

A Maximum Subspace Occupation Approach for the Study of the NEXAFS Spectra of Chemisorbed Organic Molecules Using Orthogonality Constrained Density Functional Theory: Pyrazine on Si(100) a Case Study “We may, I believe, anticipate that the chemist of the future who is interested in structures with high molecular weight will come to rely upon a new structural chemistry, involving precise geometrical relationships among the atoms in the molecules” Linus C. Pauling The near-edge X-ray absorption fine spectra (NEXAFS) at the carbon and nitrogen K-edge of pyrazine in the gas-phase and chemisorbed to the Si(100) surface are investigated using orthogonality constrained density functional theory (OCDFT). We introduce a novel approach for selectively target the 1s excitations of the adsorbate atoms based on atomic orbital subspace projection that allows for a priori specification of an atomic orbital subset relevant to the adsorbate. Using this approach we are able to simulate the NEXAFS spectrum of chemisorbed organic complexes at the K-edge of the adsorbate atoms. To analyze the calculated transitions we employ an orbital analysis based on localized intrinsic valence virtual orbitals (LIVVOs) in conjunction with OCDFT particle orbitals in order to provide a clear justification for every assignment. Utilizing the LIVVO analysis we are able to quantify the amount of π^*/σ^* mixing in each excited state and provide clear justification for changes in peak intensity.

1 Introduction

The adsorption of unsaturated hydrocarbons on the Si(100) surface has substantial technological significance:¹⁻³ exposed surface silicon dimers react with organic molecules forming covalent Si-C bonds which alters the reactivity of the silicon surface and opens up the possibility of fabricating novel semiconductor based molecular devices.⁴⁻⁷ This promise has inspired considerable investigation into the fundamental mechanisms that drive the chemisorption process.⁸⁻¹³ For this purpose, aromatic systems compose a particularly interesting class of molecules since their interaction with the surface often involves a myriad of unique bonding geometries that depend upon intrinsic properties such as the electronic and geometric structure of the molecule¹⁴⁻¹⁶ as well as extrinsic factors like temperature, coverage, and pressure.¹⁷⁻¹⁹ Aromatic molecules that contain a heteroatom can cause notable modifications the adsorption process, for example nitrogen-containing molecules have additional configurations that involve the donation of the nitrogen lone-pair electrons to the silicon.²⁰⁻²⁵ This occurs in the case of the pyrazine ($C_4H_4N_2$) molecule, a six membered aromatic heterocycle with electronic properties similar to those of benzene and pyridine. The ring system contains two nitrogen atoms at opposite ends (*para*), which increases the number of possible binding configurations upon molecular adsorption.¹⁵ Therefore the adsorption structure of pyrazine on the Si(100) surface is a very intriguing problem to study and has been the focus of many theoretical and experimental investigations.^{15,26-32}

High-resolution electron energy loss spectroscopy (HREELS) studies by Huang et al.²⁷ show that the adsorption of pyrazine on Si(100) occurs in a highly selective manner, directly bonding to the surface through the two *para* nitrogen atoms to form a 1,4-N-N-dihydropyrazine (DHP)-like structure (Figure 1b) attached to a single Si dimer. Density functional theory (DFT) cluster model calculations¹⁵ make a case for the N-end-on adsorbed pyrazine (Figure 1a) at low temperatures, and a species that is di- σ -bonded through the 2 and 5 carbons at elevated temperatures. DFT periodic slab studies by Jung and Kang²⁶ reveal that an isolated pyrazine molecule can bind in either the N-end-on configuration or di- σ -bonded cross-row configuration (Figure 1c) with equal preference but at high coverages the cross-row structure dominates and can form a linear molecular chain across the silicon dimer rows. This suggestion of the formation of a one-dimensional (1D) molecular chain is also in accordance with room temperature scanning tunneling microscopy (STM) and photoelectron diffraction (PED) studies by Shimomura et al.³¹ that suggests pyrazine forms this 1D molecular chain through a quasi-polymerization reaction through the Si dangling bonds. Pyrazine is one of the earliest examples of this type of self-assembled 1D molecular chain structure on a clean Si(100) surface.³⁰ This molecular assembly is particularly exciting in the case of pyrazine as it creates an ordered chain of planar C=C double bonds that could assist with further reaction at the semiconductor surface.

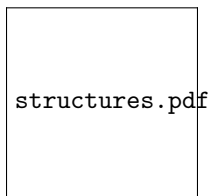


Figure 1: Schematic of three unique absorption modes of pyrazine ($C_4H_4N_2$) on Si(100), a) one with a single N atom attached at the Si surface dimer, and two structures that form a 1,4-N-N-dihydropyrazine (DHP)-like structure through forming two covalent Si-N bonds on a b) single dimer or c) across two dimer rows.

Near-edge X-ray absorption fine structure (NEXAFS) spectroscopy is a powerful experimental technique to elucidate modifications induced on the electronic and geometric structure of a molecule upon adsorption to a surface.^{33–36} One of the powerful features of NEXAFS in application to this class of molecules lies in the angular dependence of the spectral features on the light polarization. By obtaining angle resolved spectra, one can glean useful information about the orientation of the molecule with respect to the surface.^{37–39} The fully-polarization resolved C K-edge spectrum of pyrazine on the surface of Si(100) was obtained by Han-Koo et al.²⁸ at 300K and 2L coverage.⁴⁰ Their analysis showed an angular dependence of the intensity of the π^* resonance relating to the orbitals of the C=C double bonds, resulting in an average tilt angle of the adsorbate with respect to the surface of approximately $34 \pm 5^\circ$. This result coupled with data from X-ray Photoelectron Spectroscopy, effectively eliminates all possible adsorption configurations except for the DHP-like cross-row configuration (See Figure 1c).

Analysis of the NEXAFS spectral features benefit greatly from a theoretical treatment. Calculations can provide useful insight into the relationship between the spectral features and structure.⁴¹ In this regard, NEXAFS simulations of adsorbed species are particularly challenging because they require an accurate description of the adsorbate/substrate geometry and the core excited states of large systems.^{20,42,43} Finite cluster models are often employed to represent the adsorbed system as they are particularly adept at modeling phenomena that are localized on the surface or within the bulk of a solid.^{42,44} Interaction between the adsorbate and substrate model causes overlap of their molecular orbitals (MOs) and induces a rehybridization of the valence level. The position and intensity of features in the core spectrum will show modifications relative to the free molecule and can thus capture the influence that the surface has on the local electronic structure of the adsorbing molecule. Prior NEXAFS simulations for benzene and pyridine on Si(100) have employed finite models with as few as 9 to as many as 21 Si atoms.⁴⁵ The goal is to obtain an accurate description of the limited region around the adsorbate to capture the dominant effects on the NEXAFS spectrum due to the localization of core excited states near the core hole. Previous work by Rangan et al. has investigated the effect of increasing cluster size on the the core excitation spectrum of organic heterocycles adsorbed to Si(100).^{46,47} These studies show that increasing the number of Si atoms in the cluster model had a very small effect on the resulting spectrum with regard to peak position and intensity. Recent work by Romeo et al. on pyridine/Si(100) compared the results of indepently optimized small clusters to larger clusters cut from periodic slab calculations.²⁰ These results highlighted differences in the individual polarized components of the spectrum, however, the effect on the total spectrum appears to be minimal. The computational efficiency and previous succesful applications of finite cluster models has motivated their use for modeling the surface geometry in the current study with the understanding that any potential long-range interactions of the adsorbate with the extended surface will be neglected.

Aside from the treatment of the adsorbate/substrate geometry, computational approaches must calculate core excitation energies and properties in order to simulate the NEXAFS spectrum. The accuracy of calculated energies depends upon a reliable treatment of correlation, relaxation, and relativistic effects. Highly accurate many-body approaches have had considerable success in this area including coupled cluster theory,^{48–52} configuration interaction,^{53–55} and Green’s-function approaches.^{56,56–58} However, the computational cost associated with these methods limits their application to smaller systems. Considering the number of atoms necessary to construct cluster model surfaces it is desirable to consider methods that are computationally feasible for application to larger problems. Approaches that offer reduced computational cost include linear response time-dependent density functional theory (TDDFT),^{42,59,60} Hartree–Fock static exchange method, configuration interaction singles,⁶¹ Δ SCF, and real-time TD-DFT.⁶² We have recently proposed an alternative DFT based approach known as orthogonality constrained DFT (OCDFT)⁶³ and have shown that it yields highly accurate core excitation energies.^{64,65} It is a Δ SCF based Kohn–Sham DFT method which avoids variational collapse of the SCF solution by explicitly invoking an orthogonality constraint between the ground and excited states. Preliminary benchmark studies showed that OCDFT can accurately treat core excitations, providing absolute excitation energies that are more accurate than TDDFT.

Core excitation energies are calculated in OCDFT through the use of the constrained multiple hole particle (CMHP) algorithm. Similar to other excited state algorithms, CMHP is built to sequentially target excited state solutions starting from the lowest (highest) energy solution in a bottom-up (top-down) fashion. This presents an issue when targeting core excitations of organic adsorbates since the highest energy solutions are valence excitations and the lowest energy solutions are core excited states related to atomsof the cluster model. Thus, starting from either extrema requires the calculation of multiple unwanted solutions before reaching the adsorbate core. Having the ability to bypass these unwanted states and specifically target the core orbitals of the adsorbate is highly desirable.

Strategies to target specific core orbitals of interest are routinely employed to circumvent algorithmic challenges that exist in the calculation of core excited states. These methods typically fall into the category of either restricted excitation window (REW)^{62,66,67} or energy-specific (ES) techniques.^{60,68} These two methods differ by the way they solve the linear response equations. In the case of REW, the molecular orbital (MO) space from which excitations

are allowed is restricted using either an orbital energy cutoff or MO Mulliken populations. In ES methods, the linear response equations are solved in the full MO space targeting eigenvalues that lie above a predefined energy threshold. To the best of our knowledge, similar strategies for targeting specific excitations have not been explored within the context of variational DFT methods.

Another theoretical challenge that is encountered by all methods is appropriately assigning transitions in chemisorbed complexes. Often times when classifying the NEXAFS spectra of gas-phase molecules it is sufficient to simply classify the transition as σ or π , which can easily be done by inspecting the character of the molecular orbitals involved in the final state. In the case of chemisorbed molecules it is also important to specify which atoms are major contributors to the final state orbital. This is crucial, since a large part of interpreting chemisorbed spectra is deciphering which peak features are a result of interactions with the surface and which features are largely localized on the adsorbate. Determining this through simple inspection of the MOs can be ambiguous, and thus a more quantitative approach toward determining this is desirable.

In this paper we introduce a new approach for targeting core-excitations in adsorbed molecules within OCDFT based on an atomic orbital subspace occupation analysis. Using this method we can circumvent the algorithmic difficulty of calculating core excitations from multiple low lying surface atoms and specifically focus on the adsorbate atoms of interest. By developing this method around specification of atomic orbitals it requires little prior knowledge of the electronic structure of the molecule unlike other reduced subspace methods where exact energy ranges are required. We also address the issue of assigning transitions using a classification method based on localized intrinsic valence virtual orbitals.⁶⁹ We utilize these tools in order to analyze the NEXAFS spectrum of pyrazine chemisorbed on a Si(100) surface.

2 Method

Orthogonality Constrained Density Functional Theory

OCDFT is a time-independent variational density functional method⁷⁰ used to calculate electronic excited states. The original formulation of the theory can be found in ref 63, while details on its extension to treat core excitations and the algorithm used to calculate multiple excited states to simulate full NEXAFS spectra can be found in ref 64. In OCDFT, a generalized Kohn–Sham picture is assumed, where to the n^{th} electronic state there exists an auxiliary system of noninteracting electrons with wave function $\Phi^{(n)}$ and density $\rho^{(n)}$ that corresponds to the real interacting system $\Psi^{(n)}$. The wave function for the auxiliary system is a single Slater determinant, $\Phi^{(n)} = [1]\phi_1^{(n)}\phi_2^{(n)}\cdots\phi_N^{(n)}$, where the set of orbitals $\{\phi_i^{(n)}\}$ are different for each electronic state.

$$\langle\Phi^{(m)}|\Phi^{(n)}\rangle = \delta_{mn} \quad \forall m, n \quad (1)$$

It can be shown, that given this constraint, one-electron excited states can be characterized by variationally optimized hole (ϕ_h) and particle (ϕ_p) orbitals, which must span the occupied and virtual spaces of the ground state wave function, respectively. In this work we employ our constrained multiple hole/particle (CMHP) algorithm for the solution of multiple orthogonally constrained excited states.⁶⁴ This allows us to readily compute multiple excited states by enforcing mutual orthogonality between the hole and particle orbitals of each subsequent excited state. At the same time, the CMHP scheme fully accounts for relaxation of all orbitals.

Maximum Subspace Occupation

Here we introduce our approach for selectively targeting the core excitations of interest by combining OCDFT with a scheme to target specific hole orbitals. We start by considering a set of atomic orbital (AO) basis functions φ_s that are ordered by atom center, principal quantum number, and angular momentum. Based on this criteria, we can specify a subset of AO basis functions s that are of interest in a given system. For NEXAFS K-edge applications this is going to be the 1s core orbital(s) centered on the atom(s) of interest. The specified basis functions can be used to build an operator ($\hat{\Gamma}_s$) that projects onto the AO subset.

$$\hat{\Gamma}_s = \sum_s [1]\varphi_s[1]\varphi_s \quad (2)$$

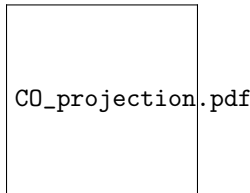


Figure 2: Example application of the maximum subspace occupation method within OCDFT to target the C 1s core excitations in CO at the B3LYP/3-21G level of theory: a) the 7 occupied orbitals of CO in C_{2V} symmetry, with the corresponding subspace occupation number (Ω_i) for a subspace composed of the C 1s atomic orbital. b) These projections are used to build the subset of occupied orbitals, in this example a hole threshold of 0.2 is used, and thus orbital $2A_1$ is chosen as the target hole orbital. This allows for direct access to the relevant C K-edge transitions.

Table 1: First 3 carbon core excitation energies and oscillator strengths (f_{osc}) in a CCl_4 molecule calculated with and without the maximum subspace occupation (MSO) method in OCDFT. All calculations were performed using the B3LYP functional and aug-cc-pvdz basis set.

State	OCDFT		MSO-OCDFT	
	ω (eV)	f_{osc}	ω (eV)	f_{osc}
1	292.0002	0.0000	292.0005	0.0000
2	292.7093	0.0384	292.7097	0.0384
3	292.6669	0.0385	292.6659	0.0385

Utilizing this operator, we can evaluate the atomic orbital occupation number (Ω_i) of each molecular orbital ϕ_i within the desired subspace.

$$\Omega_i = [1]\phi_i\hat{\Gamma}_S[1]\phi_i = \sum_S [1]\phi_i\varphi_S[1]\varphi_S\phi_i. \quad (3)$$

This is a number between 0 (no occupation in the AO subset) and 1 (full occupation in the AO subset). We have implemented this maximum subspace occupation (MSO) scheme in OCDFT for the selection of the hole orbital. Now instead of being chosen from the full occupied set, the hole orbitals ϕ_h are chosen from a subset of occupied orbitals ϕ_i with $\Omega_i \geq \omega$ where ω is a user-defined occupation threshold parameter. Figure 2 shows an example application of this subspace projection scheme to the calculation of carbon core excitations in carbon monoxide. In this example, the $C_{1s} \rightarrow \pi^*$ transition is targeted without needing to calculate any O_{1s} excitations beforehand. This projection scheme which is based on the nature of the atomic orbital provides a distinct advantage to traditional REW and ES methods, as it requires no prior knowledge of the orbital energies or relevant energy range of absorption for a given system.

Table 1 shows the first 3 carbon core excitations in a CCl_4 molecule calculated with and without the subspace projection method. Comparison of these two approaches shows a negligible difference between the excitation energies (less than 0.001 eV) and oscillator strengths. The case of targeting the C K-edge of CCl_4 is comparable in this sense to targeting the C K-edge of chemisorbed organics. In order to compute three transitions of the C K-edge using the standard algorithm, one must first compute twelve excited states related to the Cl K-edge (three for each Cl atom). However, when applying the subspace projection technique, the C 1s orbitals can be targeted in an efficient manner without the need to calculate any Cl core excitations.

2.1 Assigning Transitions Based on Localized Intrinsic Valence Virtual Orbitals

In this section, we summarize our method for classifying excited states based on a novel localized orbital representation known as the *localized intrinsic valence virtual orbitals* (LIVVOs), only a brief summary is given here, for full details on this technique we refer the reader to Ref. 69. This unique set of orbitals is a special case of the valence virtual orbitals (VVOs) developed by Rudenberg and coworkers.⁷¹ The goal of VVOs is to obtain a more chemically meaningful orbitals through evaluation of the valence character of the canonical molecular orbitals. This is done by identifying an accurate atomic minimal basis set (AAMBS) and evaluating the overlap of the canonical virtual orbitals with the AAMBS functions ψ_ρ^{AAMBS} such that:

$$(\mathbf{S})_{a\rho} = \langle \phi_a | \psi_\rho^{AAMBS} \rangle. \quad (4)$$

For our implementation we have chosen to use the intrinsic atomic orbitals of Kinizia⁷² as the necessary AAMBS functions. Following the suggestion of Ref. 73 a singular value decomposition (SVD) is performed on \mathbf{S} :

$$\mathbf{S} = \mathbf{U}\boldsymbol{\sigma}\mathbf{V}^\dagger \quad (5)$$

to yield the orthogonal transformation matrices \mathbf{U} and \mathbf{V} . These are rotations of the canonical virtual space and the AAMBS space, respectively that bring the two orbital sets into maximum coincidence. The transformation matrix \mathbf{U} is then applied to the virtual set in the following way:

$$|\psi_\nu^{\text{VVO}}\rangle = \sum_a^{N_{\text{vir}}} |\phi_a\rangle U_{a\nu}, \nu = 1, 2, \dots, N_{\text{VVO}} \quad (6)$$

where N_{vir} is the number of virtual orbitals while N_{VVO} is the number of VVOs. Finally, the set of LIVVOs (ψ_l^{LIVVO}) are formed by using Pipek-Mezey localization,⁷⁴ we arrive as a set of localized VVOs that span the antibonding interactions of the molecular environment.

For every excited state in OCDFT, we analyze the character of the particle orbital $\phi_p^{(n)}$ by evaluating its overlap with each LIVVO:

$$\Omega_{\text{pl}}^{(n)} = |\langle \phi_p^{(n)} | \psi_l^{\text{LIVVO}} \rangle|^2 \quad (7)$$

The individual overlaps $\Omega_{\text{pl}}^{(n)}$ are used to assign the character of the particle orbital to the l -th LIVVO. In addition, we also define the total valence character $t_p^{\text{val},(n)}$ for any given particle orbital as the sum of its overlap with all LIVVOs:

$$t_p^{\text{val},(n)} = \sum_l^{N_{\text{LIVVO}}} \Omega_{\text{pl}}^{(n)}. \quad (8)$$

With this set of orbitals it is possible to quantify the most important local contributions to the particle orbitals to provide a specific description of its character to make a robust assignment for the excited state.

3 Computational Details

The gas-phase geometry of pyrazine was optimized using the B3LYP^{75–78} functional and the 6-31G*^{79–81} basis set. All core excited states were calculated using OCDFT in the PSI4 *ab-initio* quantum chemistry package.⁸² Core excited state calculations on the free molecule were done at the PBE/cc-pVTZ^{83,84} level of theory. Pyrazine is then attached in the double dimer bridging configuration to a Si₂₃H₂₄ cluster model of the Si(100) surface reconstruction. Covalent Si-N bonds are formed between adjacent dimer rows and the *para* nitrogen atoms of pyrazine. Core excitation calculations on chemisorbed pyrazine utilize the PBE functional⁸³ and a custom basis set where the cc-pVTZ basis set⁸⁴ is used for all carbon and hydrogen atoms and the 6-31G basis set⁸⁵ for all silicon atoms. The AO subspace is chosen as the C 1s or N 1s orbitals to simulate their respective K-edges with a hole threshold of 0.2. This threshold ensures that the only hole orbitals chosen are MOs corresponding to carbon core orbitals.

We note that pyrazine is a highly symmetrical molecule containing symmetry equivalent C and N atoms. Molecules that contain symmetry equivalent cores present a challenging theoretical issue.^{86,87} In practice, symmetry restricted calculations produce a core hole that is delocalized evenly amongst all symmetry equivalent atoms. In contrast, a symmetry unrestricted calculation may produce core holes that are localized on each individual atom. For calculations on the free pyrazine molecule, we have studied both the symmetry restricted and unrestricted solutions. To obtain a state where the core hole is localized on each atom we utilize a wavefunction with broken spatial and spin symmetry by mixing the coefficients of the alpha HOMO and LUMO orbitals of pyrazine.

For spectral simulations, the peak intensities are based on the oscillator strength (f_{osc}):

$$f_{\text{osc}} = \frac{2}{3} |\mu_{n0}|^2 \omega_n \quad (9)$$

where μ_{n0} represents the transition dipole moment for transition from the ground state to excited state n and ω_n is the excitation energy for that transition. The spectrum is simulated by centering Gaussian functions at each excitation energy with a peak height scaled by the oscillator strength of that transition. The peak is then broadened by a constant value in order to simulate natural spectroscopic broadening effects, for the purpose of pyrazine we utilized a Gaussian width of 0.3 eV.

pyrazine_nexafs_gas.pdf

Figure 3: Near-edge X-ray absorption spectrum of pyrazine ($C_4H_4N_2$) at the a) C K-edge and b) N K-edge computed with OCDFT (blue line) at the PBE/cc-pVTZ level of theory. Experimental results (black line) are obtained from Ref. 88 and appear here in their original peak position with no applied energy shifts. The pyrazine geometry optimized at the B3LYP/6-31G* level is shown in the inset, the atoms circled (C_1 and N_4) are the core holes considered in further discussion of the results in this section.

pyrazine_gas_parts_and_Livvos.pdf

Figure 4: Particle orbitals for the core excited states that contribute to the a) carbon and b) nitrogen 1s excitation spectrum. Also shown are c) the LIVVOs that are used to analyze the particle orbitals.

4 Results and discussion

NEXAFS Spectrum of Gas-Phase Pyrazine

In this section we present an OCDFT simulation of the carbon and nitrogen 1s excitation spectrum for the free pyrazine molecule. Figure 3 shows the experimental and calculated NEXAFS spectrum along with corresponding peak labels. The relevant particle orbitals and the LIVVOs used to characterize each excited state are shown in Figure 4. Calculated excitation energies, oscillator strengths, and assignments are shown in Table 2.

Pyrazine, similar to other unsaturated molecules, follows a common pattern to their core excitation spectrum, a dominant edge onset associated with a π^* resonance localized near the core hole followed by a series of weaker transitions that are a mix of σ^* transitions, Rydberg transitions, and π^* transitions localized away from the core hole. The experimental C 1s spectrum shown in Figure 3a follows this general trend. The Experimental spectrum is characterized by four main peaks at 285.3, 285.8, 288.2, and 289.1 eV. The simulated spectrum using OCDFT agrees well with corresponding peak features at 284.5, 285.2, 288.1, and 289.1 eV respectively.

The OCDFT spectrum for the C K-edge was calculated considering excitations from carbon atom C_1 (see inset in Figure 3), due to the symmetry of the molecule, it is obvious that the equivalent carbon cores, would produce an identical core excitation spectrum. Limiting our discussion to a single core also aides in our discussion of localization when analyzing our LIVVO assignments. The first peak A in the carbon 1s spectrum occurs at 284.5 eV with a small shoulder feature A' ocuring at 285.2 eV. These features can both be attributed to π^* transitions that are a combination of the $\pi_{C_1-N_3}^*$ and $\pi_{C_7-C_8}^*$ LIVVOs. In the case of Peak A, 78.2% of the particle orbital can be attributed to $\pi_{C_1-N_3}^*$ which is localized near the core hole on atom C_1 . On the contrary, the particle orbital for the shoulder feature A' only 51.1% of the particle orbital can be attributed to $\pi_{C_1-N_3}^*$, while 43.9% is attributed to $\pi_{C_7-C_8}^*$ which is localized away from the core hole. The calculated oscillator strength is sensitive to this change in localization, with state 1 having an oscillator strength of 0.0207 and state 2 having an oscillator strength of 0.0020. State 3 is the sole contribution to Peak B, with an excitation energy of 288.1 and an orbital character that can be assigned as $\sigma_{C_1-H_5}^*$.

States 3 and 4 are quasi-degenerate both yielding excitation energies of 289.1 eV. Looking at their particle orbitals $\phi_p^{(4)}$ and $\phi_p^{(5)}$ in Figure 4, it is clear that they are orbitals of mixed π^* and σ^* character. This orbital mixing was noted in a previous study⁸⁸ however never quantified in any manner, using the LIVVO analysis we are able to quantify this orbital mixing, this data is shown in Table 2. State 4 is 65% π^* and 35% σ^* , while state 5 is 61% π^* and 39% σ^* . This similarity in the composition of their orbital character explains the similarities between their energy and oscillator strength. A weak transition ascribed to a $\sigma_{C_2-H_6}^*$ at 289.4 eV also contributes to Peak C. States 7 and 8 compose peak D at 290.1 and 290.2 eV. They are assigned to $\sigma_{C_1-N_3}^*$ and $\sigma_{C_8-H_{10}}^*$ respectively.

Table 2: Calculated (PBE/cc-pVTZ) core excitation energies, oscillator strengths, and LIVVO assignments for the C and N 1s excitation spectrum of free pyrazine. The LIVVO assignments shown are the two with the highest percentage overlap with the particle orbital. For each calculated transition we report the π^*/σ^* mixing ratio (see text for details) and the spectral feature that this state contributes to (see Figure 3).

State	energy (eV)	osc	LIVVO Assignments	π^*/σ^* Mixing	t_p^{val}	Peak Contribution
C 1s Core Excitations						
1	284.5	0.0207	78.2% $\pi_{C_1-N_3}^*$ 19.3% $\pi_{C_7-C_8}^*$	100% π^*	97.9%	A
2	285.2	0.0020	51.1% $\pi_{C_1-N_3}^*$ 43.9% $\pi_{C_7-C_8}^*$	100% π^*	97.3%	A'
3	288.1	0.0083	45.4% $\sigma_{C_1-H_5}^*$ 12.0% $\sigma_{C_7-H_9}^*$	100% σ^*	64.1%	B
4	289.1	0.0042	28.6% $\pi_{N_3-C_7}^*$ 7.4% $\sigma_{C_1-H_5}^*$	65% $\pi^*/35\%\sigma^*$	76.4%	C
5	289.1	0.0041	26.0% $\pi_{C_7-C_8}^*$ 7.9% $\sigma_{C_1-H_5}^*$	61% $\pi^*/39\%\sigma^*$	74.6	C
6	289.4	0.0003	25.7% $\sigma_{C_2-H_6}^*$ 14.6% $\sigma_{C_7-H_9}^*$	100% σ^*	54.0	C
7	290.1	0.0027	20.5% $\sigma_{C_1-N_3}^*$ 12.8% $\sigma_{C_2-N_4}^*$	100% σ^*	59.5	D
8	290.2	0.0037	17.4% $\sigma_{C_8-H_{10}}^*$ 13.7% $\sigma_{C_7-H_9}^*$	100% σ^*	49.9	D
N 1s Core Excitations						
1	397.8	0.0211	59.4% $\pi_{C_2-N_4}^*$ 34.8% $\pi_{C_1-N_3}^*$	100% π^*	98.4%	A
2	399.3	0.0000	60.3% $\pi_{C_7-N_8}^*$ 22.2% $\pi_{N_4-C_8}^*$	100% π^*	96.9%	
3	402.4	0.0003	19.3% $\sigma_{C_2-H_6}^*$ 19.3% $\sigma_{C_8-H_{10}}^*$	100% σ^*	64.1%	B
4	402.8	0.0042	38.3% $\pi_{C_1-N_3}^*$ 6.0% $\sigma_{C_2-H_6}^*$	88% $\pi^*/12\%\sigma^*$	85.4	B
5	403.2	0.0010	17.9% $\sigma_{C_8-H_{10}}^*$ 8.9% $\pi_{C_1-N_3}^*$	29% $\pi^*/71\%\sigma^*$	58.7%	C
6	403.6	0.0034	21.6% $\sigma_{C_2-N_4}^*$ 21.6% $\sigma_{N_4-C_8}^*$	100% σ^*	68.3	C
7	403.8	0.0001	13.2% $\sigma_{C_7-H_9}^*$ 12.9% $\sigma_{C_1-H_5}^*$	100% σ^*	59.5	D
8	403.9	0.0024	16.0% $\sigma_{C_1-H_5}^*$ 15.6% $\sigma_{C_7-H_9}^*$	100% σ^*	52.4	D

The N 1s core excitation spectrum shown in Figure 3b shows a similar general structure to the C 1s spectrum. Peak A occurs at 397.8 eV and can be assigned to $\pi_{C_2-N_4}^*$ which is the π^* orbital localized near the core hole. Interestingly, state 2, which has an excitation energy of 399.3 eV has an oscillator strength of zero providing no contribution to the spectrum. The LIVVO analysis shows that 60.3% of this partial orbital can be attributed to the $\pi_{C_7-N_8}^*$ orbital, which is localized away from the nitrogen core hole at N₄. Also, the particle orbital $\phi_p^{(2)}$ shown in Figure 4 shows that the orbital has no contribution localized at the nitrogen atoms. State 3 is a small contribution to Peak B at 402.4 eV attributed evenly to $\sigma_{C_2-H_6}^*$ and $\sigma_{C_8-H_{10}}^*$. As with State 2, much of the orbital character is localized away from the nitrogen resulting in a much weaker transition. As seen in the C 1s spectrum, States 4 and 5 are of mixed π^* and σ^* character. However, unlike the mixed states in the C spectrum, the composition of these orbitals in the N spectrum are unique. State 4, which contributes to Peak B at 402.4 eV, contains 88% π^* character and only 12% σ^* character. This state is assigned primarily to the $\pi_{C_1-N_3}^*$ orbital localized near the core hole, resulting in an appreciable oscillator strength of 0.0042. Although State 5 is also a mixed state it contains far more σ^* character, with only 29% π^* character and 71% σ^* character. This state can be assigned to the $\sigma_{C_8-H_{10}}^*$ LIVVO, which explains its weak intensity in the N 1s spectrum. State 6 is a more intense contribution to Peak C at 403.6 eV and is assigned to $\sigma_{C_2-N_4}^*$ and $\sigma_{C_8-N_4}^*$, the two C-N σ -bonds associated with the core hole N atom. States 7 and 8 contribute to Peak D at 403.8 and 403.9 eV respectively and are composed of transitions to σ^* orbitals associated with C-H bond along the ring system.

NEXAFS Spectrum of Chemisorbed Pyrazine on Si(100)

The OCDFT simulated NEXAFS spectrum of pyrazine chemisorbed on the Si(100) surface is shown in Figure 5. The LIVVOs shown in Figure 6 are only those that are directly involved with the adsorbed pyrazine molecule. In total, this system produces 70 LIVVOs however many of these are $\sigma_{\text{Si-Si}}^*$ and $\sigma_{\text{Si-H}}^*$ orbitals related to the Si(100) cluster model. When comparing the LIVVOs of the chemisorbed complex to that of free pyrazine, many of the orbitals are remain unchanged except for three critical differences. First, the most obvious difference is the appearance of the $\sigma_{\text{N}_3-\text{Si}_{11}}^*$ and $\sigma_{\text{N}_4-\text{Si}_{12}}^*$ LIVVOs to reflect the covalent bonds formed between the cluster and the para nitrogens on pyrazine. Secondly, the π system of the molecule has changed and the LIVVOs reflect that, the π^* orbitals are not localized along the N-C bonds, as is the case in free pyrazine, they are now localized along each C=C double bond with clear contributions from the N lone pair electrons. Lastly the $\sigma_{\text{Si}_{13}-\text{Si}_{14}}^*$ represents the dangling bonds that exist at each open Si site on the dimers. These changes in the character of the LIVVOs gives us some insight into the effect of the cluster on the adsorbate.

Since pyrazine maintains two double bonds upon binding on the surface, the NEXAFS spectrum can be expected to maintain the same general spectral pattern of a sharp π^* peak and a weaker σ^* manifold. The experimental NEXAFS spectrum of chemisorbed pyrazine reports the initial spectral feature at 285.9 attributed to a π^* excitation. Our results show an initial excitation at 279.4 eV that is attributed to a $\sigma_{\text{Si}_{13}-\text{Si}_{14}}^*$, which is the orbital associated with he open Si atoms. The appearance of this state in our calculations is purely an artifact of the cluster model. In reality, pyrazine forms a 1D molecular chain along the surface, leaving few open Si sites on the surface and thus quenching this state. State 2 at 283.8 eV agrees well with the experimental onset feature, is assigned to π^* LIVVVOs, and is the sole contribution to Peak A in the spectrum. Considering the core hole at atom C₈ State 2 is localized heavily toward the core hole, with 73.4% of the orbital character attributed to the $\pi_{\text{C}_7-\text{C}_8}^*$ LIVVO, this is similar to the localization along a single π^* bond seen in the gas phase spectrum where the first state had 78.2% overlap with a single π^* LIVVO. The similarity between the LIVVO localization of the π^* state in the gas and condensed phases considered here shows that the localization of this π^* state is not significantly affected upon adsorption to the surface. Due to this, there is a large difference in intensity at the C K-edge of chemisorbed pyrazine between the π^* peak and the σ^* manifold, that didn't exist in the spectrum of free pyrazine. The crucial difference here is the interaction and subsequent rehybridization of the pyrazine valence orbitals with the Si cluster. Looking at the particle orbital $\phi_p^{(2)}$ it is clear that this initial π^* state interacts very little with the Si cluster, meanwhile every other particle orbital has significant contribution from Si cluster orbitals. This can also be seen quantitatively by looking at the LIVVO assignments in Table 3, for the π^* state (State 2), a single VVO constitutes 76% of the total valence character, in contrast no other state has a single VVO contribution of over 20%. As a result, the σ^* manifold for the C 1s spectrum is orders of magnitude less intense

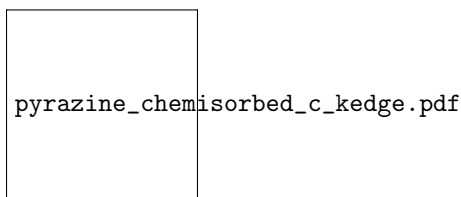


Figure 5: Near-edge X-ray absorption spectrum of pyrazine ($\text{C}_4\text{H}_4\text{N}_2$) chemisorbed on a $\text{Si}_{23}\text{H}_{24}$ cluster model of Si(100) at the a) C K-edge and b) N K-edge computed with OCDFT at the PBE level. All C, N, and H atoms use the cc-pVTZ basis set while the 6-31G basis set is used for the Si atoms of the cluster model. The inset in the C spectrum magnifies the σ^* manifold of the spectrum. The pyrazine geometry optimized at the B3LYP/6-31G* level is shown in the inset in the N spectrum, the atoms circled (C₈ and N₄) are the core holes considered in further discussion of the results in this section.

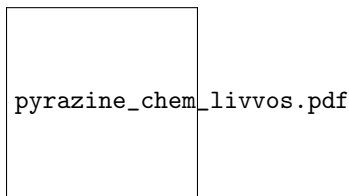


Figure 6: LIVVOs used to analyze particle orbitals in pyrazine ($\text{C}_4\text{H}_4\text{N}_2$) chemisorbed on a $\text{Si}_{23}\text{H}_{24}$ cluster model of Si(100).

Figure 7: Particle orbitals for the core excited states that contribute to the a) carbon and b) nitrogen 1s excitation spectrum of pyrazine ($C_4H_4N_2$) chemisorbed on a $Si_{23}H_{24}$ cluster model of Si(100).

than the π^* feature.

The N 1s spectrum shown in Figure 5b, does not show the same stark contrast in intensity as the C spectrum. This is due to the fact that the π^* orbitals are now localized away from the N atoms upon the formation of the covalent Si-N bonds. In fact, in contrast to all other spectra considered here, the pure π^* state (State 10) is the highest in energy at 401.6 eV and has an oscillator strength of zero. Interestingly this is not the only state to provide no intensity to the spectrum. States 6 and 7 also provide no intensity to the spectrum due to their interaction with the cluster, both states have no contributions from LIVVOs that are localized on the pyrazine molecule and their respective particle orbitals in Figure 7b ($\phi_p^{(6)}$ and $\phi_p^{(7)}$) clearly shows no interactions from pyrazine.

Peak A of the N 1s spectrum has contributions from States 1 and 2, similar to the C 1s spectrum this first state is an artifact of the cluster from the open Si sites. State 2 however is a contribution from the Si-N bond, $\phi^{(2)}_p$ has 39.7% overlap with the $\sigma^*_{N_4-Si_{12}}$ LIVVO. Peak B has contributions from States 3 and 4 both of which are mixed π^*/σ^* states. Similar to the C 1s spectrum, this ratio favors σ^* character due to interaction with the cluster model. The sole contribution to Peak C is a weak transition attributed to the π^* orbitals with 5.2% overlap with $\pi^*_{C_7-C_8}$ and 4.9% overlap with $\pi^*_{C_1-C_2}$.

5 Conclusions and Future Work

In this work we have introduced a subspace projection method for orthogonality constrained density functional theory in order to calculate the NEXAFS spectrum of chemisorbed organic molecules. Core excitation energies computed with subspace projected OCDFT are in agreement with those computed using full OCDFT due to the localized nature of these excited states they are flawlessly decoupled from other excitations. Throughout the chemisorbed molecules considered here we notice similar agreement to experiment as was observed for gas-phase organic molecules. We also highlight a powerful tool for analyzing the character of the transitions involved in NEXAFS spectra using an atomic decomposition of the transition dipole moment. These can assist in more complex cases where simply categorizing an orbital as π or σ is not sufficient. Our analysis of the NEXAFS spectrum in this way is more thorough than the typical directionality based arguments that are generally made to classify transitions of chemisorbed molecules.

We utilize this approach in order to critically analyze the NEXAFS spectrum of pyrazine chemisorbed on the Si(100) surface. Utilizing our subspace projection approach we were able to efficiently target the organic core orbitals of the pyrazine molecule without the need to calculate any lower lying Si core excited states. Applying our LIVVO analysis to assign excited states we were able to track local changes related to the pyrazine molecule upon chemisorption to the surface. Providing, for the first time, a critical theoretical analysis of the NEXAFS spectrum for pyrazine chemisorbed on Si(100). Using LIVVOs we were able to quantify the amount of π/σ mixing in each state. Mixing of σ orbitals from the cluster results in fewer states that are purely π^* or purely σ^* with most states showing some degree of mixing. This analysis was most useful in identifying and quantifying differences and changes in intensity. For example, the OCDFT C 1s spectrum for chemisorbed pyrazine showed a large discrepancy between the π and σ manifolds of the spectrum and analysis of the LIVVOs revealed that the π^* state had 73.4% overlap with a single π^* LIVVO while all other states have, 18.5% overlap or less with a single pyrazine LIVVO resulting in significantly weaker transitions.

This subspace projection method in OCDFT lends itself well to future work, the most obvious of which would be the computation of the NEXAFS spectrum of solvated organic molecules. These systems suffer from the same computational hurdle as chemisorbed molecules, where the organic core orbitals of interest are often times higher in energy than those of the solvent molecules. Also for those applications it could be useful to extend this subspace projection formalism to also constrain the choice of the particle orbital to only include unoccupied MOs relevant to the solute. Another potential application is to explore the use of our subspace projection scheme in the context of other excited

Table 3: Calculated (PBE/cc-pVTZ) core excitation energies, oscillator strengths, and LIVVO assignments for the C and N 1s excitation spectrum of pyrazine ($C_4H_4N_2$) chemisorbed on a $Si_{23}H_{24}$ cluster model of Si(100). The LIVVO assignments shown are the two with the highest percentage overlap with the particle orbital. For each calculated transition we report the π^*/σ^* mixing ratio (see text for details) and the spectral feature that this state contributes to (see Figure 5).

State	energy (eV)	osc	LIVVO Assignments	π^*/σ^* Mixing	t_p^{val}	Peak Contribution
C 1s Core Excitations						
1	279.4	0.0002	83.8% $\sigma_{Si_{13}-Si_{14}}^*$	1% $\pi^*/99\%\sigma^*$	98.2%	
2	283.8	0.0210	73.4% $\pi_{C_7-C_8}^*$ 10.5% $\pi_{C_1-C_2}^*$	86% $\pi^*/24\%\sigma^*$	97.0%	A
3	284.9	0.0005	11.8% $\sigma_{N_3-Si_{11}}^*$ 10.1% $\sigma_{N_4-Si_{12}}^*$	100% σ^*	92.2%	B
4	285.3	0.0002	18.5% $\pi_{C_1-C_2}^*$ 13.2% $\sigma_{N_3-Si_{11}}^*$	21% $\pi^*/79\%\sigma^*$	92.2%	C
5	285.6	0.0006	9.7% $\pi_{C_1-C_2}^*$ 4.0 % $\sigma_{C_8-H_{10}}^*$	12% $\pi^*/88\%\sigma^*$	89.5%	C
6	285.8	0.0005	11.4% $\pi_{C_7-C_8}^*$ 5.7 % $\sigma_{N_3-Si_{11}}^*$	14% $\pi^*/86\%\sigma^*$	89.9%	C
7	286.4	0.0003	10.5% $\sigma_{C_8-H_{10}}^*$ 1.7 % $\sigma_{N_3-Si_{11}}^*$	100% σ^*	87.2%	D
8	286.3	0.0000	2.1% $\pi_{C_7-C_8}^*$ 0.6% $\pi_{C_7-C_8}^*$	3% $\pi^*/97\%\sigma^*$	91.0%	D
9	286.9	0.0005	13.3% $\sigma_{C_8-H_{10}}^*$ 1.6% $\sigma_{C_7-H_9}^*$	2% $\pi^*/98\%\sigma^*$	83.9%	D
10	287.8	0.0007	12.0% $\sigma_{N_4-Si_{12}}^*$ 3.7 % $\sigma_{N_4-C_8}^*$	4% $\pi^*/96\%\sigma^*$	89.9%	E
N 1s Core Excitations						
1	398.2	0.0026	85.1% $\sigma_{Si_{13}-Si_{14}}^*$ 1.3 % $\sigma_{N_4-Si_{12}}^*$	1% $\pi^*/99\%\sigma^*$	98.7%	A
2	398.4	0.0016	39.7% $\sigma_{N_4-Si_{12}}^*$ 2.1 % $\sigma_{N_4-C_8}^*$	2% $\pi^*/98\%\sigma^*$	93.3%	A
3	399.2	0.0010	12.3% $\sigma_{N_3-Si_{11}}^*$ 7.6 % $\pi_{C_7-C_8}^*$	15% $\pi^*/85\%\sigma^*$	92.8%	B
4	399.5	0.0010	11.0% $\pi_{C_7-C_8}^*$ 6.3 % $\sigma_{N_3-Si_{11}}^*$	22% $\pi^*/78\%\sigma^*$	92.2%	B
5	399.6	0.0005	9.9% $\sigma_{N_3-Si_{11}}^*$ 5.5 % $\pi_{C_7-C_8}^*$	12% $\pi^*/88\%\sigma^*$	89.9%	B
6	399.7	0.0003	5.6% $\sigma_{N_3-Si_{11}}^*$ 2.3 % $\pi_{C_7-C_8}^*$	5% $\pi^*/95\%\sigma^*$	89.9%	B
7	400.2	0.0000	N/A	100% σ^*	91.3%	
8	400.5	0.0000	N/A	100% σ^*	91.3%	
9	400.9	0.0003	5.2% $\pi_{C_7-C_8}^*$ 4.9 % $\pi_{C_1-C_2}^*$	12% $\pi^*/88\%\sigma^*$	87.8%	C
10	401.6	0.0000	44.2% $\pi_{C_1-C_2}^*$ 38.9 % $\pi_{C_7-C_8}^*$	87% $\pi^*/13\%\sigma^*$	95.7%	

state methods in order to target specific excitations. Our formulation of this method is general enough to be readily used in this manner.

Appendix A: Implementation of Maximum Subspace Occupation Method

The maximum subspace occupation method presented in this work is a technique that is general enough to be readily applied for other applications. We begin by specifying a subset of atomic orbitals S , a projection onto this subspace can be formally defined as:

$$\hat{\Gamma}_s = \sum_s [1]\theta_s[1]\theta_s \quad (\text{A1})$$

In order to evaluate the occupation of a given molecular orbital within this atomic orbital subspace we must evaluate the expectation value of this operator for the i^{th} MO.

$$\Omega_i = [1]\phi_i\hat{\Gamma}_s[1]\phi_i \quad (\text{A2})$$

$$= \sum_s [1]\phi_i\theta_s[1]\theta_s\phi_i \quad (\text{A3})$$

We can readily express the MOs of the KS determinant in the AO basis

$$[1]\phi_i = \sum_{\mu} C_{i\mu}[1]\chi_{\mu} \quad (\text{A4})$$

Using this definition of the MOs

$$\Omega_i = \sum_{\mu\nu} \sum_s [1]\chi_{\mu} C_{\mu i}^T [1]\theta_s [1]\theta_s C_{i\nu} [1]\chi_{\nu} \quad (\text{A5})$$

The subspace basis functions can be defined as an orthonormal transformation of the atomic orbital basis

$$[1]\theta_s = \sum_{\mu} X_{\mu s} [1]\chi_{\mu} \quad (\text{A6})$$

where X is the orthonormal transformation matrix. we can now plug this definition back into Equation A5 in order to obtain the following expression for the subspace occupation number.

$$\Omega_i = \sum_{\mu\nu} \sum_s [1]\chi_{\mu} C_{\mu i}^T X_{s\nu}^T [1]\chi_{\nu} [1]\chi_{\mu} X_{\mu s} C_{i\nu} [1]\chi_{\nu} \quad (\text{A7})$$

$$= \sum_{\mu\nu} \sum_s C_{\mu i}^T [1]\chi_{\mu} \chi_{\nu} X_{s\nu}^T X_{\mu s} [1]\chi_{\mu} \chi_{\nu} C_{i\nu} \quad (\text{A8})$$

$$= \mathbf{C}^T \mathbf{s} \mathbf{X} \mathbf{X}^T \mathbf{s} \mathbf{C} \quad (\text{A9})$$

Trivially, if we specify the subspace to include all atomic orbitals, then the subspace occupation number for every MO will be 1.

References

- [1] Tao, F.; Bernasek, S. L.; Xu, G.-Q. *Chem. Rev.* **2009**, *109*, 3991–4024.
- [2] Bent, S. F. *Surf. Sci.* **2002**, *500*, 879–903.
- [3] Filler, M. A.; Bent, S. F. *Prog. Surf. Sci.* **2003**, *73*, 1–56.
- [4] Wolkow, R. A. *Annu. Rev. Phys. Chem.* **1999**, *50*, 413–441.
- [5] Karthäuser, S. *J. Phys.: Condens. Matter* **2011**, *23*, 013001.
- [6] Mentovich, E.; Kalifa, I.; Tsukernik, A.; Caster, A.; Rosenberg-Shraga, N.; Marom, H.; Gozin, M.; Richter, S. *Small* **2008**, *4*, 55–58.
- [7] Joachim, C.; Gimzewski, J. K.; Aviram, A. *Nature* **2000**, *408*, 541–548.
- [8] Mayne, A. J.; Lastapis, M.; Baffou, G.; Soukiassian, L.; Comtet, G.; Hellner, L.; Dujardin, G. *Phys. Rev. B* **2004**, *69*, 045409.

- [9] Lu, X. *J. Am. Chem. Soc.* **2003**, *125*, 6384–6385.
- [10] Taguchi, Y.; Fujisawa, M.; Takaoka, T.; Okada, T.; Nishijima, M. *J. Chem. Phys.* **1991**, *95*, 6870–6876.
- [11] Sinniah, K.; Sherman, M. G.; Lewis, L. B.; Weinberg, W. H.; Yates, J. T.; Janda, K. C. *Phys. Rev. Lett.* **1989**, *62*, 567–570.
- [12] Zhao, J.; Noffke, B. W.; Raghavachari, K.; Teplyakov, A. V. *J. Phys. Chem. C* **2017**, *121*, 7208–7213.
- [13] Akagi, K.; Yoshinobu, J. *Surf. Sci.* **2016**, *652*, 304–311.
- [14] Konečný, R.; Doren, D. J. *Surf. Sci.* **1998**, *417*, 169–188.
- [15] Lu, X.; Xu, X.; Wu, J.; Wang, N.; Zhang, Q. *New J. Chem.* **2002**, *26*, 160–164.
- [16] Qu, Y.-Q.; Han, K.-L. *J. Phys. Chem. B* **2004**, *108*, 8305–8310.
- [17] Tao, F.; Wang, Z. H.; Xu, G. Q. *J. Phys. Chem. B* **2002**, *106*, 3557–3563.
- [18] Kong, M. J.; Teplyakov, A. V.; Lyubovitsky, J. G.; Bent, S. F. *Surf. Sci.* **1998**, *411*, 286–293.
- [19] Wang, G. T.; Mui, C.; Tannaci, J. F.; Filler, M. A.; Musgrave, C. B.; Bent, S. F. *J. Phys. Chem. B* **2003**, *107*, 4982–4996.
- [20] Romeo, M.; Balducci, G.; Stener, M.; Fronzoni, G. *J. Phys. Chem. C* **2014**, *118*, 1049–1061.
- [21] Tao, F.; Qiao, M. H.; Wang, Z. H.; Xu, G. Q. *J. Phys. Chem. B* **2003**, *107*, 6384–6390.
- [22] Miwa, J. A.; Eves, B. J.; Rosei, F.; Lopinski, G. P. *J. Phys. Chem. B* **2005**, *109*, 20055–20059.
- [23] Weier, D.; Lühr, T.; Beimborn, A.; Schönbohm, F.; Döring, S.; Berges, U.; Westphal, C. *Surf. Sci.* **2011**, *605*, 1784–1790.
- [24] Ardalan, P.; Dupont, G.; Musgrave, C. B. *J. Phys. Chem. C* **2011**, *115*, 7477–7486.
- [25] Chatterjee, A.; Zhang, L.; Leung, K. T. *Langmuir* **2013**, *29*, 9369–9377.
- [26] Jung, S. C.; Kang, M. H. *Phys. Rev. B* **2009**, *80*, 235312.
- [27] Huang, H. G.; Huang, J. Y.; Ning, Y. S.; Xu, G. Q. *J. Chem. Phys.* **2004**, *121*, 4820–4825.
- [28] Lee, H.-K.; Park, J.; Kim, I.; Kim, H.-D.; Park, B.-G.; Shin, H.-J.; Lee, I.-J.; Singh, A. P.; Thakur, A.; Kim, J.-Y. *J. Phys. Chem. C* **2012**, *116*, 722–725.
- [29] Lu, X.; Lin, M. *Int. Rev. Phys. Chem.* **2002**, *21*, 137–184.
- [30] Ng, W. K. H.; Sun, S. T.; Liu, J. W.; Liu, Z. F. *J. Phys. Chem. C* **2013**, *117*, 15749–15753.
- [31] Omiya, T.; Yokohara, H.; Shimomura, M. *J. Phys. Chem. C* **2012**, *116*, 9980–9984.
- [32] Shimomura, M. *Asian J. Chem.* **2013**, *25*, S35.
- [33] Stöhr, J. *NEXAFS Spectroscopy*; Springer: Berlin, 1992; Vol. 25.
- [34] Penner-Hahn, J. E. *Coord. Chem. Rev.* **1999**, *190–192*, 1101–1123.
- [35] Garino, C.; Borfecchia, E.; Gobetto, R.; van Bokhoven, J. A.; Lambert, C. *Coord. Chem. Rev.* **2014**, *277–278*, 130–186.
- [36] Tourillon, G.; Mahatsekake, C.; Andrieu, C.; Williams, G. P.; Garrett, R. F.; Braun, W. *Surf. Sci.* **1988**, *201*, 171–184.
- [37] Rosenberg, R. A.; Love, P. J.; Rehn, V. *Phys. Rev. B* **1986**, *33*, 4034–4037.
- [38] Shimoyama, I.; Wu, G.; Sekiguchi, T.; Baba, Y. *Phys. Rev. B* **2000**, *62*, R6053–R6056.
- [39] Stöhr, J.; Outka, D. A. *Phys. Rev. B* **1987**, *36*, 7891–7905.
- [40] This L stands for Langmuir, it is a standard unit to quantify exposure of a surface to a gaseous sample. The quantity is obtained by multiplying the pressure of the gas by the time of exposure.
- [41] Triguero, L.; Pettersson, L. G. M.; Ågren, H. *Phys. Rev. B* **1998**, *58*, 8097–8110.
- [42] Besley, N. A.; Noble, A. *J. Phys. Chem. C* **2007**, *111*, 3333–3340.
- [43] Fronzoni, G.; Balducci, G.; De Francesco, R.; Romeo, M.; Stener, M. *J. Phys. Chem. C* **2012**, *116*, 18910–18919.
- [44] De Francesco, R.; Stener, M.; Fronzoni, G. *J. Phys. Chem. C* **2007**, *111*, 13554–13563.
- [45] Coustel, R.; Carniato, S.; Rochet, F.; Witkowski, N. *Phys. Rev. B* **2012**, *85*, 035323.
- [46] Rangan, S.; Gallet, J.-J.; Bournel, F.; Kubsky, S.; Le Guen, K.; Dufour, G.; Rochet, F.; Sirotti, F.; Carniato, S.; Ilakovac, V. *Phys. Rev. B* **2005**, *71*, 165318.
- [47] Rangan, S.; Bournel, F.; Gallet, J.-J.; Kubsky, S.; Le Guen, K.; Dufour, G.; Rochet, F.; Sirotti, F.; Carniato, S.; Ilakovac, V. *Phys. Rev. B* **2005**, *71*, 165319.
- [48] Fransson, T.; Coriani, S.; Christiansen, O.; Norman, P. *J. Chem. Phys.* **2013**, *138*, 124311.
- [49] Besley, N. A. *Chem. Phys. Lett.* **2012**, *542*, 42–46.
- [50] Coriani, S.; Christiansen, O.; Fransson, T.; Norman, P. *Phys. Rev. A* **2012**, *85*, 022507.

- [51] Coriani, S.; Fransson, T.; Christiansen, O.; Norman, P. *J. Chem. Theory Comput.* **2012**, *8*, 1616–1628.
- [52] Myhre, R. H.; Coriani, S.; Koch, H. *J. Chem. Theory Comput.* **2016**, *12*, 2633–2643.
- [53] Maganas, D.; Kristiansen, P.; Duda, L.-C.; Knop-Gericke, A.; DeBeer, S.; Schlögl, R.; Neese, F. *J. Phys. Chem. C* **2014**, *118*, 20163–20175.
- [54] Maganas, D.; Roemelt, M.; Weyhermuller, T.; Blume, R.; Havecker, M.; Knop-Gericke, A.; DeBeer, S.; Schlogl, R.; Neese, F. *Phys. Chem. Chem. Phys.* **2014**, *16*, 264–276.
- [55] Grimme, S. *Chem. Phys. Lett.* **1996**, *259*, 128–137.
- [56] Wenzel, J.; Wormit, M.; Dreuw, A. *J. Chem. Theory Comput.* **2014**, *10*, 4583–4598.
- [57] Wenzel, J.; Holzer, A.; Wormit, M.; Dreuw, A. *J. Chem. Phys.* **2015**, *142*, 214104.
- [58] Wenzel, J.; Dreuw, A. *J. Chem. Theory Comput.* **2016**, *12*, 1314–1330.
- [59] Imamura, Y.; Nakai, H. *Chem. Phys. Lett.* **2006**, *419*, 297–303.
- [60] Lestranger, P. J.; Nguyen, P. D.; Li, X. *J. Chem. Theory Comput.* **2015**, *11*, 2994–2999.
- [61] Asmuruf, F.; Besley, N. *Chem. Phys. Lett.* **2008**, *463*, 267–271.
- [62] Lopata, K.; Van Kuiken, B. E.; Khalil, M.; Govind, N. *J. Chem. Theory Comput.* **2012**, *8*, 3284–3292.
- [63] Evangelista, F. A.; Shushkov, P.; Tully, J. C. *J. Phys. Chem. A* **2013**, *117*, 7378–7392.
- [64] Derricotte, W. D.; Evangelista, F. A. *Phys. Chem. Chem. Phys.* **2015**, 14360–14374.
- [65] Verma, P.; Derricotte, W. D.; Evangelista, F. A. *J. Chem. Theory Comput.* **2016**, *12*, 144–156.
- [66] Stener, M.; Fronzoni, G.; de Simone, M. *Chem. Phys. Lett.* **2003**, *373*, 115–123.
- [67] Besley, N. A.; Asmuruf, F. A. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12024.
- [68] Peng, B.; Lestranger, P. J.; Goings, J. J.; Caricato, M.; Li, X. *J. Chem. Theory Comput.* **2015**, *11*, 4146–4153.
- [69] Derricotte, W. D.; Evangelista, F. A. *Submitted to J. Chem. Theory Comput.* **2017**,
- [70] Ayers, P. W.; Levy, M.; Nagy, A. *Phys. Rev. A* **2012**, *85*, 042518.
- [71] Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *120*, 2629–2637.
- [72] Knizia, G. *J. Chem. Theory Comput.* **2013**, *9*, 4834–4843.
- [73] Schmidt, M. W.; Hull, E. A.; Windus, T. L. *J. Phys. Chem. A* **2015**, *119*, 10408–10427.
- [74] Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916–4926.
- [75] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- [76] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- [77] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- [78] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- [79] Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys.* **1998**, *109*, 1223–1229.
- [80] Hariharan, P. C.; Pople, J. A. *Theoret. Chim. Acta* **1973**, *28*, 213–222.
- [81] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- [82] Turney, J. M. et al. *WIREs Comput Mol Sci* **2012**, *2*, 556–565.
- [83] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [84] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- [85] Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- [86] Ma, Y.; Sette, F.; Meigs, G.; Modesti, S.; Chen, C. T. *Phys. Rev. Lett.* **1989**, *63*, 2044–2047.
- [87] Carravetta, V.; Ågren, H. *J. Phys. Chem. A* **2013**, *117*, 6798–6802.
- [88] Vall-Iloera, G.; Gao, B.; Kivimäki, A.; Coreno, M.; Álvarez Ruiz, J.; de Simone, M.; Ågren, H.; Rachlew, E. *J. Chem. Phys.* **2008**, *128*, 044316.