# Bayesian approach to the calculation of lateral interactions: NO/Rh(111)

*Citation for published version (APA):*
Jansen, A. P. J., & Popa, C. (2008). Bayesian approach to the calculation of lateral interactions: NO/Rh(111). *Physical Review B*, *78*(8), 085404-1/9. Article 085404. https://doi.org/10.1103/PhysRevB.78.085404

*DOI:*
10.1103/PhysRevB.78.085404

*Document status and date:*
Published: 01/01/2008

*Document Version:*
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

*Please check the document version of this publication:*

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

*General rights*
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

*Take down policy*
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

PHYSICAL REVIEW B **78**, 085404 (2008)

# Bayesian approach to the calculation of lateral interactions: NO/Rh(111)

A. P. J. Jansen and C. Popa
*Laboratory of Inorganic Chemistry and Catalysis, ST/SKA, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

(Received 9 April 2008; published 5 August 2008)

We show how Bayesian statistics and density-functional theory can be combined to compute reliable values for the interactions in a cluster expansion for adsorbates on a surface. The method is an alternative to the leave-one-out cross-validation method. We show that it easily selects which interactions can be determined even if the total number of possible interactions is very large. We have applied the method to NO/Rh(111). Based on the interactions we have determined for this system we have predicted some structures, which have been confirmed by scanning-tunneling microscopy.

DOI: 10.1103/PhysRevB.78.085404 PACS number(s): 68.43.Bc, 02.50.Tt, 34.35.+a

## I. INTRODUCTION

Interactions between adsorbates, or lateral interactions, on transition-metal surfaces have at least been known for as long as diffraction techniques have revealed that adlayers can form very well-defined structures at low temperatures. The importance of these interactions for kinetics at higher temperatures has only more recently been acknowledged but forms now an active area of research. This is understandable if one realizes that even small interactions between adsorbates can be of the same magnitude or larger than the thermal energy and can therefore change rate constants by an order of magnitude or more, especially at low temperatures.

There are various experimental techniques one can use to determine lateral interactions. Using scanning-tunneling microscopy one can determine the statistical distribution of the adsorbates over the surface. This distribution can be converted into a radial distribution function, which in turn can be converted into effective lateral interactions. One has to be able to image the exact position of each adsorbate to obtain the radial distribution function. This requires a low adsorbate mobility, which means that it is mainly applied to atoms on low-temperature surfaces.[1–3]

Low energy electron diffraction allows one to study the ordered phases of an adsorbate. One usually compares the different ordered phases found for an adsorbate and the temperature ranges in which they appear. This experimental results can then be fitted using a lattice-gas model, thus yielding values for the lateral interactions.[4–6] Several ordered phases are needed for this method to be of practical use.

Accurate estimates of (differences in) binding energies can be obtained by analyzing temperature programmed desorption traces.[7–10] The differences in desorption temperatures can be directly related to differences in binding energies. It is, however, in general difficult to relate these differences in binding energies to lateral interactions, since the local adsorbate configuration (the number of adsorbates interacting with the desorbing molecule) is unknown. One option is to fit desorption spectra using kinetic Monte Carlo simulations.[11]

In single-crystal adsorption calorimetry binding energies of adsorbates can be directly measured. The differences in binding energies due to lateral interactions can therefore also be accurately determined. The configuration of the adsorbate adlayer is unknown, however, and relating these differences in binding energies to lateral interactions is therefore—just as for temperature programmed desorption—difficult. If the local adsorbate configuration is known or can be guessed using another technique, then lateral interactions can be extracted.[12–15]

With computer hardware becoming faster it is possible to do quantum chemical calculations on quite realistic models of adsorbates on transition-metal surfaces especially using density-functional theory (DFT). Such calculations have the advantage over experiments for the determination of lateral interactions that one knows precisely the system one is dealing with, and one does not have to worry about possible effects of steps, impurities, shortcomings of apparatuses, etc.

When calculating lateral interactions they are most commonly described with the cluster expansion. The adsorption energy $E_{\text{ads}}$ of an adsorbate in a particular adlayer structure is then written as

$$E_{\text{ads}} = \sum_m c_m V_m, \tag{1}$$

with $V_m$ as the value of the interaction of type $m$ and $c_m$ as the number of interactions of type $m$ per adsorbate. The interactions $V_m$ stand for the interaction of the adsorbate with the substrate (adsorption energy of an isolated adsorbate), pair interactions between adsorbates at various distances, all possible three-particle interactions, four-particle interactions, etc. This expansion can be made to reproduce the calculated adsorption energy as accurately as one wants.[16] This generally takes however a large number of terms. Moreover, it may lead to overfitting; i.e., the cluster expansion will not only describe the interactions but also the errors one makes in the calculations of the adsorption energies. To avoid this one needs to truncate the cluster expansion.

The truncation of the cluster expansion for lateral interactions between adsorbates has so far mainly been done based on the desired accuracy with which the truncated expansion reproduces the calculated results, the number of acceptable terms in the expansion, the type of terms in the expansion (pair, three particle, etc.), the estimated accuracy of the calculated results, and possibly other factors. Often these factors involve a trade-off; e.g., one prefers a short expansion, but it should also be accurate. Researchers have usually dealt with this using their personal experience and insight, but few objective criteria have been used.

A similar problem was encountered for the calculation between atoms forming an alloy. This has led to the development of the leave-one-out cross-validation (LOO-CV) method.[17,18] This is a statistical technique that uses part of the results of a set of calculations to determine values for the interactions between atoms and the rest to test these values. Because determination and testing of the interactions is done on independent results of calculations, one obtains an estimate of how well the values for interactions one calculates will predict energies of unknown structures. This method has recently also been applied to the determination of lateral interactions.[19–21]

In the LOO-CV method one starts with a model with few interaction parameters and determines the CV score. This is a measure for how much a prediction of the energy of an adlayers based on the model will differ from the energy obtained from a calculation. Adding parameters will initially lower the CV score, which means that the model becomes better. Adding too many parameters should however increase the CV score because of overfitting. The minimum of the CV score indicates a best set of interaction parameters. The problem with the method is that often the CV score becomes almost constant when more parameters are added, and it is very hard to determine the minimum of the CV score.[22–24]

This paper presents an alternative to the LOO-CV method. It is based on Bayesian statistics. Instead of the CV score, we assign a probability to each model of lateral interactions or set of terms of the cluster expansion. The best model is the one with the highest probability. We will show that this approach does not have the drawback of the LOO-CV method. The model with the maximum probability is well defined. Moreover, it seems that the approach leads to model with fewer interaction parameters. The method also lends itself well to an analysis of the importance of parameters; it is easy to compute probabilities for individual parameters and correlation between parameters. We will show as an illustrative example the interactions between NO molecules in adlayers on Rh(111).

## II. THEORY

We assume that we have done calculations on various adlayer structures with the same substrate in all calculations and only one type of adsorbate in all adlayer structures. The calculations have resulted in the adsorption energy per adsorbate $E_{\text{ads}}^{\text{calc.}}(n)$ with $n$ as an index to distinguish the adlayer structures. We want to describe the energy $E_{\text{ads}}^{\text{calc.}}(n)$ using a cluster expansion for the lateral interactions. This means we write

$$E_{\text{ads}}^{\text{fit}}(n) = \sum_{m} c_m(n) V_m, \tag{2}$$

with $V_m$ as the value of the interaction of type $m$ and $c_m(n)$ as the number of interactions of type $m$ per adsorbate in structure $n$. The interactions $V_m$ stand for the interaction of the adsorbate with the substrate (adsorption energy of an isolated adsorbate), pair interactions between adsorbates at various distances, all possible three-particle interactions, four-particle interactions, etc. The expression $E_{\text{ads}}^{\text{fit}}(n)$ should approximate $E_{\text{ads}}^{\text{calc.}}(n)$. Note that we treat the adsorption energy of an isolated adsorbate in the same way as the lateral interactions. It is possible to single out adlayer structures with low coverage and therefore no lateral interactions to determine the adsorption energy of isolated molecules separately. The errors that are made in such determination of adsorption energies will however skew the subsequent fit of the lateral interactions. We have shown that this increases the error in the lateral interactions by a factor of $\sqrt{2}$.[20] One also misses the advantage that systematic errors in the $E_{\text{ads}}^{\text{calc.}}(n)$'s do not affect the lateral interactions as will be shown below.

The summation in the expression for $E_{\text{ads}}^{\text{fit}}(n)$ runs in principle over an infinite number of interactions. The question is which of the interactions can be determined from the adsorption energies $E_{\text{ads}}^{\text{calc.}}(n)$. This is the important difference with a straightforward multivariate linear regression. Each subset of all interactions forms a model for the lateral interactions in the system. We want to know which interaction model describes the calculated adsorption energies best. We use $S$ for a subset of all interaction parameters $V_m$. We use $V$ as a shorthand for the set of all values of the interaction parameters in $S$. For all calculated adsorption energies we use $E$. We will determine which model of the interactions is best by calculating $P(S|E)$, which stands for the probability that the calculated adsorption energies $E$ can be described by interaction parameters in $S$.

We can use Bayes's theorem to relate the probability of $S$ given $E$ [i.e., $P(S|E)$] to the probability of $E$ given $S$ [i.e., $P(E|S)$].[25–27]

$$P(S|E) \propto P(E|S)P(S). \tag{3}$$

The proportionality constant that is missing in this expression can be determined by normalizing $P(S|E)$ as a function of $S$. The probability $P(E|S)$ is often called the likelihood of $S$ given $E$, the probability $P(S)$ is called the prior (probability) of $S$, and the probability $P(S|E)$ is called the posterior (probability).

Bayes's theorem is used as follows in the selection of models for the lateral interactions. We want to calculate $P(S|E)$. How good a model is will also depend on whether we can find good values for the lateral interactions. It is important to distinguish between $S$ (the parameters in the model) and $V$ (the values of these parameters). We can introduce the values by regarding $P(E|S)$ as a marginal distribution of $P(E,V|S)$ via[27]

$$P(E|S) = \int dV P(E,V|S). \tag{4}$$

The integrand can be written as

$$P(E,V|S) = P(E|S,V)P(V|S). \tag{5}$$

Substitution in the Bayes's expression for $P(S|E)$ then gives

$$P(S|E) \propto P(S)\int dVP(E|S,V)P(V|S). \tag{6}$$

This allows us to compute $P(S|E)$ because we can make a good guess of what the calculated adsorption energies should be given the lateral interactions [i.e., $P(E|S,V)$], and it should be possible to think of reasonable priors $P(V|S)$ and $P(S)$.

Suppose we have a set of interaction parameters $S$ with values $V$. A normal way to obtain such a set is via a least-squares procedure to fit $E_{\text{ads}}^{\text{fit}}(n)$ to $E_{\text{ads}}^{\text{calc.}}(n)$. Suppose that the set has all interaction parameters to describe the system and that they have the correct values. The most likely values for $E_{\text{ads}}^{\text{calc.}}(n)$ should then be equal to $E_{\text{ads}}^{\text{fit}}(n)$. Due to errors in the calculations the calculated adsorption energies will not be exactly equal. Instead it seems reasonable to assume that the difference can be described by a Gaussian probability distribution; i.e.,

$$P(E|S,V) = \exp\left[-\frac{1}{2}\chi_1^2\right]\prod_{n=1}^{N_{\text{str}}}\frac{1}{\sqrt{2\pi\sigma_n^2}}, \tag{7}$$

with

$$\chi_1^2 \equiv \sum_{n=1}^{N_{\text{str}}}\left[\frac{E_{\text{ads}}^{\text{fit}}(n) - E_{\text{ads}}^{\text{calc.}}(n)}{\sigma_n}\right]^2, \tag{8}$$

with $N_{\text{str}}$ as the number of adlayer structures for which we have calculated adsorption energies and $\sigma_n$ as an error estimate of the calculated adsorption energies $E_{\text{ads}}^{\text{calc.}}(n)$. Note that $P(E|S,V)$ is a function of the adsorption energies $E_{\text{ads}}^{\text{calc.}}(n)$, but the integration in Eq. (6) is over $V$. This means that we should regard $\chi_1^2$ as a function of the interaction parameters $V$.

Usually one does not know much about the interaction parameters before one starts with the calculations of the adsorption energies nor about which interaction parameters to include in the model. In Eq. (6) ideas on which model $S$ is appropriate are split from ideas on which values $V$ seem reasonable. The former have a prior $P(S)$, the latter a prior $P(V|S)$. A reasonable expression for $P(V|S)$, which is also computationally convenient, is a Gaussian distribution,

$$P(V|S) = \exp\left[-\frac{1}{2}\chi_2^2\right]\prod_{m=1}^{N_{\text{par}}}\frac{1}{\sqrt{2\pi s_m^2}}, \tag{9}$$

with

$$\chi_2^2 \equiv \sum_{m=1}^{N_{\text{par}}}\left[-\frac{(V_m - V_m^{(0)})^2}{s_m^2}\right], \tag{10}$$

with $N_{\text{par}}$ as the number of parameters in the interaction model, $V_m^{(0)}$ as the most likely prior value for parameter $m$, and $s_m$ as the standard deviation. The lack of prior knowledge of the values of the interactions can be implemented by choosing large values for these deviations. We find it harder to give a general expression for $S$. If a model contains a pair interaction for adsorbates at a certain distance, then pair interactions at shorter distances should be included as well. Also, if there is a three-particle interaction in a model, then all pair interactions between these three particles having such three-particle interaction should be included too. One also want to cut off the summation in Eq. (2). Apart from these considerations it seems natural to take all interaction models equally likely.

The integration in Eq. (6) can be done easily because the integrand is a Gaussian expression in the integration variables. We define a column vector $\mathbf{c}$ via $\mathbf{c}_n \equiv E_{\text{ads}}^{\text{calc.}}(n)$, a column vector $\mathbf{e}$ via $\mathbf{e}_n \equiv E_{\text{ads}}^{\text{fit}}(n)$, and a matrix $\mathbf{A}$ via $\mathbf{A}_{nm} \equiv \sigma_n^{-2}\delta_{nm}$. With these we can write $\chi_1^2 = (\mathbf{e}-\mathbf{c})^T\mathbf{A}(\mathbf{e}-\mathbf{c})$. The fitted adsorption energies $\mathbf{e}$ can be written as $\mathbf{e}=\mathbf{M}\mathbf{v}$, with $\mathbf{v}$ as a column vector with the interaction parameters ($\mathbf{v}_m = V_m$) and the matrix $\mathbf{M}$ contains the coefficients of the interaction parameters in Eq. (2). Similarly we write $\chi_2^2 = (\mathbf{v}-\bar{\mathbf{v}})^T\mathbf{B}(\mathbf{v}-\bar{\mathbf{v}})$, with $\bar{\mathbf{v}}_n \equiv V_n^{(0)}$ and $\mathbf{B}_{nm} \equiv s_m^{-2}\delta_{nm}$. We can combine $\chi_1^2$ and $\chi_2^2$ and write the result as a quadratic function of the interaction parameters. This gives us

$$\chi_1^2 + \chi_2^2 = (\mathbf{v} - \tilde{\mathbf{v}})^T[\mathbf{M}^T\mathbf{A}\mathbf{M} + \mathbf{B}](\mathbf{v} - \tilde{\mathbf{v}}) + \mu, \tag{11}$$

with

$$\tilde{\mathbf{v}} = [\mathbf{M}^T\mathbf{A}\mathbf{M} + \mathbf{B}]^{-1}(\mathbf{M}^T\mathbf{A}\mathbf{c} + \mathbf{B}\bar{\mathbf{v}}) \tag{12}$$

and

$$\mu = \mathbf{c}^T\mathbf{A}\mathbf{c} + \bar{\mathbf{v}}^T\mathbf{B}\bar{\mathbf{v}} - \tilde{\mathbf{v}}^T[\mathbf{M}^T\mathbf{A}\mathbf{M} + \mathbf{B}]\tilde{\mathbf{v}}. \tag{13}$$

With this expression the integration in Eq. (6) then becomes

$$P(S|E) = \frac{P(S)}{P(E)}\frac{1}{\sqrt{\prod_n (2\pi\sigma_n^2)}}\prod_{m=1}^{N_{\text{par}}}\left[\frac{\tilde{s}_m}{s_m}\right]e^{-\mu/2}, \tag{14}$$

with $1/\tilde{s}_m^2$ being an eigenvalue of the matrix $\mathbf{M}^T\mathbf{A}\mathbf{M}+\mathbf{B}$.

Equation (14) can be interpreted as follows. The probability of an interaction model is higher when $\mu$ is smaller. If we assume that the prior distribution $P(V|S)$ is very broad, so that we can set $\mathbf{B}=0$, then $\chi_2^2=0$, $\tilde{\mathbf{v}}$ minimizes $\chi_1^2$, and $\mu$ is the least-squares sum of the difference between the calculated and fitted adsorption energies. This means that the probability $P(S|E)$ becomes higher if the fit becomes better, as expected.

The probability $P(S|E)$ also becomes higher when the prior errors $s_m$ become smaller. This is to be expected too because it more or less means that we know a parameter already in advance. Less easy to interpret is the dependence on the $\tilde{s}_m^2$'s. It seems that a high value for $\tilde{s}_m$ would improve the model. This however would be incorrect and also counterintuitive. As will be shown below the $\tilde{s}_m$'s are error estimates for a set of statistically independent interaction parameters. So a high value for $\tilde{s}_m$ would mean a parameter that is ill-defined, which one would not expect to improve an interaction model. The solution of this paradox is hidden in $\mu$ which also depends on the $\tilde{s}_m$'s. To see this let us compare two interaction models $S^{(1)}$ and $S^{(2)}$ with the difference that $S^{(2)}$ has one additional parameter compared to $S^{(1)}$. Let us also assume, for simplicity, they have the same prior $\{P[S^{(2)}]=P[S^{(1)}]\}$, that $\mathbf{B}=0$, and that this additional parameter is independent from the others. This means that each $\tilde{s}_m$ of $S^{(1)}$ is also found for $S^{(2)}$, but $S^{(2)}$ has an additional factor in the product of Eq. (14) which we call $\tilde{s}_N$. We then get

$$\frac{P(S^{(2)}|E)}{P(S^{(1)}|E)} = \frac{\tilde{s}_N}{s_N} e^{-(\mu^{(2)}-\mu^{(1)})/2}, \tag{15}$$

with $\mu^{(n)}$ as the least-squares sum for $S^{(n)}$. When the additional parameter in $S^{(2)}$ is independent from the others, the matrix $\mathbf{M}^T\mathbf{A}\mathbf{M}$ blocks and $\mu^{(2)}-\mu^{(1)}=-(\mathbf{M}^T\mathbf{A}\mathbf{M})_{NN}(\tilde{\mathbf{v}}_N)^2$ holds. This leads to

$$\frac{P(S^{(2)}|E)}{P(S^{(1)}|E)} = \frac{\tilde{s}_N}{s_N} e^{(\tilde{\mathbf{v}}_N/\tilde{s}_N)^2/2}. \tag{16}$$

This function is for small values of $\tilde{s}_N$ a decreasing function. Consequently, if the interaction parameters are defined better (i.e., smaller $\tilde{s}_m$), then the probability $P(S|E)$ becomes higher, again as expected. Equation (16) shows also how adding a parameter can reduce the probability of an interaction model because the ratio $\tilde{s}_N/s_N$ will usually be smaller than one and because the error estimates in the prior $P(V|S)$ will be large.

The results above depend on the error estimates $\sigma_n$. This means that the posterior really has these estimates as a parameter and should be written as $P(S|E,\sigma)$. Because we need to know the $\sigma_n$'s but have no good information on them, they are often called nuisance parameters. There are various ways to deal with this.[26,27] We find it the most convenient to set $\sigma_n=\sigma$ for all $n$ and to get a value for $\sigma$ by determining the maximum of its probability distribution $P(\sigma|E)$. This probability can be related to probability distributions that we have dealt with before.

$$P(\sigma|E) \propto P(\sigma)\sum_S \int dV P(E|S,V,\sigma)P(S,V|\sigma). \tag{17}$$

$P(E|S,V,\sigma)$ is given by Eq. (7) and $P(S,V|\sigma)$ by Eq. (9). The proportionality constant can be determined by normalization. Only $P(\sigma)$ is new. However, if we assume that it is a uniform distribution then we only need to do the integral and sum over all models. The integral has already been done before, so we only need to add all results for the different models.

The parameters $V_m^{(0)}$ and $s_m$ can also be regarded as nuisance parameters, but we handle them differently. We do not want to use any prior information for the interaction parameters, so we take large values for the error estimates $s_m$. As a consequence the values that we take for the $V_m^{(0)}$'s have then only a negligible effect on the final results (see Sec. III).

Once we have determined $S$ we also need the parameters $V$. For this we can used $P(S,V|E)$, which equals $P(E|S,V)P(S,V)$. We have already seen the two factors in this product; the only difference with what we have done before is that we do not need to integrate out the interaction parameters. The interaction parameters themselves can be obtained from $P(S,V|E)$ by computing the expectation values of $V$. The derivation above shows that the parameters should be chosen equal to components of the vector $\tilde{\mathbf{v}}$. $P(S,V|E)$ can give us also error estimates for the parameters: the covariance matrix is given by $(\mathbf{M}^T\mathbf{A}\mathbf{M}+\mathbf{B})^{-1}$.

We compare the Bayesian approach with the LOO-CV method.[17,18,28] This method works as follows. We take all structures except structure $k$. We then do a multivariate linear regression for a particular interaction model. This gives us a set of interaction parameters that we can use to predict the adsorption energy for structure $k$. We then compare this energy $E_{\text{ads}}^{\text{pred}}(k)$ with the calculated energy $E_{\text{ads}}^{\text{calc.}}(k)$. We do this not just for one structure $k$ but for all structures and define the cross-validation score or leave-one-out error,

$$R_{\text{CV}}^2 = \frac{1}{N_{\text{str}}}\sum_k [E_{\text{ads}}^{\text{pred}}(k) - E_{\text{ads}}^{\text{calc.}}(k)]^2. \tag{18}$$

This error indicates how well an interaction model predicts the energy. Adding parameters will not necessarily decrease this error. An increase indicates overfitting.

## III. LATERAL INTERACTIONS BETWEEN NO MOLECULES ON Rh(111)

In a previous paper we have used the LOO-CV method to determine the lateral interactions in O/Pt(111).[19] The results obtained with the Bayesian approach are essentially the same. The lateral interactions to be included in the cluster expansion are the same, and also the values of these interactions are the same if we assume that the prior errors of the interactions are large. This is because **B** in Eq. (12) can then be ignored and the expression reduces to the same expression we would obtain in the case of LOO-CV or a multivariate linear regression. The only difference between the Bayesian approach and LOO-CV is that they give slightly different values for the error estimates of the interactions, and the former also gives probabilities for the inclusion of parameters in the interaction model. The advantages of the Bayesian approach are better demonstrated for NO/Rh(111), which is a system with many more parameters. In O/Pt(111) the oxygen atoms only occupy threefold fcc sites. The NO molecules on Rh(111) prefer the threefold hcp site, but the difference in energy with an NO on the fcc site is small.[29,30] Moreover, at coverages above 0.5 ML the repulsion between the NO molecules becomes so large that even top sites become occupied.[31,32] This means that compared to O/Pt(111) there are three times the number of interaction parameters between adsorbates on like sites plus parameters for the interaction parameters for adsorbates on different types of sites.

We have done DFT calculations of various adlayer structures of NO/Rh(111) with the VASP code,[33] which uses a plane-wave basis set and the (relativistic) ultrasoft pseudopotentials introduced by Vanderbilt[34] and generated by Kresse and Hafner.[35] The generalize gradient approximation of Perdew and Wang (PW-91) has been used.[36] All calculations on NO/Rh(111) were done with a surface model consisting of a supercell with a slab of five metal layers separated by five metal layers replaced by vacuum. Grids of size 3×3×1–9×9×1 (depending on the unit cell of the adlayer) for Brillouin-zone sampling obtained via the Monckhorst-package were used and a cutoff of 400 eV. This yielded in all cases adsorption energies per NO molecule converged to within 5 kJ/mol with respect to $k$-point sampling, energy cutoff, number of slab and vacuum layers, and cell size.

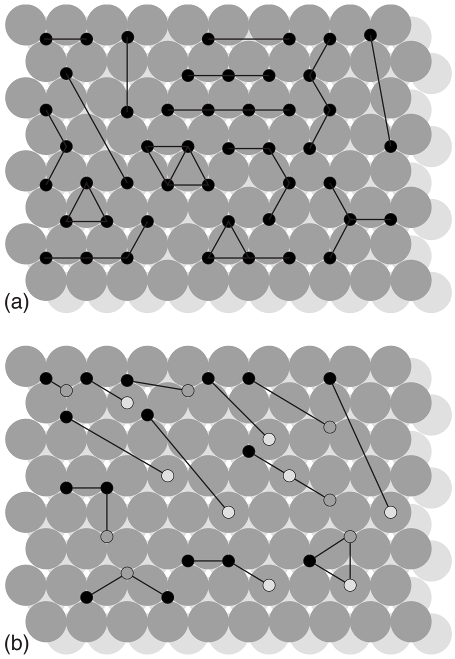

FIG. 1. Definition of all lateral interactions. The large circles are Rh atoms. Only the first two layers of the substrate are shown. (a) shows the interactions between NO molecules on hcp sites. There is a similar set of interactions for fcc and top sites. (b) shows the interactions between NO molecules on different types of sites. Only a selection of all of these interactions is shown. The others have the same distances between the molecules and angles between the intermolecular vectors but have molecules at different sites.

The adsorption energy per NO molecule was computed for 74 different structures of the adlayer. Of these structures 19 contained adsorbates only on hcp sites, 18 adsorbates only on fcc sites, 14 adsorbates only on top sites, and 23 adsorbates on different sites. An initial set of about 65 structures was created to have a good representation of all interactions that might possibly be present in the system. It also contained all experimentally observed structures. The structures not in this initial set were added because these structures were observed in kinetic Monte Carlo simulations with interactions obtained from using a limited set of structures or related to these structures by shifting the whole adlayer so that the adsorbates became adsorbed at different sites.[37–39]

For each adsorption site (hcp, fcc, and top) we have one interaction parameter, the adsorption energy, for an isolated NO molecule at such site. Figure 1 shows the lateral interactions that we have considered as well. In all there are 91 candidate parameters for the interaction model. The pair interactions between NO molecules on different types of sites, a distance $a/\sqrt{3}$ apart, with $a$ as the nearest distance between two Rh atoms, were not included. The reason is that at this distance there is such a strong repulsion between the NO molecules that the molecules in the geometry optimization of the DFT calculations are pushed to other sites. In the kinetic Monte Carlo simulations we used somewhat arbitrarily a value for this interaction of 100 kJ/mol, which prevented the molecules getting so close. In the real system this approach is also avoided.

To get some idea what to choose for the priors $P(V|S)$ we did some linear regressions using only a small number of interaction parameters. These showed that the most important interactions were the adsorption energies and the nearest-neighbor interaction for NO molecules at like sites. Based on these calculations we chose for $V_m^{(0)}$ in the prior $-250$ kJ/mol for all three adsorption energies and 15 kJ/mol for the interaction between nearest neighbors on like sites and next-nearest neighbors on different types of sites. For the other lateral interactions we chose 2.5 kJ/mol. As these estimates are rather crude, and we did not want the prior to affect the final values of the interaction parameters, we used a value of 100 kJ/mol for all error estimates $s_m$.

Because the number of interaction parameters is so large the total number of subsets $S$ is enormous $(2^{91}-1)$. It is therefore not possible to generate all terms in summations over $S$ as in Eq. (17). It is also not possible to consider each interaction model of course. To deal with this problem we did the following. We started with a guess for the accuracy of the DFT calculations $\sigma$. We used $\sigma=10$ kJ/mol. Then we took as a first set of interaction parameters only the three adsorption energies of the isolated NO molecules. Next we added in turn each of the other interaction parameters and determined which one gave the largest increase in $P(S|E)$. This one was then added to the set. With this extended set this procedure of adding each interaction parameter in turn and determining the change in $P(S|E)$ was repeated. This led to an iterative procedure of adding again and again one parameter to the set until a maximum for $P(S|E)$ was obtained. To check that the model of the interaction parameters did not correspond to a local maximum of $P(S|E)$, we also looked how the probability changed when two interaction parameters instead of one were added, and we looked at the changes in $P(S|E)$ upon removing interaction parameters. Because we never considered an interaction model with more than about 25 parameters, we never had the problem that we did not have sufficient structures to determine the parameters.

This procedure gave us a smaller set of parameters for which we could generate all subsets $S$. We used this subset to determine $\sigma$ by maximizing $P(\sigma|E)$ in Eq. (17). This implies that we assumed that for interaction models with other parameters $P(S|E)=0$. With this value for $\sigma$ the procedure in the previous paragraph was repeated. Occasionally this led to the inclusion of new interaction parameters. In that case the whole procedure had to be repeated. We also repeated the whole procedure when new interaction parameters had to be included because we added new adlayer structures.

The final results gave us an error estimate for the DFT calculations of $\sigma=3.5$ kJ/mol. This is about the same as the root-mean-square deviation of the fitted adsorption energies with respect to the calculated ones. In fact, it seems that this root-mean-square deviation can be used as an error estimate for $\sigma$ that is somewhat easier to obtain. The value of $\sigma$ is rather small. We have noted before that this is in part due to the fact that there is probably a systematic component in the error of the DFT adsorption energies.[20] Such a systematic component does not affect $\sigma$ or the root-mean-square deviation. This can be seen best when we have only one type of site. In Eq. (2) the adsorption energy of that site then has $c_m(n)=1$ for all structures $n$. Consequently, a systematic error then only shifts the adsorption energy but does not affect

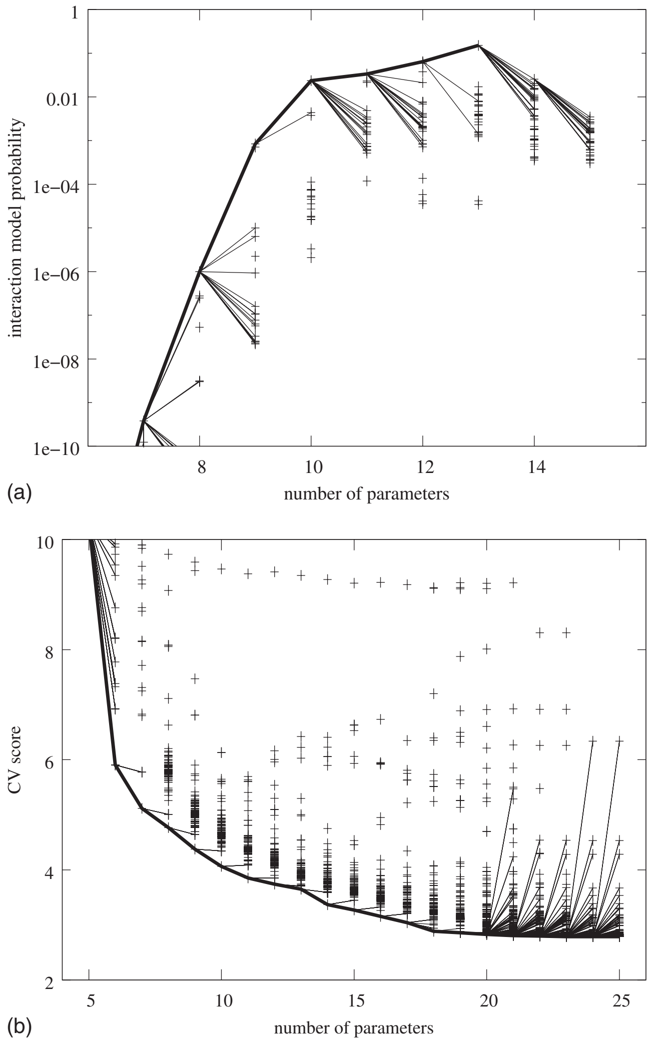

FIG. 2. The probabilities of the (a) interaction models and (b) CV score (in kJ/mol) as a function of the number of parameters in the model. The lines are guides for the eyes and show what happens when a parameter is added. Models not connected by a line were obtained by removing a parameter. The fat lines indicate that adding a parameter improves the model.

the other interaction parameters. Because we have three adsorption sites in NO/Rh(111), the situation is a bit more complicated, but a systematic error will again only cause a shift of the adsorption energies, which moreover is the same for all sites. This is because the sum of the coefficients of the adsorption energies in Eq. (2) is equal to 1 for all structures. Note that by systematic error we mean a contribution to $E_{\text{ads}}^{\text{calc.}}$ that is numerically the same for all structures. This may, but need not, be related to some particular physical mechanism of an interaction that DFT does not properly take into account.

Figure 2 shows how the probabilities $P(S|E)$ change when parameters are added to and removed from the interaction model. Adding parameters initially increases a model's probability, but there is a well-defined maximum after which adding parameters only decreases the probability. The fact that there is such a clear maximum is an important advantage of the Bayesian approach over the LOO-CV method. The results of that method are shown in the figure too. The CV score is shown to decrease initially, but then it becomes almost constant when more parameters are added. It is very hard to determine the minimum of the CV score. Moreover, we found that the interaction model corresponding to the minimum depends quite sensitively on adding or removing structures in the fit. So it is difficult to determine the interaction model with the LOO-CV method. This problem has been observed before for the determination of interactions in alloys.[22–24]

Figure 2 shows that the model with the maximum probability has 13 parameters. It is also seen that the CV score is still clearly decreasing for that model. The Bayesian approach thus gives a cluster expansion with fewer terms than LOO-CV. The smallest CV score is 2.7 kJ/mol, which is only a bit smaller than the 3.5 kJ/mol for the root-mean-square deviation for the 13 parameter models. As this score is achieved at the cost of almost doubling the number of interaction parameters, one may wonder if LOO-CV really prevents overfitting.

Because of the difficulty of determining the minimum in the CV score the LOO-CV method takes also somewhat more time to do. Compared to a single calculation of a probability of an interaction model a calculation of a single CV score takes more time. This is because for the Bayesian approach Eqs. (12)–(14) have to be calculated only once, whereas for the CV score such a calculation has to be done for each calculation of $E_{\text{ads}}^{\text{pred}}(k)$ in Eq. (18). On the other hand the determination of the error estimate $\sigma$ in Eq. (8) means that the whole procedure in the Bayesian approach has to be repeated several times. The time to do any of these methods is negligible however compared to the time it takes to do the DFT calculations.

The interaction model with the maximum probability has a probability of 0.15. This may seem low, but one should realize that there are many interaction models that also have a relatively high probability (but still much lower than the one with the maximum probability) and that differ in only a few parameters from the interaction model with the highest probability. There are three interaction models that can be obtained from the most likely model by leaving out one of the next-nearest pair interaction for NO molecules on different types of sites with probabilities of 0.06, 0.06, and 0.04. Adding a next-next-nearest pair interaction for one NO on an hcp and one NO on an fcc site gives a model with probability of 0.06. All other models have lower probabilities, but there are 17 models with probabilities of 0.01 or higher.

It is more insightful to look at the probabilities of the individual parameters. We can calculate these probabilities by summing the probabilities of all interaction models that contain a particular parameter. The results are shown in Table I. From that table we see that there is a clear distinction between parameters that should be included in the interaction model and those that can be left out. There are only a few exceptions: the next- and next-next-nearest pair interactions for NO's on hcp sites and the next-next-nearest pair interaction for one NO on hcp and one NO on fcc. In the most likely model the first two are included, but the last one is not.

From Eq. (15) it seems that adding an interaction parameter can be made to lead to a lower probability $P(S|E)$ by

TABLE I. Probabilities for the some important interactions. Interaction parameters that are not shown have a small probability; the largest being 0.03 for the linear four-particle interactions for hcp and fcc sites. Note that the nearest pair interaction between NO molecules on different types of site has not been determined (see text).

| | hcp | fcc | top | hcp-fcc | hcp-top | fcc-top |
|---|---|---|---|---|---|---|
| Adsorption energy | 1.00 | 1.00 | 1.00 | | | |
| Nearest pair interaction | 1.00 | 1.00 | 1.00 | | | |
| Next-nearest pair interaction | 0.67 | 0.01 | 0.01 | 0.87 | 0.90 | 0.91 |
| Next-next-nearest pair interaction | 0.59 | 0.00 | 0.00 | 0.30 | 0.02 | 0.02 |
| Linear three-particle interaction | 0.99 | 1.00 | 0.02 | | | |
| Triangular three-particle interaction | 0.04 | 0.06 | 0.04 | | | |
| Bent three-particle interaction | 0.01 | 0.01 | 0.01 | | | |

simply taking a large value of the prior error estimate $s_N$. This is indeed the case. However, the probability $P(S|E)$ varies by many orders of magnitude as can be seen in Fig. 2. To change the ordering of the probabilities of the interaction models appreciably, the prior error estimates would have to be increased by an order of magnitude or more as well, giving exaggerated values for them. Reducing the values of these error estimates can lead to the acceptance of more interaction parameters, but then one is effectively saying that one already knows the value of a parameter in advance. We have found that there is a large range for the prior error estimates that hardly fixes the values of the interaction parameters in advance, yet always leads to the same preferred interaction model.

Table II shows the values of the interaction parameters for the most likely interaction model. All the lateral interactions are repulsive. Note that some parameters have values close to the value $V_m^{(0)}$ of the prior, but some differ quite substantially. In particular the adsorption energy of the top site differs about 60 kJ/mol. The error estimates in the table are obtained from the diagonal elements of the matrix $\mathbf{M}^T\mathbf{A}\mathbf{M}+\mathbf{B}$, which is the inverse of the covariance matrix for the interaction parameters. The relative error can be quite large for the weaker interactions, but this is only because these interactions are so weak. The covariances are generally in absolute value smaller than 1 $\text{kJ}^2/\text{mol}^2$ except the one for the nearest-neighbor pair interaction and the linear three-particle interaction which are $-2.2$ $\text{kJ}^2/\text{mol}^2$ for NO's on hcp and $-2.5$ $\text{kJ}^2/\text{mol}^2$ for NO's on fcc sites. This means that if we would leave out the three-particle interactions, we would find a too high nearest-neighbor pair interaction for NO's on hcp and fcc sites. This was indeed found in a fit without these three-particle interactions.

Initially we did not use all the 74 structures. We started with a somewhat smaller set. We used this set to determine the lateral interactions and then did kinetic Monte Carlo simulations. These simulations showed some structures that were not in our initial set. So we did DFT calculations for these structures as well and calculated the lateral interactions again. This caused however only very small changes in the interaction parameters, which indicates that the method is robust; i.e., the results do not depend on details of the calculations. Because these structures show up in simulations, it should be possible to observe them in experiments as well. Indeed, this has been done in the meantime with scanning-tunneling microscopy.[40]

Figure 3 shows the calculated and fitted adsorption energies as a function of coverages. The figure confirms that the difference between the calculated and fitted results is small. The exceptions can be mainly found among the less relevant structures with a high energy. These are structures that contain structures with NO molecules on top sites at low coverage. Stable structures with adsorbates on top site can only be found above 0.5 ML, but then the fit is good. The tangent construct shows the stable phases at 0 K. The calculated and the fitted results show that there are phase transitions at 0.5 and 0.778 ML. The structures at which the transition takes place are the $c(4\times2)-2\text{NO}$ (Refs. 31 and 41) and the 3

TABLE II. Values for the interaction parameters with standard deviations in parentheses (in kJ/mol) for the most likely interaction model. The parameter with the largest probability (0.30) that has not been included in this model is the next-next-nearest pair interaction between one NO molecule on an hcp and one NO on an fcc site.

| | hcp | Fcc | top | hcp-fcc | hcp-top | fcc-top |
|---|---|---|---|---|---|---|
| Adsorption energy | −254.4 (1.4) | −245.6 (1.1) | −190.4 (1.2) | | | |
| Nearest pair interaction | 15.6(1.8) | 17.1(1.6) | 19.1(1.0) | | | |
| Next-nearest pair interaction | 3.0(1.0) | | | 4.1 (1.4) | 5.7 (2.0) | 5.8 (1.9) |
| Next-next-nearest pair interaction | 3.1(0.9) | | | | | |
| Linear three-particle interaction | 6.7(1.9) | 8.5(1.8) | | | | |

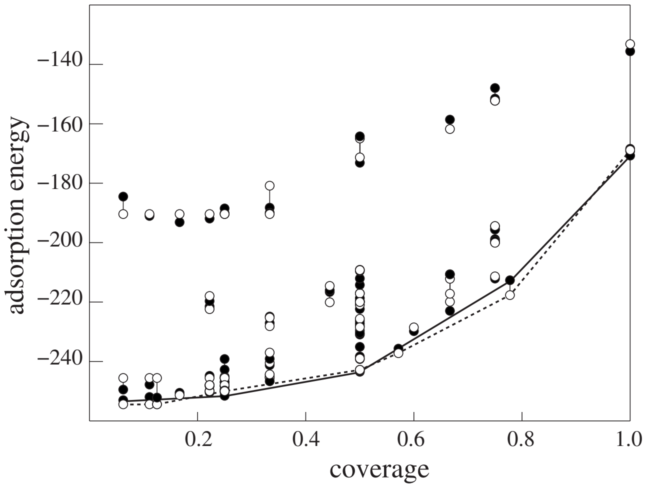

FIG. 3. Adsorption energy (in kJ/mol) per NO molecule as a function of coverage. The closed circles are DFT results; the open circles are results from the most likely interaction model. Corresponding structures are connected by thin lines. The fat lines are the tangent constructs for the DFT results (solid) and the model results (dashed).

$\times 3-7$NO structures.[32] Both have been observed experimentally before. Other structures have been observed as well, but they can be shown to be entropy stabilized.[40]

## IV. SUMMARY AND CONCLUSIONS

We have presented an alternative to the LOO-CV method for the truncation problem of the cluster expansion for lateral interactions when using DFT calculations. The method uses Bayesian statistics. The main advantage seems to be that the method clearly shows which parameters include in the interaction model even if the set of all candidate parameters becomes large. The Bayesian approach seems to yield a more compact interaction model with fewer terms in the cluster expansion. The resulting model fits the DFT results only slightly worse than the model obtained with LOO-CV. As the latter yields many more terms of the expansion, it might be that we still get overfitting with LOO-CV. The Bayesian method also has the advantage that the results are easier to analyze. Apart from the values of the parameters we get probabilities that they should be included in an interaction model and correlations between these parameters. A drawback of the Bayesian approach is that the theory is more complicated than that of LOO-CV, and it requires the determination of some parameters (prior probabilities and error estimates of the data to be fitted) that is not needed in LOO-CV.

We used the Bayesian approach to compute the interactions for NO/Rh(111). This is a complicated system because the NO molecules adsorb at three different sites depending on coverage. The approach yields an interaction model with 13 parameters out of a total of 91 possible parameters. Comparison of the resulting structures using these parameters in kinetic Monte Carlo simulations shows good agreements with established experimental results. We also found structures that we could later confirm using scanning-tunneling microscopy.[40]

We have taken a system from surface science as an illustrative example because we are working in that field. However, the Bayesian approach should also be useful for the study of alloys. In fact, it might even be more useful there because the number of terms in the cluster expansion for alloys increases more rapidly than for layers of adsorbates.

## ACKNOWLEDGMENTS

This research was financially supported by the Foundation for Fundamental Research on Matter (FOM), The Netherlands, and by the Deutsche Forschungsgemeinschaft (DFG), Germany, in the frame of the DFG-priority program No. SPP 1091, “Bridging the gap between ideal and real systems in heterogeneous catalysis.”

---

[1] J. Trost, T. Zambelli, J. Wintterlin, and G. Ertl, Phys. Rev. B **54**, 17850 (1996).

[2] L. Österlund, M. O. Pedersen, I. Stensgaard, E. Lægsgaard, and F. Besenbacher, Phys. Rev. Lett. **83**, 4812 (1999).

[3] S. Renisch, R. Schuster, J. Wintterlin, and G. Ertl, Phys. Rev. Lett. **82**, 3839 (1999).

[4] N. V. Petrova and I. N. Yakovkin, Surf. Sci. **519**, 90 (2002).

[5] C. Schwennicke and H. Pfnür, Phys. Rev. B **56**, 10558 (1997).

[6] P. Piercy, K. De’Bell, and H. Pfnür, Phys. Rev. B **45**, 1869 (1992).

[7] D. L. S. Nieskens, A. P. van Bavel, and J. W. Niemantsverdriet, Surf. Sci. **546**, 159 (2003).

[8] L. J. Whitman and W. Ho, J. Chem. Phys. **90**, 6018 (1989).

[9] M. Kiskinova, A. Szabo, and J. T. Yates, Jr., J. Chem. Phys. **89**, 7599 (1988).

[10] A. Szabo, M. Kiskinova, and J. T. Yates, Jr., J. Chem. Phys. **90**, 4604 (1989).

[11] A. P. J. Jansen, Phys. Rev. B **69**, 035414 (2004).

[12] J. T. Stuckless, N. Al-Sarraf, C. Wartnaby, and D. A. King, J. Chem. Phys. **99**, 2202 (1993).

[13] L. Vattuone, Y. Y. Yeo, and D. A. King, J. Chem. Phys. **104**, 8096 (1996).

[14] Y. Y. Yeo, L. Vattuone, and D. A. King, J. Chem. Phys. **106**, 1990 (1997).

[15] R. Kose, W. A. Brown, and D. A. King, J. Phys. Chem. B **103**, 8722 (1999).

[16] J. N. Murrell, S. Carter, P. Huxley, S. C. Farantos, and A. J. C. Varandas, *Molecular Potential Energy Functions* (Wiley, Chichester, 1984).

[17] A. van der Walle and G. Ceder, J. Phase Equilib. **23**, 348 (2002).

[18] V. Blum and A. Zunger, Phys. Rev. B **69**, 020103(R) (2004).

[19] A. P. J. Jansen and W. K. Offermans, in *Computational Science and its Applications—ICCSA-2005, LNCS 3480*, edited by O. Gervasi (Springer, Berlin, 2005).

[20] C. G. M. Hermse and A. P. J. Jansen, in *Catalysis*, edited by J. J. Spivey and K. M. Dooley (Royal Society of Chemistry, London, 2006), Vol. 19.

[21] Y. Zhang, V. Blum, and K. Reuter, Phys. Rev. B **75**, 235406 (2007).

[22] N. A. Zarkevich and D. D. Johnson, Phys. Rev. Lett. **92**, 255702 (2004).

[23] R. Drautz and A. Díaz-Ortiz, Phys. Rev. B **73**, 224207 (2006).

[24] D. E. Nanu, Y. Deng, and A. J. Böttger, Phys. Rev. B **74**, 014113 (2006).

[25] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).

[26] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Chapman & Hall, London/CRC, Boca Raton, FL, 2003).

[27] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, Oxford, 2006).

[28] D. M. Hawkins, J. Chem. Inf. Comput. Sci. **44**, 1 (2004).

[29] D. Loffreda, D. Simon, and P. Sautet, Chem. Phys. Lett. **291**, 15 (1998).

[30] D. Loffreda, F. Delbecq, D. Simon, and P. Sautet, J. Chem. Phys. **115**, 8101 (2001).

[31] I. Zasada, M. A. V. Hove, and G. A. Somorjai, Surf. Sci. **418**, L89 (1998).

[32] K. B. Rider, K. S. Hwang, M. Salmeron, and G. A. Somorjai, Phys. Rev. Lett. **86**, 4330 (2001).

[33] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[34] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[35] G. Kresse and J. Hafner, J. Phys.: Condens. Matter **6**, 8245 (1994).

[36] J. P. Perdew, in *Electronic Structure of Solids '91*, edited by P. Ziesche and H. Eschrig (Akademie, Berlin, 1991), p. 11.

[37] J. J. Lukkien, J. P. L. Segers, P. A. J. Hilbers, R. J. Gelten, and A. P. J. Jansen, Phys. Rev. E **58**, 2598 (1998).

[38] A. P. J. Jansen, an introduction to Monte Carlo simulations of surface reactions, arXiv:cond-mat/0303028.

[39] A. P. J. Jansen, The kinetic Monte Carlo website, http://www.catalysis.nl/chembond/kMC/, 2006.

[40] J. H. A. Hagelaar, *The Role of the Electron Trajectory in Scanning Tunneling Microscopy: Elastic and Inelastic Tunneling Through NO On Rh(111)* (Eindhoven University of Technology, Eindhoven, 2008).

[41] C.-T. Kao, G. S. Blackman, M. A. V. Hove, G. A. Somorjai, and C.-M. Chan, Surf. Sci. **224**, 77 (1989).