbonate electrolytes. ACS Applied Materials & Interfaces, 12(36):40879–40890, 2020.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Elisa Molinari and Deborah Prezzi, for their constant guidance, unconditional support, and encouragement throughout the course of my PhD. Their invaluable advice and expertise have been fundamental in shaping this work. I am also grateful to UNIMORE and CNR for providing the institutional and financial support that made this research possible.

I would like to extend my appreciation to my group colleagues, Peter Gillespie and Ivan Marri, for their invaluable collaboration and support during this journey.

A special thanks to Michele Ceriotti for hosting me in his group at EPFL and guiding me through the initial stages into the machine learning field. I am also indebted to Federico Grasselli and Davide Tisi, who played a crucial role in training me in machine learning techniques. Thanks to the entire Laboratory of Computational Science and Modelling (COSMO) for making my experience at the EPFL so enriching.

Many thanks to the CEA members Sandrine Lyonnard and Diana Zapata for the fruitful collaboration on X-ray Raman scattering (XRS) studies of silicon nanoparticles. I truly appreciate their expertise and dedication throughout the project.

I also want to acknowledge the members of PSI, especially Giovanni Pizzi, Iurii Timrov, and Xing Wang, for their insightful discussions on core-level spectroscopies and their assistance with workflow implementation.

I am grateful to the Modeling Modena Group, including Andrea Ferretti, Daniele Varsano, and Claudia Cardoso, as well as my fellow PhD students Miki Bonacci, Simone Vacondio, and Andrea Vezzosi, for creating a stimulating and collaborative environment. A special mention goes to my office colleagues, Riccardo Pozzi and Luca Bursi, for their camaraderie and support.

I am also deeply thankful to my fellow countrymen, Dario Leon, Gabriel Gil, Frank Quintela, and Daniel Lopez, for their friendship and support during this journey and to my friend Emanuele Piazza for his companionship and guidance into the Italian ways. In particular, I would like to express my heartfelt thanks to my friend Dario, whose unconditional help with my PhD application, accommodation, and adaptation process in Italy was crucial.

Throughout this journey the presence of Olive has been truly invaluable. I am

profoundly grateful for her unwavering support and for enlivening this experience with her love and constant encouragement.

Finally, I want to express my gratitude to my family, specially to my mom and late grandmother, for their support and love, their sacrifices and dedication have profoundly shaped my path.

Thank you all.