# Spatial Probability Distribution of Adsorbate Atoms

Ph. Hofmann,* O. Schaff, and K.-M. Schindler
*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany*
(Received 17 August 1995)

In order to access the dynamics of surface diffusion or of surface chemical reactions it is necessary to determine the spatial probability distributions of the adsorbed atoms or molecules rather than their average positions. Here we demonstrate how such a distribution can be determined from scanned-energy mode photoelectron diffraction data using the maximum entropy method. As a test case we have chosen the distribution of the carbon atom in the methoxy species adsorbed on Ni(111). Both simulated and experimental data are used.

PACS numbers: 61.14.–x, 68.35.Ja, 79.60.Dp, 82.65.My

Surface crystallographic techniques such as low energy electron diffraction (LEED), surface extended x-ray absorption fine structure (SEXAFS), or photoelectron diffraction usually determine the average positions of atoms or molecules adsorbed on a surface. The analysis procedures normally assume that only low amplitude (harmonic) vibrations occur. Clearly this rather simplified view is inadequate when the dynamics of surface diffusion and of surface reactions have to be addressed. Pendry and Heinz have therefore suggested that the concept of surface structure determination should be extended to include the spatial probability distributions $P(\mathbf{r})$ of the adsorbed atoms or molecules rather than just their average position [1,2]. In this Letter we demonstrate how such a distribution can be determined from scanned-energy mode photoelectron diffraction data using the maximum entropy method.

In scanned-energy mode photoelectron diffraction (PHD) the photoemission intensity from a core level of an atom within the adsorbate is measured in a defined direction as a function of photon, and thus photoelectron, energy. Modulations in this intensity of up to 40% can be observed due to the interference of the directly emitted part of the photoelectron wave field and the parts scattered at the substrate atoms. From the intensity $I(E)$ the so-called modulation function $\chi(E)$ is determined according to

$$\chi(E) = \frac{I(E) - I_0(E)}{I_0(E)}, \tag{1}$$

where $I_0(E)$ is the slowly varying part of $I(E)$ containing, for example, the energy dependence of the photoemission cross section and the monochromator flux. The modulation function $\chi(E)$ provides information about the local bonding geometry around the emitting atom [3]. In order to enlarge the data set modulation functions are normally determined for several different emission directions. In the remainder of this paper we use the symbols $I(\mathbf{k})$ and $\chi(\mathbf{k})$ for a set of $I(E)$ and $\chi(E)$ distributions taken in different emission directions. The structure is usually determined by comparing the $\chi(\mathbf{k})$ with multiple-scattering spherical-wave calculations for a model system which is then modified until good agreement is achieved (quantified in terms of a suitably defined $R$ factor) [4]. The most important parameter is the position of the emitter atom on which all scattering pathways depend. PHD is therefore a suitable technique for determining the spatial probability distribution $P(\mathbf{r})$ of the emitter.

The core level intensity measured at the detector is the (incoherent) sum over very many emitter atoms which are distributed at any single time according to the desired $P(\mathbf{r})$. In principle, $P(\mathbf{r})$ could be extracted from the experimental data by calculating the intensity $I_{\mathrm{th}}(\mathbf{r}, \mathbf{k})$ for the emitter atom at any position $\mathbf{r}$ above the surface, integrating over these intensities weighted with a test distribution $P_{\mathrm{trial}}(\mathbf{r})$, according to

$$I_{\mathrm{th}}(\mathbf{k}) = \int P_{\mathrm{trial}}(\mathbf{r}) I_{\mathrm{th}}(\mathbf{r}, \mathbf{k}) d^3\mathbf{r}, \tag{2}$$

normalizing the resulting intensity to give a theoretical modulation function $\chi_{\mathrm{th}}(\mathbf{k})$ and comparing it to the experimental data. $P_{\mathrm{trial}}(\mathbf{r})$ could then be varied in an iterative fashion until the best agreement between theory and experiment is achieved. An important prerequisite would be that the local substrate geometry around the emitter is precisely known, so that only the influence of $P_{\mathrm{trial}}(\mathbf{r})$ on the level of agreement between experiment and theory has to be considered.

The left-hand side of Fig. 1 shows the adsorption geometry of the methoxy species on Ni(111). Both the adsorption site and the local relaxations of the surface atoms were determined by PHD data taken from the oxygen 1$s$ core level [5]: The molecular fragment is adsorbed with the oxygen atom in an "fcc" threefold symmetric hollow site 1.16(5) Å above the first Ni layer. Furthermore, the vibrational mean square deviation from the equilibrium position of the oxygen atom both parallel and perpendicular to the surface have been determined assuming a Gaussian probability distribution [6]. The vibrational amplitude is, as expected for a high coordination adsorption site, small and isotropic (the mean square deviation from equilibrium position is $3 \times 10^{-3}$ Å$^2$). More details on the relaxation of the underlying nickel atoms can be

 0031-9007/96/76(6)/948(4)$06.00 

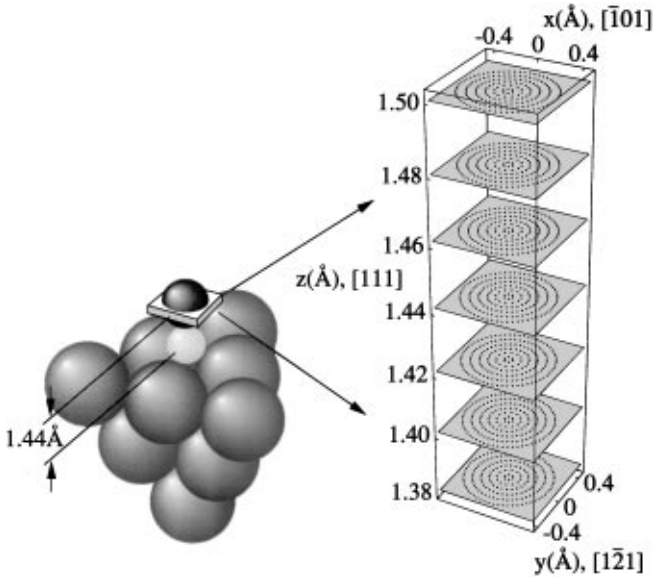


FIG. 1. The geometric structure of the surface methoxy species on Ni(111) after Ref. [5]; substrate relaxations are not included. On the right-hand side the positions are shown at which C 1*s* PHD simulations were performed. The diagram is expanded by a factor of 40 in the *z* direction with respect to the left-hand diagram. The *z* axis gives the height of the carbon atom above the oxygen atom.

found in Ref. [5]. The carbon atom represents an ideal candidate for determining its spatial probability distribution since the only parameter which has to be determined by the C 1*s* PHD data is the carbon atom position; all other parameters have already been fixed by the oxygen data. Using a data set of seven C 1*s* modulation functions measured in different emission directions Schaff *et al.* have found the carbon-oxygen bond length to be 1.44(5) Å with the O-C axis perpendicular to the surface resulting in an *R* factor of 0.15 [5]. The vibration of the carbon atom is highly anisotropic, which is presumably due to its lower coordination. Assuming a Gaussian distribution the mean-square deviations from the equilibrium position for vibrations perpendicular and parallel to the surface were found to be $2 \times 10^{-3}$ and $1 \times 10^{-2}$ Å$^2$, respectively. The combined *R* factor of 0.15 for seven different emission directions implies remarkably good agreement between theory and experiment, which in turn stresses the reliability of the structure determination. Therefore, it is justified to restrict the search to the space in the vicinity of the carbon atom position found in Ref. [5]. The continuous spatial probability distribution $P(\mathbf{r})$ is represented by discrete $P_i$ according to $P_i = P(\mathbf{r}_i)$. A suitable grid for the $\mathbf{r}_i$'s turns out to be of cylindrical shape with seven planes parallel to the surface, each with 186 positions. This grid is shown on the right-hand side of Fig. 1 and has been expanded in the *z* direction by a factor of 40 in order to make all points visible. The separation of the grid points has been chosen such that the modulation functions do not change appreciably on going from one point to the next. Note that the technique is less sensitive to a motion of the adsorbate parallel to the surface [4], hence the lower point density. For all of these 1302 positions the C 1*s* photoemission intensities were calculated and combined according to the discrete form of Eq. (2):

$$I_{\mathrm{th}}(\mathbf{k}) = \sum_i P_i I_{\mathrm{th},i}(\mathbf{k}) \,. \qquad (2')$$

Taking into account the $C_{3v}$ symmetry of the adsorption complex and the $C_s$ symmetry of the experiment leads to $1302/(3/2) = 217$ parameters $P_i$ for fitting this $I_{\mathrm{th}}(\mathbf{k})$ to the photoelectron diffraction data. However, this still represents an ill-posed problem, so that simple least squares fitting leads to unreasonable, discontinuous spatial probability distributions.

We have therefore adopted a probabilistic approach which allows us to incorporate prior knowledge about the form of the solution [7]. We seek the most likely distribution $\{P_i\}$ given the experimental data and all our prior knowledge *K* about the solution. From Bayes' theorem it can be derived that

$$\mathrm{prob}[\{P_i\}\,|\,\chi(\mathbf{k}), K] \sim \mathrm{prob}[\chi(\mathbf{k})\,|\,\{P_i\}, K] \times \mathrm{prob}[\{P_i\}\,|\,K] \,. \qquad (3)$$

The term on the left is the probability of the distribution $\{P_i\}$ given the experimental modulation function(s) $\chi(\mathbf{k})$ and our prior knowledge about the solution. Our aim is to find $\{P_i\}$ such that it maximizes this *posterior* probability. In the absence of any experimental data the posterior probability would be equal to the *prior* probability $\mathrm{prob}[\{P_i\}\,|\,K]$. The experiment modifies this prior state of knowledge by the so-called *likelihood function* $\mathrm{prob}[\chi(\mathbf{k})\,|\,\{P_i\}, K]$, i.e., the probability of measuring $\chi(\mathbf{k})$ if $\{P_i\}$ was the true probability distribution of the carbon atom. Assuming independent data points, which are subject to additive Gaussian noise only, the likelihood function is proportional to $\exp(-\chi^2/2)$ where $\chi^2$ is the usual sum of squared misfit residuals [8]. In PHD the agreement between theory and experiment is usually not expressed as $\chi^2$ but in the form of an *R* factor. In the vicinity of good agreement, however, the two are proportional [4]. The likelihood function can therefore be expressed using the *R* factor

$$\mathrm{prob}[\chi(\mathbf{k})\,|\,\{P_i\}, K] \sim \exp(-R) \,. \qquad (4)$$

Since the spatial probability distribution of the carbon atom is positive and additive, a reasonable assumption for the prior probability is [7,9,10]

$$\mathrm{prob}[\{P_i\}\,|\,K] \sim \exp(\alpha S) \,, \qquad (5)$$

with the scaling parameter $\alpha$ and the generalized Shannon Jaynes information theory entropy $S$ for $i = 1, \ldots, N$ cells:

$$S = \sum_i P_i - M_i - P_i \ln(P_i/M_i), \tag{6}$$

where $\{M_i\}$ is the model for which $S$ becomes zero, i.e., attains its maximum value. In the absence of experimental data $P_i = M_i$ maximizes the prior probability and thus the posterior probability. Hence, $M_i$ is often named default model.

The posterior probability becomes proportional to

$$\exp(-R + \alpha S). \tag{7}$$

The task of maximizing the posterior probability is now reduced to minimizing $R - \alpha S$. Clearly the choice of $\alpha$ determines the relative importance of the prior knowledge and the experimental data. There are several concepts to determine $\alpha$, such as historic and classic maximum entropy. However, several of the underlying assumptions may not be valid in our case, such as a well-defined $\chi^2$, independence of neighboring data points, or exact fit of the experiment by the simulation. Therefore, we have adopted a concept where a large starting value of $\alpha$ is chosen, the posterior probability is maximized, $\alpha$ is decreased, and the procedure repeated for a large range of $\alpha$. This enables us to observe the behavior of $R$, $S$, and the spatial probability distribution as a function of $\alpha$. We have tested two different starting models $\{M_i\}$ for calculating $\{P_i\}$. The first was the anisotropic Gaussian distribution of the carbon atom giving the best $R$ factor in Ref. [5], the second a uniform distribution over all positions in the cylinder. The choice of the starting model had only minor influence on the resulting $\{P_i\}$ so that in all examples given here the uniform distribution was taken as starting model.

We have calculated the probability distribution of the carbon atom for three sets of C 1$s$ PHD data each with seven modulation functions taken in normal emission in the $\langle 110 \rangle$ azimuth as well as in 10° and 20° off-normal emission in the $\langle 110 \rangle$ and both $\langle 112 \rangle$ azimuths. Two of the data sets are simulated and superimposed with artificial noise; the other consists of the measured modulation functions used also in Ref. [5]. The stimulated data sets represent a carbon atom fixed at 1.44 Å above the oxygen atom without any vibrations and a carbon atom at the same height with a Gaussian probability distribution and mean square deviations from the equilibrium position of $5 \times 10^{-4}$ and $2 \times 10^{-2}$ Å$^2$ perpendicular and parallel to the surface, respectively.

Figure 2 shows the $R$ factor and $-S$ as a function of $\alpha$ for the measured data set as input where $R - \alpha S$ is minimized for each $\alpha$. The calculated input data sets show the same type of behavior. For large values of $\alpha$ the entropy is zero and the $\{P_i\}$ is equal to the starting model

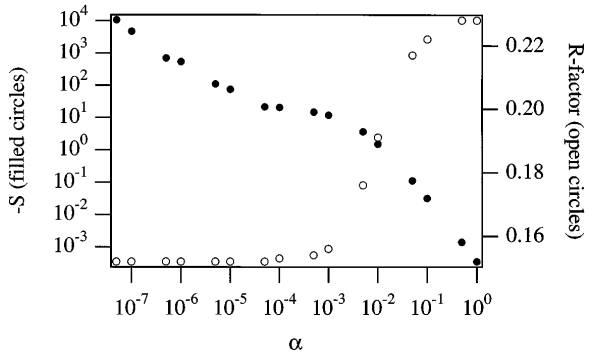


FIG. 2. The $R$ factor (open circles) and $-S$ (filled circles) as a function of $\alpha$, where $R - \alpha S$ is minimized for each $\alpha$ taking the experimental data as input.

$\{M_i\}$. With decreasing $\alpha$, $R$ also decreases at first and then remains constant. Three regions of behavior for $S$ can be distinguished: first a strong decrease, then a region where it is almost constant, and finally another strong decrease. The corresponding distributions $\{P_i\}$ are very similar to $\{M_i\}$ in the first region and more and more discontinuous and then unphysical in the third. For the $\alpha$ values where the entropy is almost constant the distributions are very similar to each other and smooth (with the exception of the point-emitter case, of course). Here replacing a smaller $\alpha$ by a larger $\alpha$ results only in a smearing out of the details of the distribution but does not change the general shape. In historic maximum entropy the stop criterion for the decrease of $\alpha$ is that $\chi^2$ drops below the number of data. Here we do not have a clear criterion concerning which value of $\alpha$ to choose since it is not clear at which $R$ factor we may regard the resulting $\{P_i\}$ as "true." However, it is reasonable to assume that the true solution lies in the range of $\alpha$ where the entropy is constant and the $R$ factor close to its minimum value because a further reduction of $\alpha$ results only in a smaller $S$ due to "overfitting" but not necessarily in a lower $R$ factor.

Figure 3 shows the resulting probability distributions for all three data sets with $\alpha$ values such that $S$ is in the constant region. The two distributions determined from the simulated data show the expected form: The first corresponds to a point emitter 1.44 Å above the oxygen atom, the other to an anisotropic Gaussian distribution with approximately the same parameters as in the input data. $\{P_i\}$ for the experimental data shows an umbrellalike shape with a small azimuthal anisotropy indicating that the residence probability of the carbon atom is higher towards the bridge sites than towards the atop sites. Note that the $R$ factor for the final distribution is only marginally smaller than the value found by Schaff *et al.* [5]. It should be remembered, however, that we are seeking the maximum of the posterior probability rather than the smallest possible $R$ factor. Indeed the spatial probability distribution is exactly that which might be expected. It should reflect the superposition of all the vibrational modes of the molecule on the surface. If we neglect the hydrogen atoms, the

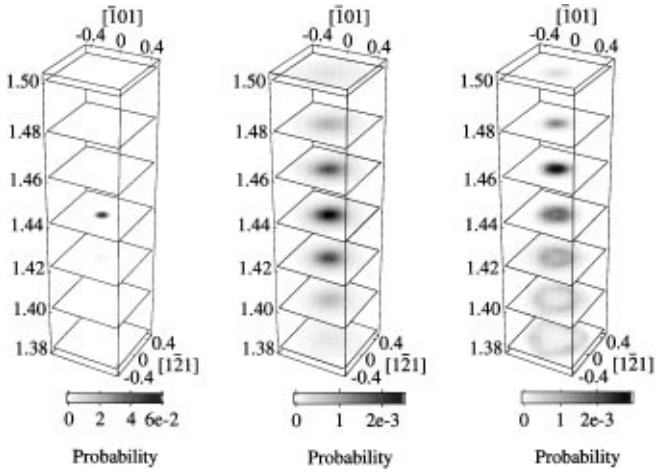


FIG. 3. Final probability distributions interpolated on a fine mesh for the three systems: left-hand side: simulated data with superimposed noise for a point emitter 1.44 Å above the oxygen atom; center: simulated data with superimposed noise for an emitter with an anisotropic Gaussian distribution; right-hand side: experimental data for the methoxy species on Ni(111).

most important high amplitude contributions will come from the hindered rotations and translations [11]. The probability distribution of the carbon atom should be particularly strongly influenced by the hindered rotation, i.e., the dynamic tilt of the C-O axis, upon which the much higher frequency C-O stretch vibration is superimposed. Furthermore, the direction of the observed anisotropy is also consistent with intuition. A higher potential barrier for a tilt of the C-O axis is expected towards the atop sites than towards the bridge sites. Note that the contours of highest probability in each plane do not follow the half sphere centered on the oxygen atom which might be expected from a rigid O-C bond. This is probably due to the vibration of the oxygen atom itself. While its in-plane motion is strongly confined by the threefold coordination, the amplitude of the oxygen-metal stretch vibration might still be considerable. Together with the bending mode it could lead to the deviation from the spherical shape.

In summary, we have demonstrated that it is possible to determine the spatial probability distribution of adsorbed atoms from PHD data using the maximum entropy method. Because of the approximation (4) in our approach, no error bars for the reconstruction can be given at present. An extension to classic maximum entropy where the parameter $\alpha$ is also determined in a probabilistic fashion from the data would require a rigorous statistical treatment of the experimental data avoiding the normalization (1). Furthermore, the theory would have to describe perfectly the measurements, a requirement which is not fulfilled in electron scattering experiments such as PHD or, indeed, LEED. However, the application of these concepts considerably broadens the scope of surface structure analysis. It should even be possible to include small surface relaxations which might be treated within the linear PHD framework [12]. While the computational effort at present strongly limits the size of the grid, future calculation could determine the probability distribution in a whole surface unit cell. This would open the possibility to study surface diffusion: The tail of the distribution into equivalent adsorption sites gives a measure of the probability that the adsorbate will migrate into these sites. A direct access to chemical reactions might not be possible because of the long data collection time. However, observing the change in the probability distribution of the reactant(s) as the temperature is increased might still give a hint as to the reaction mechanism.

The authors are pleased to acknowledge stimulating discussions with A. M. Bradshaw, V. Fritzsche, D. P. Woodruff, W. von der Linden, and V. Dose. We also thank V. Fritzsche for the possibility of using his simulation program and C. Hirschmugl for proofreading. K. M. S. thanks the Deutsche Forschungsgemeinschaft for a Habilitationsstipendium. This work has been supported by the German Federal Minister for Education, Science, Research and Technology (BMBF) under Contract No. 05 625EBA 6 and the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich No. 290.

*Present address: Department of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996-1200 and Solid State Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831-6057.

[1] J. B. Pendry and K. Heinz, Surf. Sci. **230**, 137 (1990).
[2] J. B. Pendry, Catal. Lett. **9**, 189 (1991).
[3] D. P. Woodruff and A. M. Bradshaw, Rep. Prog. Phys. **57**, 1029 (1994).
[4] Ph. Hofmann, K.-M. Schindler, S. Bao, V. Fritzsche, D. E. Ricken, A. M. Bradshaw, and D. P. Woodruff, Surf. Sci. **304**, 74 (1994).
[5] O. Schaff, G. Hess, V. Fritzsche, V. Fernandez, K.-M. Schindler, A. Theobald, Ph. Hofmann, A. M. Bradshaw, R. Davis, and D. P. Woodruff, Surf. Sci. **331–333**, 201 (1995).
[6] V. Fritzsche, K.-M. Schindler, P. Gardner, A. M. Bradshaw, M. C. Asensio, and D. P. Woodruff, Surf. Sci. **269/270**, 35 (1992).
[7] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988).
[8] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering, *Numerical Recipes* (Cambridge University Press, Cambridge, 1989).
[9] J. E. Gubernatis, M. Jarrell, R. N. Silver, and D. S. Sivia, Phys. Rev. B **44**, 6011 (1991).
[10] W. von der Linden, M. Donath, and V. Dose, Phys. Rev. Lett. **71**, 899 (1993).
[11] N. V. Richardson and A. M. Bradshaw, Surf. Sci. **88**, 255 (1979).
[12] V. Fritzsche and J. B. Pendry, Phys. Rev. B **48**, 9054 (1993).