



HAL
open science

Une étude théorique de l'hydrogénation et l'oligomérisation de l'acétylène sur des catalyseurs de palladium

Emanuele Vignola

► **To cite this version:**

Emanuele Vignola. Une étude théorique de l'hydrogénation et l'oligomérisation de l'acétylène sur des catalyseurs de palladium. Catalysis. Université de Lyon, 2017. English. NNT : 2017LYSEN054 . tel-01968019

HAL Id: tel-01968019

<https://theses.hal.science/tel-01968019>

Submitted on 2 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Numéro National de Thèse : **2017LYSEN054**

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée par
l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N° 206
Ecole Doctorale de Chimie (Chimie, Procédés,
Environnement)

Spécialité de doctorat : Chimie Théorique
Discipline : Chimie

Soutenue publiquement le 29/09/2017, par :
Emanuele VIGNOLA

**A Theoretical Perspective on
Hydrogenation and Oligomerization of
Acetylene over Pd Based Catalysts**

**Une étude théorique de l'hydrogénation et l'oligomérisation de l'acétylène sur
des catalyseurs de palladium**

Devant le jury composé de :

Studt, Felix	Professeur	Karlsruhe Institut of Technology	Rapporteur
Honkala, Karoliina	Professeure	University of Jyväskylä	Rapporteuse
Chizallet, Céline	Ingénieure de recherche	IFPEN	Examinatrice
Filhol, Jean-Sébastien	Professeur	Université de Montpellier	Examineur
Curulla, Daniel	Chercheur	TOTAL Research and Technology Feluy	Examineur
Sautet, Philippe	Professeur	University of California at Los Angeles	Directeur de thèse
Steinmann, Stephan	Chargé de recherche	Ecole Normale Supérieure de Lyon	Co-encadrant

Table of Contents

Chapter 1. Introduction	1
Industrial context	1
Ethylene: a key building block in the petrochemical industry	1
Removal of the acetylenic impurity	1
Selective hydrogenation of acetylene: Experiments and Theoretical Modelling	2
Reaction network complexity.....	2
Chemical physics of the catalyst: hydrides and carbides	4
Pd-Ag alloy catalysts for selective hydrogenation of acetylene.....	5
Statement of Purpose	7
Chapter 2. Methodologies	9
The ensembles chemical equilibrium.....	9
Ensemble description of acetylene adsorbed on Pd-Ag alloys	9
Pd-Ag triangular ensembles as species in chemical equilibrium.....	10
Cluster expansions	12
Monte Carlo methods	15
A first skirmish with Chapter 6: machine learning and adsorbed layers.....	16
Chapter 3. C ₂ H ₂ -Induced Surface Restructuring of Pd-Ag Catalysts: Insights from Theoretical Modelling	18
Chapter 4. A Monte Carlo Study of Acetylene Adsorption on Pd-Ag Alloys	19
Introduction.....	19
Computational Details.....	21
Results and Discussions.....	21
Conclusions.....	27
References.....	27
Chapter 5. Evaluating the Risk of C-C Bond Formation during Selective Hydrogenation of Acetylene on Palladium.....	29
Introduction.....	29
Computational Details.....	32
Results and Discussions.....	32
Conclusions.....	38
References.....	40
Chapter 6. A Machine Learning Approach to Graph-Theoretical Cluster Expansions of the Energy of Adsorbate Layers	42
I. INTRODUCTION	42
II. THE CORRESPONDENCE BETWEEN REAL AND GRAPHICAL INTERACTIONS.....	44

III. A RECOGNITION MEASURE TO MAP MOLECULAR CONFIGURATIONS INTO PATTERNS ON THE GRAPHICAL LATTICE	48
IV. THE OCCURRENCE MATRIX AND ITS INVERSION	50
V. APPLICATION TO THE ETHYLENE-PALLADIUM SYSTEM	52
V.1 THE TRAINING SET	53
V.2 PATTERN RECOGNITION	54
V.3 CLUSTER EXPANSION OF THE INTERACTION ENERGY	57
VI. CONCLUSIONS	61
VII. SUPPLEMENTARY MATERIAL	62
REFERENCES	62
Conclusions and Outlook	66
Annex	69
Annex of Chapter 4	69
Annex of Chapter 5	70

Chapter 1. Introduction

Industrial context

Ethylene: a key building block in the petrochemical industry

In a world changed by materials such as plastics, detergents, fertilizers, the role of the petrochemical industry does not require any emphasis. At the core of the petrochemical industry lies ethylene, a small organic molecule whose chemical functionality allows for a wide range of specialty chemicals on today's market, from low and high density polyethylene to ethylene glycol, PVC, polyesters, to name a few. A 2006 report¹ shows ethylene dominated the global organic chemistry production in the pre-crisis era; 11 years later, with recession basically at their shoulders, the main actors in the American petrochemical industry still plan to open new ethylene plants as the market's demand increased again.²

Production of ethylene is provided by industrial plants known as crackers.^{3,4} Originally, cracking technology arose from the need of gasoline to fuel aircraft and general inner-combustion engines in the 1920s; until then, simple distillation of crude oil was indeed limited to give a medium-high boiling derivative, kerosene, to be burnt in lamps. The idea lying behind cracking is to break high molecular weight components of crude oil into smaller fragments, which then rearrange to give useful products. There are two kinds of cracking. The one kind is promoted by catalysts – usually zeolites – and enhances the production of propylene and butylene, materials of great interest but outside of the scope of the present discussion; the other kind is thermally activated instead, and favours ethylene. Thermally activated cracking of heavy oil to lighter substances proceeds via radicalic mechanism. As it is well-established in chemistry, thermally-activated radicalic reactions are not very selective;⁵ thermodynamically stable ethylene emerges at relatively high temperatures (> 900 °C) but it is unavoidably accompanied by small impurities of methane and acetylene.⁶ The latter, typically detected by gas-chromatography,⁷ nowadays also spectroscopically,⁸ is of special harm to ethylene quality, due to its elevated and analogous reactivity. Indeed, acetylene compromises the main application of the ethylenic feedstock, namely ethylene polymerization, both participating to the polymerization reaction and poisoning the Ziegler-Natta catalyst thereof; this results into lower-grade polyethylene and slow polymerization rates.

Removal of the acetylenic impurity

Chemical technology has provided the industry with tools to suppress the noxious acetylenic impurity in ethylene.⁹ Apart from the not very common extraction with solvents like acetone or dimethylformamide, the most frequent method is selective hydrogenation, i.e. the conversion of acetylene into ethylene with no further reduction to ethane. Refineries realize this transformation in fixed-bed catalytic units of two configurations.¹⁰ In front-end configurations, selective hydrogenation

¹ *Chem. Eng. News*, **2006**, 84 (28), 59-68

² *Chem. Eng. News*, **2017**, 95 (2), 26-33

³ Speight, J. G. *The Chemistry and Technology of Petroleum*, 5th ed.; CRC Press; Boca Raton (FL); 2014

⁴ Wessermel, K.; Arpe, H.-J. *Industrial Organic Chemistry*, 3rd ed.; Wiley-VCH; Weinheim (Germany); 1997

⁵ Bar, G.; Parsons, A. F. *Chem. Soc. Rev.* **2003**, 32, 251-263

⁶ <http://www.sigmaaldrich.com/catalog/product/supelco/7940124?lang=fr®ion=FR>. Access: 16/05/2017

⁷ Paylor, R. A. L.; Feynland R. *Anal. Chem.* **1961**, 33, 808-809

⁸ Kluczynski, P.; Jahjah, M.; Nähle, L.; Axner, O.; Belahsene, S.; Fischer, M.; Koeth, J.; Rouillard, Y.; Westberg, J.; Vicet, A.; Lundqvist, S. *Appl. Phys. B*, **2011**, 105, 427-434

⁹ http://www.klmtechgroup.com/PDF/Articles/acetylene_converter.pdf. Access: 16/05/2017

¹⁰ Hou, R. *Catalytic and Process Study of the Selective Hydrogenation of Acetylene and 1,3-Butadiene*, PhD Thesis; Springer; Singapore; 2017

takes place while the ethylenic stream still contains additional impurities of methane and carbon monoxide, which are removed afterwards; back-end configurations purify the stream first, allowing only an inlet of C_2 to the reductive unit.

The process is catalytic in nature. Though the amount of catalytic formulations is rather numerous, catalyst for acetylene partial reduction can be collected into two major classes. First of these is Ni based, pure or alloyed to Co; selectivity towards ethylene is not high, and the feed must be kept sulphured constantly to sustain catalytic activity: the reactor configuration must therefore be of the front-end type. The other class of catalysts is Pd-based, again pure or alloyed to *sp* or transition metals. The development of Pd technology improved the reaction selectivity so as to guarantee the maximum acetylenic impurity of high quality ethylene (5 ppm). Research on Pd-based catalysts is still active and focuses mainly on the control of the catalyst's morphology and on the use of different second metals (also called *promoters* in the literature). For the purposes of the present work, the focus will stay on pure Pd and Pd-Ag alloys, as commissioned by our industrial partner, Total Research and Technology Feluy.

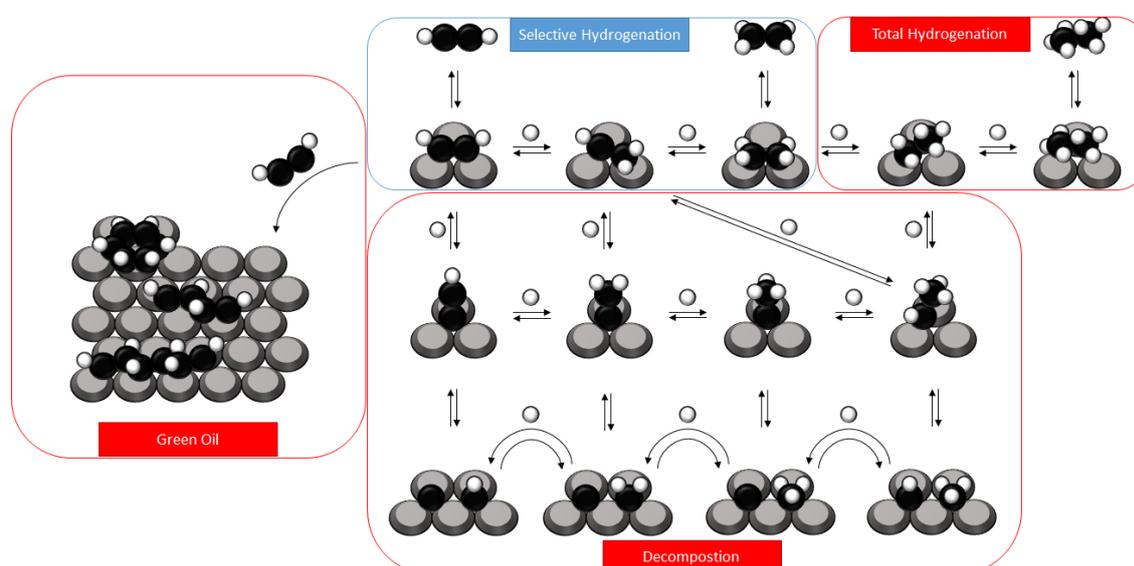


Figure 1. Reaction mechanism overview

Selective hydrogenation of acetylene: Experiments and Theoretical Modelling

Reaction network complexity

On the basis of the reaction stoichiometry, it might be inferred that the main issue of partial hydrogenation of acetylene resides in the reactant and the desired product having the same functional group; since they are both unsaturated molecules, they show analogous chemical affinity towards hydrogen. As a result, this may lead to the total hydrogenation product, ethane. Though total hydrogenation is certainly a problematic issue in the process, it is accompanied with additional side reactions which compromise the catalyst lifespan. The complex scenario of acetylene selective hydrogenation, reviewed in 1993¹¹ and 2008,¹² is best visualized and discussed by inspecting the scheme proposed by Mei *et al.*,¹³ which is displayed and revisited in Figure 1. As the Figure shows, hydrogen atoms are added to acetylene in a consecutive fashion. This mechanism, known as Horiuti-

¹¹ Bos, A. N. R.; Westerterp K. R. *Chem. Eng. Proc.* **1993**, *32*, 1-7

¹² Borodzinski, A.; Bond, G. C. *Catal. Rev.* **2008**, *50*, 379-469

¹³ Mei, D.; Neurock, M.; Smith, C. M. *J. Catal.* **2009**, *268*, 181-195

Polanyi mechanism,¹⁴ does not seem to indicate any selectivity in the transformation, which is traditionally explained by the different interaction of acetylene and ethylene with the catalytic surface.^{15,16,17} Addition of atomic hydrogen to acetylene leads to the vinyl intermediate. If vinyl is hydrogenated symmetrically – i.e. each of the two carbon atoms are hydrogenated with one hydrogen atom – it reaches ethylene, which is then reduced to ethyl, then ethane.

Dissymmetric hydrogenation of vinyl leads to ethylidene, which is hydrogenated to ethyl thereby ending into the total hydrogenation product, or can be de-hydrogenated to ethylidyne. This route was matter of speculation until the first computational study of 2002;¹⁸ then, the refining work of Moskaleva *et al.*¹⁹ motivated an experimental validation which finally arrived in the form of an IR spectroscopy analysis.²⁰ De-hydrogenation of ethylidyne forms vinylidene, an isomer of acetylene. De-hydrogenation pathways are also shown to decompose acetylene in a drastic way, till formation of graphitic layers.²¹

Despite the stability of ethylidyne, dissymmetric by-products are not as harmful as the evolution of oligomers. The so-called “green oil” is the major responsible of catalyst deactivation and forces ethylene producers to run parallel operations to allow for catalyst renewal.²² The oligomers formed are an intricate mixture of aliphatic hydrocarbons with even numbers, ranging from C₄ to C₂₈, sometimes reaching as high as C₆₀ and even more. Three main fractions can be identified as constituting green oil, a C₄–C₆ light-end fraction which gets carried away with the stream, a liquid fraction made of droplets and a sticky fraction, the heaviest, which accumulates on the surface.²³ A few groups tried to analyse green oil to establish its composition, with results rarely in agreement. Two industrial studies done using slipstream test units in ethylene plants in Russia and China using unmodified Pd catalysts can be cited. Yayun *et al.*²⁴ for instance found that liquid polymers have an average composition C_nH_{(1.8-1.9)n} with a number of carbon atoms between 14 and 17, and an olefin/paraffin ratio between 0.1 and 0.4. Most remarkably, no aromatics were detected, although palladium is famed to boost acetylene trimerization.²⁵ Gandman *et al.*²⁶ reported different green oil composition, with a content of aromatic compounds as high as 74.5% in weight, 23.0% of olefins, and a small quantity of di-olefins. The origins of the oligomers is also controversial. Most authors agree that they exclusively derive from acetylene, and that they are produced only in the presence of hydrogen; therefore, the reaction is often referred to as acetylene hydro-oligomerization. Nevertheless, according to the results obtained by Yayun *et al.* the oligomerization of ethylene cannot be ignored. Oligomer formation has only been timidly tackled by computational scientists. Lopez and

¹⁴ Mattson, B.; Foster, W.; Greimann, J.; Hoette, T.; Le, N.; Mirich, A.; Wankum, S.; Cabri, A.; Reichenbacher, C.; Schwanke, E. *J. Chem. Educ.* **2013**, *90*, 613-619

¹⁵ Al-Ammar, A. S.; Webb, G. *J. Chem. Soc., Faraday Trans. 1* **1978**, *74*, 195-205

¹⁶ Al-Ammar, A. S.; Webb, G. *J. Chem. Soc., Faraday Trans. 1* **1978**, *74*, 657-664

¹⁷ Al-Ammar, A. S.; Webb, G. *J. Chem. Soc., Faraday Trans. 1* **1979**, *75*, 1900-1911

¹⁸ Pallassana, V.; Neurock, M.; Lusvardi, V. S.; Lerou, J., J.; Kragten, D. D.; A. van Santen R. A. *J. Phys. Chem. B* **2002**, *106*, 1656-1669

¹⁹ Moskaleva, L. V.; Chen, Z.-X.; Aleksandrov, H. A.; Mohammed, A. B.; Sun, Q.; Rösch, N. *J. Phys. Chem. C* **2009**, *113*, 2515

²⁰ Stacchiola, D.; Tysoe W. T. *J. Phys. Chem. C*, **2009**, *113* (19), 8000-8001

²¹ Kozlov, S. M.; Yudanov, I. V.; Aleksandrov, H. A.; Rosch, N. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10955-10963

²² Curulla, D.; Chassard, O. Total Internal Report RC-N-2014-CRBC-033

²³ Curulla, D.; Chassard, O. Total Internal Report RC-N-2014-CRBC-033

²⁴ Yayun, L.; Jing, Z.; Xueru M. *AIChE Beijing* **1982**, *2*, 688-702

²⁵ Gentle, M.; Muettterles E., L. *J. Chem. Phys.* **1983**, *87*, 2469–2472.

²⁶ Gandman, Z. E.; Aerov, M. E.; Men'shchikov, V. A.; Getsmantsev V. S. *Intl. Chem. Eng.* **1975**, *15* (1), 183-185

Vargas-Fuentes²⁷ proposed an acetylene-acetylene cis coupling step to C₄H₄ requiring an ensemble of at least 4 active metal atoms. Yang *et al.*²⁸ pursued the investigation further comparing three paths to 1,3-butadiene on plain and stepped surfaces, in presence of subsurface carbon, and of alloyed silver atoms; it emerged that, on a clean Pd(111) surface, acetylene-acetylene coupling followed by hydrogenation is the quickest way to butadiene, while the Pd(211) favours vinyl-vinyl coupling.

Chemical physics of the catalyst: hydrides and carbides

There is general agreement in the literature that the palladium catalyst undergoes structural transformations under the reaction conditions. The Pd/H system with its two pure hydride phases is well-known to the scientific community.²⁹ The first, hydrogen-poor phase is the α -Pd/H, which intakes a few percent of hydrogen without any significant modification of the face-centred cubic structure of palladium. The other phase is the β -Pd/H phase (>50% hydrogen). This phase has a sodium chloride crystalline structure and is hydrogen rich. A lot of work has been done on the Pd/H phase portrait. Anyway, this is a deceiving picture when discussing the catalytic process under consideration. In fact, those studies were led in an atmosphere of hydrogen only and do not take into account the effect of the reaction. The reaction of acetylene with hydrogen subtracts material to the Pd/H phase; as a result, the hydrogen pressure required to sustain the β -Pd/H phase at reaction conditions is ten times as huge as the one of ordinary hydrogen adsorption studies. The need for *in situ* studies of the catalyst thus became apparent. At the time of the aforementioned review, the year 2006, there was a few *in situ* studies of acetylene hydrogenation under C₂H₂/H₂ gas pressure, while the C₂H₄/H₂/C₂H₂ was tackled only once. Along with the hydride phases, formation of Pd/C phase was extensively proved. The carbon atoms, coming from the cleavage of the organic reactive species, occupy the octahedral sites of the palladium unit cell in their most stable configuration. It was suggested that a competition between the carbide and hydride phases takes place inside of the bulk material, and that thermodynamics should favour the former. Apart from that, even though the maximum carbon intake of palladium was found to be 11.5%, the catalytic surface at steady state conditions is always about 1.6%.

The picture above presented was going to change in a very profound way as new experimental techniques of higher sensitivity and precision shed light on the palladium catalytic system. In fact, XPS of the system under realistic conditions showed that the Pd/C phase enclosed into the first three subsurface layers was much richer than 1.6%.³⁰ A novel, fundamental role was then attributed to the Pd/C phase. Firstly, alloying carbon with the catalytic surface atoms exerts an inhibitory effect over all the hydrogenation steps of Figure 1, due to modification of the catalyst electronic structure. This effect is more pronounced for ethylene hydrogenation than for acetylene hydrogenation; therefore, selectivity towards ethylene is favoured.³¹ Moreover, since this effect is more significant for ethylene, this molecule tends to desorb faster from the surface, thus resulting in an improved selectivity.

²⁷ López, N.; Vargas-Fuentes, C. *Chem. Comm.* **2012**, 48 (10), 1379-1391

²⁸ Yang, B.; Burch, R.; Hardacre, C.; Hu, P.; Hughes, P. *J. Phys. Chem. C* **2014**, 118, 1560-1567

²⁹ Borodzinski, A.; Bond, G. C. *Catal. Rev.* **2006**, 48, 91-114

³⁰ Teschner, D.; Borsodi, J.; Wootsch, A.; Révay, Z.; Hävecker, M.; Knop-Gericke, A.; David Jackson, S.; Schlögl, R. *Science* **2008**, 320, 86-89

³¹ Torres, D.; Cinquini, F.; Sautet, P. *J. Phys. Chem. C* **2013**, 117, 11059-11065

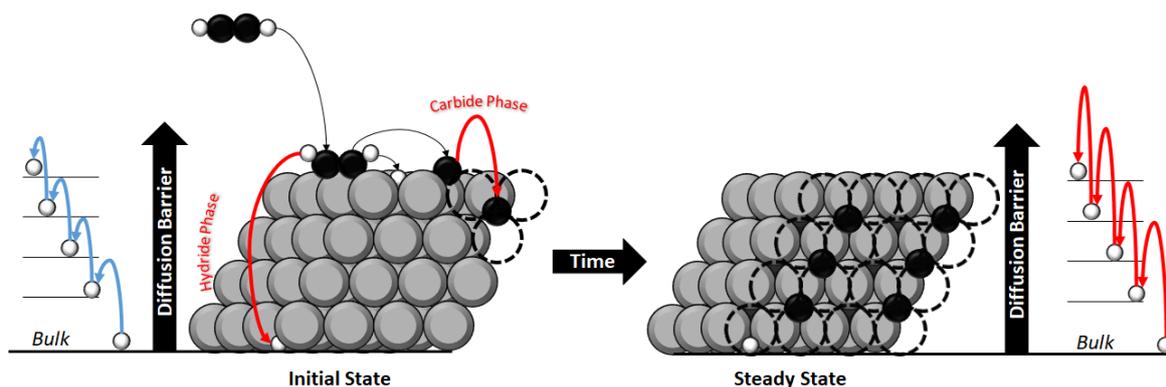


Figure 2. Pd/C phase and its effect on hydrogen diffusion

Selectivity is also ensured by another effect. It is in fact recognized that, even though bulk hydrogen can pass through the subsurface carburized layers, its diffusion is more hindered than in the pure catalyst,³² as shown in Figure 3. Since bulk hydrogen is unselective,³³ this diffusive obstacle prevents total hydrogenation to ethane.

Surface modifications of palladium to give a carburized phase is explained by the chemical potential of carbon as stored into acetylene and as an interstitial element of a Pd-C alloy. This concept is of course general and applies to other transition metals,³⁴ so that a trend is obtained spanning the d block of the periodic table. As it is well-known from the metallurgy of steel, iron shows great affinity towards carbon and even to formation of bulk alloys. Palladium is less prone than iron, and only generates a subsurface carbide phase. Explaining surface carburization is easy if the contribution to the carbon chemical potential as solute is decomposed into a bond term and a surface strain term. Noble metals such as copper, silver, and gold pay a very little energy penalty for structural deformation, yet their carbon subsurface chemical potential is indeed high due to weak interaction with the absorbed atoms; at the same time, early transition state metals give stronger bonds, but their distortion occurs at high price. Palladium is at the edge of these two extremes; its distortive and bonding contributions to the chemical potential into the mixed phase are not as significant as early or late transition metals and this leads to its subsurface (but not bulk) surface Pd-C phase.

Pd-Ag alloy catalysts for selective hydrogenation of acetylene

Though a still intriguing issue and a matter of considerable debate in the physical and in the chemical research community, the use of bimetallic systems and alloys for catalytic purposes has a long history. First attempts to summarize the early results obtained in this field are found in two excellent review papers from the 70s^{35,36}. Three general concepts emerge from the literature, which are reviewed shortly in the following.

Forming bonds between atoms is what chemistry is made of. Let us imagine to create such a bond decreasing the distance between two initially well-separated and chemically compatible atoms; if the electronic energy is plotted versus the distance, a minimum is observed at bond distance. The electronic configuration is also sensitive to this experiment; indeed, it changes as the atoms are moved

³² Teschner, D.; Borsodi, J.; Kis, Z.; Szentmiklósi, L.; Révay, Z.; Knop-Gericke, A.; R. Schlögl; Torres, D.; Sautet P. *J. Phys. Chem. C* **2010**, *114*, 2293-2299

³³ García-Mota, M.; Bridier, B.; Pérez-Ramírez, J.; López, N. *J. Catal.* **2010**, *273*, 92-102

³⁴ Sautet, P.; Cinquini, F. *ChemCatChem* **2010**, *2*, 636-639

³⁵ Clarke, J. K. A. *Chem. Rev.* **1975**, *75*, 291-305

³⁶ Sinfelt, J. H. *Acc. Chem. Res.* **1977**, *10*, 15-20

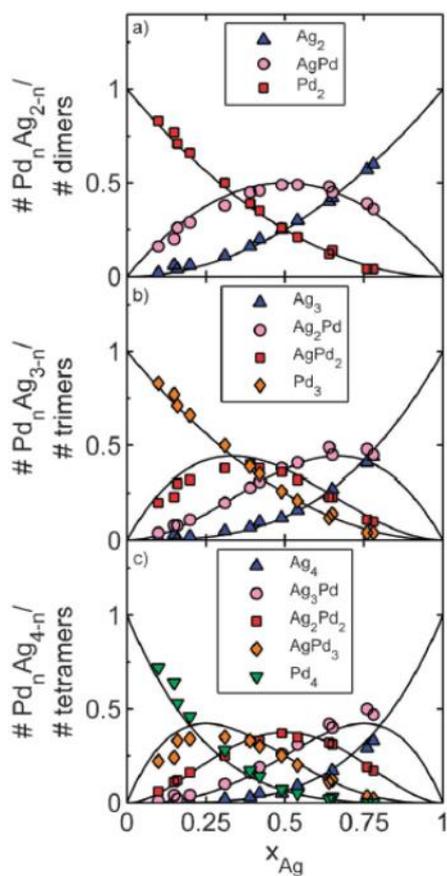


Figure 3. Ensembles concentration on the Pd-Ag (111) surface layer in UHV conditions as function of the Ag surface molar fraction (taken from ref. 42)

along the reaction path. The electronic configuration at bond distance and those at infinite separation are usually much different; due to constructive and destructive interference of their wavefunctions. Within a small approximation (neglecting higher-order effects), it can be said that a linear combination of two orbitals lying on different atoms generates two molecular orbitals placed at different energies. Now, if a third atomic orbital is added to the molecular configuration, again interactions of constructive and destructive nature takes place; theoretically, the process can be iterated at will to generate lots of electronic states, whose energetic spacing gets smaller and smaller so that a description of extended bonds in terms of *electronic bands*³⁷ is convenient.

The most important band when dealing with transition metals is the band formed by atomic d orbitals, also called the *d-band*. The d-band of a particular transition metal allows or forbids adsorption of molecules, enhances catalysis of given processes, etc. A question follows naturally: if Pd and Ag have their respective d-bands, what a band results by their mixing? This question is far from being trivial. Two scenarios can be imagined. The resulting d-

band could be an average of the two pure-metal bands, maybe weighted by composition; thus the resulting band is a linear combination of the two bands. Alternatively, a non-linear evolution could follow mixing, exalting some aspects and depressing the others. All of this is crucial to correctly explain why alloying Ag with Pd improves selectivity towards hydrogenation and suppresses oligomer formation; since Ag is indeed not prone to accommodate acetylene molecules, it is very likely that all adsorption events are located on Pd atoms. Now, if after alloying the electronic d-band structure of Pd gets as much modified as to lower the coverage of the surface, then the catalyst is selective no matter what the disposition of atoms on the surface is and selectivity is basically an *electronic effect*. Else, the d-band is left unaltered, which means that Ag acts simply as a spacer, hindering collisions of molecules on the surface; in this case, selectivity is purely an *ensemble effect*. Experimental evidence has been brought in favour of the ensemble effect,^{38,39} also corroborated by calculations.⁴⁰ In any event, neither experiments were led at the industrial scale, nor *in situ* studies were performed in the laboratory to evaluate adsorbate-induced restructuring of Pd-Ag catalysts.

³⁷ Hoffmann, R. *Angew. Chem. Int. Ed. Engl.* **1987**, *26*, 846-878

³⁸ Soma-Noto, Y.; Sachtler, W. M. H. *J. Catal.* **1974**, *32*, 315-324

³⁹ Ma, Y.; Diement, T.; Bansmann, J.; Behm, R. *J. Phys. Chem. Chem. Phys.* **2011**, *13*, 10741-10754

⁴⁰ Mancera, L. A.; Jürgen Behm, R.; Gross, A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1497-1508

The ensemble effect in alloys is best visualized by their *ensemble description*.⁴¹ This description defines first an ensemble as an arrangement of sites in a given geometry, e.g. triangular, square, pentagonal, etc. ensembles. Let us now consider a binary Pd-Ag alloy; if the ensemble is made of n atoms, its composition ranges from the pure Pd _{n} to the pure Ag _{n} ensemble. If the alloy topmost layer – the essential layer for catalytic purposes – is randomized, its ensembles should also be distributed in a random fashion. Building these random distributions is easy; they are Bernoulli distributions whose probability is given by the layer composition. For an ensemble of composition Pd _{$n-i$} Ag _{i} , the probability is $p_i = \frac{n!}{i!(n-i)!} \times \varphi^i (1 - \varphi)^{n-i}$. Engstfeld *et al.*⁴² studied distributions of triangular ensembles on the topmost layer of the Pd-Ag (111) surface plane; their STM study proved that the surface is randomized in ultrahigh vacuum conditions, see Figure 3. Operative conditions were not explored, unfortunately. The distributions of ensembles provide hints to appraise the ensemble effect quantitatively; predominance of ensembles of mixed composition implies dilution of Pd atoms, which impedes coupling reactions.

As just seen a few lines above, the Ag layer molar fraction φ is a principal quantity in the ensemble description of alloys and will be employed in the present work as a catalyst control parameter. This means that the stability of catalyst phases are scanned via systematic increases of φ , keeping other control parameters constant, i.e. the pressure and the temperature. Now, let us consider again the pure constituents of the Pd-Ag alloy and imagine to cut a surface model out of their bulks. This operation will certainly need some energy to be performed (you do not get things for free in chemistry), which will in general differ from material to material. The difference in the surface formation energy between the alloyed metals is responsible for a phenomenon known as *surface segregation*: the metal showing the least surface formation energy tends to accumulate at the top of the surface. This is the case of Pd-Ag alloys, as proven by the segregation profiles obtained by Shu *et al.*⁴³ The picture gets even more complicated by the reversal of segregation operated by adsorbates,^{44,45} which extract Pd atoms from the subsurface. In fact, adsorption events have free energy contributions, whose total energy parts are much higher when adsorbates lie on active atoms; therefore, adsorbates show a tendency to pull active atoms out of the subsurface and maximize the energy loss.

Statement of Purpose

The precedent Sections have emphasized the role of two crucial aspects in the selective hydrogenation of acetylene, namely the complexity of the reaction mechanism, which spans from decomposition of the substrate to single carbon atoms to its polymerization up to C₆₀ species, and the impact of the catalyst's physics on the reaction selectivity towards ethylene. The design of experiments to understand both these aspects is difficult, and computational studies have often played an important role in this respect. While a basic understanding of the catalyst's phases under realistic conditions has been achieved, much is left to the comprehension of Pd-Ag alloys; as for the formation of green oil, analyses are sparse and controversial, while theory has not been prompted much further than acetylene and vinyl homo- and hetero-couplings.

The following chapters address these two points of inquiry. In particular, the thermodynamics of Pd-Ag alloys under hydrogenation conditions is explored to assess the topology of the catalyst at

⁴¹ Ponc, V.; Bond, G. C. *Catalysis by Metals and Alloys*; Elsevier; Amsterdam; 1995

⁴² Engstfeld, A. K.; Hosterw, H. E.; Behm, R. J. *Phys. Chem. Chem. Phys.*, **2012**, *14*, 10754–10761

⁴³ Shu, J.; Bongondo, B.E.W.; Grandjean, B.P.A.; Adnot, A.; Kaliaguine S. *Surf. Sci.* **1993**, *291*, 129-138

⁴⁴ Svenum, I.-H.; Herron, J.A.; Mavrikakis, M.; Venvik, H.J. *Catal. Tod.* **2012**, *193*, 111– 119

⁴⁵ Padama, A. A. B.; Kasai, H.; Budhi, Y. H. *Int. J. Hydrogen Energ.* **2013**, *38*, 14715-14724

high pressures of acetylene, the distribution of ensembles, and the effect of segregation. Then, early stages of green oil formation are investigated from the point of view of chemical kinetics to understand the relative importance of hydrogenation and oligomerization steps in the overall mechanism.

Chapter 2. Methodologies

The ensembles chemical equilibrium

Ensemble description of acetylene adsorbed on Pd-Ag alloys

In Chapter 1, Section 1.2.2, the concept of ensemble was introduced as a way to quantify the degree of randomness showed by an alloyed surface layer. It turns out that this measure is not only of use in surface science, but describes the interaction of reactants with catalytic surfaces as well.⁴⁶ This is particularly fruitful in the case of Pd-Ag alloys. A TPD study by Ma *et al.*⁴⁷ is a clear example of this. In that study, thin surface layers of Pd-Ag (111) were prepared in different Pd/Ag ratio, so as to scan the whole compositional range. CO was allowed to flow and adsorb on surfaces, with the temperature increasing of 4 K per second. CO shows three adsorption modes on pure Pd (111), namely linear (one molecule interacts with one metallic atom), bridge (interaction with two metallic atoms), and three-fold bridge (interaction with three metallic atoms arranged in a triangle). These adsorption features are captured by IR spectroscopy,⁴⁸ and recovered in TPD experiments. It is noteworthy that alloying Ag does not significantly alter the profiles of TDP measurements, but only their intensities; this was explained claiming that the electronic effect is not significant in this system, which is also observed by computational studies of Sheth *et al.*⁴⁹ and Mancera *et al.*⁵⁰ After 75% Ag, the high-temperature peak, attributed to three-fold bridge adsorption, is suppressed, which has been proposed as evidence for disappearance of triangular ensembles of Pd atoms. Single Pd atoms are expected to populate the surface at higher Ag molar fractions. The TPD study of Ma *et al.* is of course of great value but misses several points of industrial relevance. Industrial plants need prediction over long periods of time and huge pressures of adsorbate; model surfaces should also be considered as approximations for nanoparticles with defects and of different size. One issue of special importance to TOTAL, as an example, is green oil formation on an Ag-rich catalyst: if large ensembles are suppressed, how is it possible for oligomers to appear on the surface? There simply could not be room enough for it. An answer to this problem might come from statistical mechanics.

Acetylene adsorbs on triangular sites of Pd (111) preferentially.⁵¹ Analogously to CO, it exhibits a variety of adsorption modes: it presents itself as di- σ bonded (interacting with three metallic atoms) or π -bonded (interaction with only one metallic atom). The former is definitely the most stable mode, with an adsorption energy estimated by computational studies to be around -1.90 eV per molecule (the value might fluctuate due to lateral interactions or on the choice of functional). This interaction is much greater of the negligible interaction between Pd and Ag atoms.⁵² On the basis of what has been said, mixing Ag to the catalyst is expected to depress acetylene adsorption. Let us perform an ideal experiment on two triangular ensembles of Pd-Ag (111) layer. A molecule of acetylene is posited in its equilibrium geometry, while the other ensemble stays empty. All of the 6 metallic atoms are Pd but one. Assume the covered ensemble contains Ag; if the two ensembles are allowed to exchange matter via some mechanism, the most stable configurations will be the ones where Ag is transferred to the empty site. Since at operating conditions the adsorption energy on a Pd-pure ensemble is much

⁴⁶ Ponc, V.; Bond, G. C. *Catalysis by Metals and Alloys*; Elsevier; Amsterdam; 1995

⁴⁷ Ma, Y.; Diemant, T.; Bansmann, J.; Behm, R. J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10741-10754

⁴⁸ Soma-Noto, Y.; Sachtler, W. M. J. *J. Catal.* **1974**, *32*, 315-324

⁴⁹ Sheth, P.; Neurock, M.; Smith, C. M. *J. Phys. Chem. B* **2005**, *109*, 12449-12466

⁵⁰ Mancera, L. A.; Behm, R. J.; Gross, A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1497-1508

⁵¹ Dunphy, J. C.; Rose, M.; Behler, S.; Ogletree, D. F.; Salmeron, M.; Sautet, P. *Phys. Rev. B* **1998**, *57* (20), R12705-12708

⁵² Engstfeld, A. K.; Hosterw, H. E.; Behm, R. J. *Phys. Chem. Chem. Phys.*, **2012**, *14*, 10754-10761

greater than the thermal quantum, induced-segregation of a Pd particle should be the most stable thermodynamics state. This is precisely what Chapters 3 and 4 are going to show.

Pd-Ag triangular ensembles as species in chemical equilibrium

The ensemble picture of alloys is a conceptual framework conceived by experimentalists to describe an alloyed catalyst topmost layer. In order to characterize layer thermodynamics, its ensembles distributions must be known as function of external constraints such as the temperature, the pressure of the gaseous reservoir, and the molar fraction of Ag (or Pd, for what matters) contained into it. This is a situation much close to the one scientists meet in statistical mechanics. Statistical mechanics predicts averages of physical quantities if the so called partition function is known.⁵³ The partition sum is a sum of exponentials over all the possible states of the system. In the case of the surface topmost layer in equilibrium with a gaseous reservoir, is written in the form of equation 1

$$Q = \sum_{i=1}^N \sum_{s=1}^S e^{\frac{\mu_i - H(s)}{k_B T}} \quad (1)$$

Here, T stands for the temperature, k_B for the Boltzmann constant, N for the number of adsorption sites on the layer, S for the number of states available to the surface layer, i.e. the number of possible non-identical configurations obtained by continuously exchanging positions of metal atoms. The symbol μ stands for the chemical potential of acetylene in the gas phase; under the hypothesis of ideal gas, statistical mechanics provides an easy formula for it (equation 2)

$$\mu = \varepsilon_{C_2H_2} - k_B T \ln \left(\frac{k_B T}{h^3} \times \sqrt{(2\pi M k_B T)^3} \right) - k_B T \ln \left(\frac{4\pi^2 I k_B T}{h^2} \right) - k_B T \ln \left(\frac{p}{p_0} \right) \quad (2)$$

The first term is the formation energy of the molecule, h is the Planck constant, M and I are the molecule's mass and moment of inertia, respectively, p is the gas pressure with reference to the pressure p_0 . One term of equation 1 remains unexplained, the configurational energy H . While all other terms are straightforward, the configurational energy is hopelessly intricate. The major challenge is to find a recursive and compressible procedure to evaluate the energy of all the configurations. This is a very complicated and still unsolved problem in statistical mechanics (at least in general);⁵⁴ sampling procedures for all practical purposes realized by numerical recipes (e.g. Monte Carlo sampling). In any event, one can draw some estimates with approximate schemes, such as mean-field theory. An approach close to the spirit of mean field theory was indeed developed to explore the ensemble populations of topmost layers by Vignola *et al.*⁵⁵ and is discussed in Chapter 3.

The approximation made by Vignola *et al.* considers all the triangular ensemble as independent adsorption sites, as shown in the lateral view of Figure 9. Each ensemble spans the compositional range from the pure Pd₃ to the pure Ag₃ site, stepping through the Pd₂Ag and the PdAg₂

⁵³ Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications; New York; 1960

⁵⁴ Mussardo, G. *Statistical Field Theory. An Introduction to Exactly Solved Models in Statistical Physics*; Oxford University Press; Oxford; 2010

⁵⁵ Vignola, E.; Steinmann, S. N.; Vandegehuchte, B. D.; Curulla, D.; Sautet, P. *J. Phys. Chem. C* **2016**, *120* (46), 26320-26327

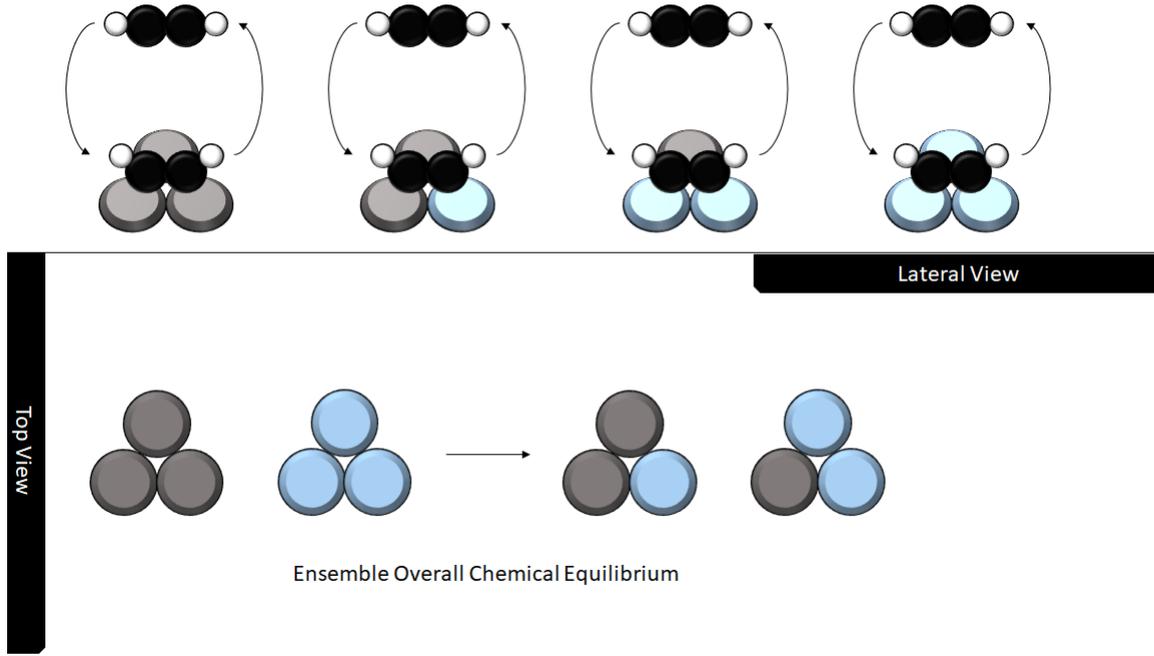


Figure 4. Lateral view of top layer ensembles illustrating the adsorption chemical equilibrium; top view of ensembles showing the overall chemical equilibrium

configuration. Now, the following positions are made: the total population of Pd_3 ensembles in the layer is labelled with x_0 ; the total population of Pd_2Ag ensembles in the layer is labelled with x_1 ; the total population of $PdAg_2$ ensembles in the layer is labelled with x_2 ; the total population of Ag_3 ensemble in the layer is labelled with x_3 . In general, the variable x_a corresponds to the $Pd_{3-a}Ag_a$ ensemble population. For each of these variables, another variable θ_a is added to represent the covered fraction relative to ensemble $Pd_{3-a}Ag_a$. Now, since all adsorption sites are independent, the adsorption isotherm of a given layer state may be cast in the form of equation 3

$$\theta_a = x_a \frac{e^{\frac{\mu - \varepsilon_a}{k_B T}}}{1 + e^{\frac{\mu - \varepsilon_a}{k_B T}}} \quad (3)$$

Here, the adsorption energy on site type $Pd_{3-a}Ag_a$ was labelled with ε_a . This provides a first set of four relationship between pairs of unknowns. Another constraint comes from the chemical equilibrium sustained by uncovered sites. Provided there is a mechanism equilibrating the layer (by diffusion of matter, for instance), the system reaches an equilibrium state described by the law of mass action: the ratio of uncovered sites partition functions equals to an equilibrium constant. These partition functions are in the form of sums of Boltzmann factors over single ensemble states. Since Pd and Ag do not interact in a significant way, the energy terms into the exponentials can be approximated to zero; thus, partition functions of uncovered sites are merely the number of states associated to them. Pure ensembles only have one state, while mixed ensembles terms have a multiplying pre-factor 3 due to multiplicity, i.e. the impurity can appear on one out of three sites. Now, the ensemble overall chemical equilibrium reaction is $Pd_3 + Ag_3 \rightarrow Pd_2Ag + PdAg_2$ (see top view of Figure 4 for reference); the law of mass action in this case reads as equation (4)

$$\frac{(x_1 - \theta_1) \times (x_2 - \theta_2)}{(x_0 - \theta_0) \times (x_3 - \theta_3)} = \frac{3 \times e^{-\frac{\varepsilon_1}{k_B T}} \times 3 \times e^{-\frac{\varepsilon_2}{k_B T}}}{e^{-\frac{\varepsilon_0}{k_B T}} \times e^{-\frac{\varepsilon_3}{k_B T}}} = 9 \quad (4)$$

The quotient $e^{\varepsilon_0 + \varepsilon_3 - \varepsilon_1 - \varepsilon_2}$ was put to 1. An additional constraint comes from mass balance, equation 5

$$\begin{cases} x_0 + \frac{2x_1}{3} + \frac{x_2}{3} = 1 - \varphi \\ \frac{x_1}{3} + \frac{2x_2}{3} + x_3 = \varphi \end{cases} \quad (5)$$

In equation 5, φ is the molar fraction of Ag in the layer. This makes up to 7 equations in 8 unknowns. The last constraint is obtained resorting to an *ad hoc* assumption: the distributions of ensembles inside of the uncovered, random part of the layer are Bernoulli distributed. This is mathematically formulated in the terms of equation 6

$$\begin{cases} y_0 = \left[x_0 - \theta_0 + \frac{2(x_1 - \theta_1)}{3} + \frac{(x_2 - \theta_2)}{3} \right]^3 - x_0 + \theta_0 = 0 \\ y_1 = 3 \left[x_0 - \theta_0 + \frac{2(x_1 - \theta_1)}{3} + \frac{(x_2 - \theta_2)}{3} \right]^2 \left[\frac{(x_1 - \theta_1)}{3} + \frac{2(x_2 - \theta_2)}{3} + x_3 - \theta_3 \right] - x_1 + \theta_1 = 0 \\ y_2 = 3 \left[x_0 - \theta_0 + \frac{2(x_1 - \theta_1)}{3} + \frac{(x_2 - \theta_2)}{3} \right] \left[\frac{(x_1 - \theta_1)}{3} + \frac{2(x_2 - \theta_2)}{3} + x_3 - \theta_3 \right]^2 - x_2 + \theta_2 = 0 \\ y_3 = \left[\frac{(x_1 - \theta_1)}{3} + \frac{2(x_2 - \theta_2)}{3} + x_3 - \theta_3 \right]^3 - x_3 + \theta_3 = 0 \end{cases} \quad (6)$$

The system of equations so-obtained is heavily non-linear and is solved by numerical methods.

Under the assumptions of Vignola *et al.* the configurational energy is written as a linear combination of ensemble terms, namely

$$H(s) = \sum_{a=0}^3 n_a(s) \varepsilon_a \quad (7)$$

In equation 7, $n_a(s)$ denotes the number of sites of composition $\text{Pd}_{3-a}\text{Ag}_a$ in state s . This expression corresponds to the energy of an “ideal gas” of ensembles in a given state s . In order to improve the model as to describe interactions among ensembles, an interaction term has been added to equation 7 resulting into equation 8

$$H(s) = \sum_{a=0}^3 n_a(s) \varepsilon_a + f(N) \quad (8)$$

The function $f(N)$ represents the averaged behaviour of interactions; as explained in more detail in Chapter 3, it was modelled by interactions on a pure Pd surface at different acetylenic coverages. This choice was motivated by the fact that acetylene is expected to adsorb on Pd in the interesting range of thermodynamic parameters. Adsorption and interaction energies, and the interaction function are all tabulated and discussed in Chapter 3.

Cluster expansions

The method described in the previous Section is computationally cheap and allows for quick tests of intuitions. In the present context, as an example, it was employed to assess the ability of the reaction’s main reactant to induce reconstruction of the topmost catalytic layer. As the reader finds out in Chapter 3, this was the prediction at low temperatures (say, under 600 K at 1 bar of acetylenic pressure). The competitiveness of the method comes at the sacrifice of precision and accuracy, with many of the system’s features lost in the simplification. For instance, the description of lateral interactions as in equation 8 has clearly room for improvement and should be described as a function of the adsorbates positions; moreover, the limitation of the precedent model to triangular ensembles neglects the propagation of interactions through them. One should therefore improve his

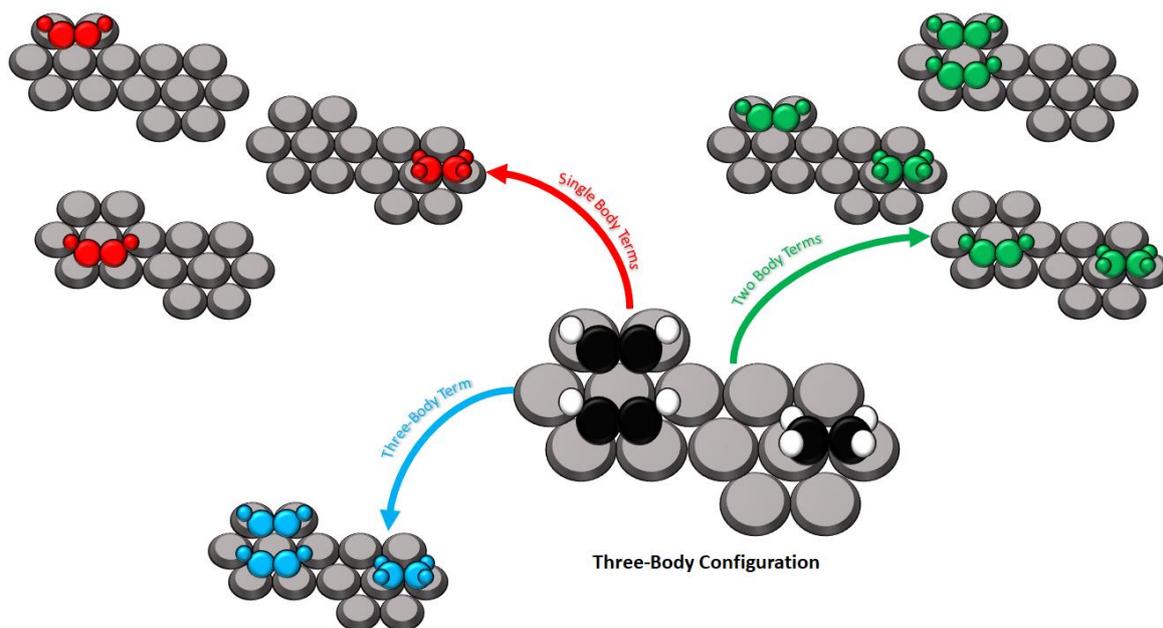


Figure 5. Cluster expansion of a three-body configuration

methods and models so as to cast prediction of a more rigorous level. In Chapter 4, a detailed surface model is constructed that fits *ab initio* data. The technique at the core of our treatment is cluster expansion; even though there are outstanding reviews on the subject (noteworthy are De Fontaine's⁵⁶ and Inden's⁵⁷ book chapters), the essential concepts are summarized in the following to provide the reader with the minimal background.

Firstly, let us consider a system of points $V = \{v_1, v_2, \dots, v_{|V|}\}$, a set of spin-like variables $S = \{s_1, s_2, \dots, s_{|V|}\}$, and a set of spin-like values $M = \{m_1, m_2, \dots, m_{|M|}\}$. Let us then assign a spin-like variable to each of the points in V via a one-to-one mapping, say, $\omega_1 : V \rightarrow S$ and also propose a second mapping $\omega_2 : S \rightarrow M$ so as to guarantee that a spin-value is assigned to each variable. If the set M is furnished with the usual operation of multiplication and with the operation of sum modulo $|M|$, not only it is possible to describe each configuration as a $|V|$ -tuple in a $|V|$ -dimensional space, but the vector algebra formalism can also be applied to this space (formally, a vector space has been established on the ring structure of M).⁵⁸ Scalar and vector functions of the space coordinates in S can be proposed as functions of configurations; among these, there is the Hamiltonian function $H(\vec{s})$ discussed in Section 2.1 (mind that the state is now symbolized as a vector consistently).

Now, it is known that, under rather mild conditions, state functions can be expanded on a basis of orthonormal functions.⁵⁹ Though there is freedom to choose whatever set of functions, the much popular "SDG" set of Sanchez *et al.*⁶⁰ is usually discussed because its orthogonality and completeness are easily proved. The SDG set is made of Chebyshev polynomials as orthogonalized and normalized by the Gram-Schmidt procedure. Mind that this was not the choice of Chapter 4,

⁵⁶ De Fontaine, D. "Cluster Approach to Order-Disorder Transformation in Alloys" in *Solid State Physics vol. 47* (eds. H. Ehrenreich and D. Turnbull); Academic Press; New York; 1994

⁵⁷ Inden, G. "Atomic Ordering," in *Phase Transformations in Materials* (ed. G. Kostorz); Wiley-VCH; Weinheim; 2001

⁵⁸ Biggs, N. L. *Interaction Models*; Cambridge University Press; New York; 1977

⁵⁹ Book on Fourier series or review paper

⁶⁰ Sanchez, J.-M.; Ducastelle, F.; Gratias, D. *Physica A*, **1984**, 128 (1), 334-350

where the basis of Stamatakis and Vlachos was used instead;⁶¹ this basis is a set of delta functions of molecular clusters appearing on the surface, i.e. they spike to 1 if a given molecular arrangement is detected. The use of orthonormal functions allows for an expansion in the form of equation 9

$$H(\vec{s}) = \sum_{i=1}^{|M|^{|V|}} \varepsilon_i \varphi_i(\vec{s}) \quad (9)$$

Here, the function $\varphi_i(\vec{s})$ represents a general cluster function. The completeness of this expansion must be remarked: as it is, the Hamiltonian has a complete lattice representation. For practical purposes, however, cluster expansion need to be truncated to a given order. As an example, the cluster expansions discussed in Chapters 4 and 6 are truncated to second order (pair interactions). An example of this is displayed in Figure 5.

The chemically-oriented reader might find the formalism above a bit confusing, but there is nothing to be afraid of, for such formalism is at the basis of many spectroscopic analyses in the realm of organic and inorganic chemistry. Basically, all the coefficients ε_i of equation 9 are interactions not very dissimilar from coupling constants encountered in NMR spectroscopy. In the present context, however, the fine structure of the surface layer cannot be obtained by pulsing the model with radio waves but must be acquired with *ab initio* calculations instead. A set of DFT calculations is therefore generated so as to span part of the configurational space; if the calculated values of the total energy are listed for each of the sampled configurations, a linear system of equations of the form of 10 is obtained

$$\hat{\varphi} \cdot \vec{\varepsilon} = \vec{H} \quad (10)$$

The matrix $\hat{\varphi}$ is a square matrix made of vectors whose elements are orthonormal functions as calculated at the given configuration. As shown by Connolly and Williams for the first time,⁶² this system may be inverted to give the cluster coefficients

$$\vec{\varepsilon} = \hat{\varphi}^{-1} \cdot \vec{H} \quad (11)$$

The total energy of the system can then be calculated at will for an arbitrary configuration, without recurring to *ab initio* methods any further.

As it will be noted in later Chapters, sampling of the configurational space often requires a greater amount of configurations to specify cluster coefficient in full; in fact, the truncation of the system of equations 10 introduces uncertainties in the coefficients. This means that the matrix $\hat{\varphi}$ will be a rectangular matrix for all practical applications. This raises a question about its invertibility; in fact, there is no way to invert a rectangular matrix. An inversion scheme is nevertheless available if the definition of inversion is intended in the sense of the least squares. That is to say, the unknown vector is interpolated to the available data. The inversion operation, known as the *Moore-Penrose pseudoinversion*,⁶³ keeps the formalism of equation 11; it only needs to add products involving the matrix itself and its conjugate $\hat{\varphi}^+$

$$\vec{\varepsilon} = (\hat{\varphi}^+ \cdot \hat{\varphi})^{-1} \cdot \hat{\varphi}^+ \cdot \vec{H} \quad (12)$$

Numerical problems associated to this operation are discussed in Chapter 6 in depth, as well as the necessary techniques to ensure the statistical validity of models thereof.

⁶¹ Stamatakis, M.; Vlachos, D. *J. Chem. Phys.* **2011**, *134* (21), 214115

⁶² Connolly, J. W. D.; Williams, A. R. *Phys. Rev. B*, **1983**, *27*, 5168

⁶³ Barata, J.; Hussein, M. S. *Braz. J. Phys.*, **2012**, *42*, 146-165

Monte Carlo methods

As stated in the precedent Section, it is possible to represent Hamiltonians with reliable lattice model. It might seem that the most critical and time consuming goal was achieved: the Hamiltonian is easily connected to the partition function, which directly give the free energy or other thermodynamic potentials. Caution is suggested before getting to such a conclusion; in fact, a subtle problem is nested inside of the vector \vec{s} when calculating the partition function. The problem is stated in very simple terms: to get thermodynamically meaningful results, the cardinality $|V|$ of sets V and S must be large enough. Let us assume for the sake of the argument that there are only two spin-like values in the set M – this is the case of a two-metal system like pure Pd, pure Ag, and their alloys; this means that the full configurational space spans $2^{|V|}$ configurations, in other words, that the configurational space increases exponentially with the number of points in the model – the general formula is $|M|^{|V|}$. This configurational conundrum is way too much to be sampled for realistic systems; interactions in the system were compressed into the model Hamiltonian, but not the states themselves. This poses the additional challenge to sample the remaining states in an intelligent, time-saving way. Such a method is provided by Monte Carlo algorithms,⁶⁴ which sample the most relevant part of the configurational space.

The general mechanism of a Monte Carlo is of striking simplicity. With reference to equation 2, the partition sum starts with an arbitrary state $\vec{s} = \vec{m}_1$. Then, a transition to another state \vec{m}_2 is evaluated in the following way: a number is picked randomly from a set $[0,1]$; the values $p_1 = e^{-\frac{\mu i_1 - H(m_1)}{k_B T}}$ and $p_2 = e^{-\frac{\mu i_2 - H(m_2)}{k_B T}}$ are calculated; if the ratio p_2/p_1 is smaller than the random number picked up, the transition is forbidden and the state vector rests in its initial position; else, it gets accepted. This procedure is iterated for a large enough number of steps, at each of which thermodynamic averages are accumulated. This statement implicitly shows the power of the Monte Carlo method; calculation of the partition sum is bypassed through the evaluation of relative weights, leading to averages in a faster way. Apart from unlikely cases (fluctuations), the algorithm sticks close to the most influential points in the space.

Though it is not immediately referred to the present work, a few words about a special kind of Monte Carlo algorithm could be helpful in view of further investigation. The algorithm described above refers to calculation of equilibrium averages but does not tell much about the mechanism by which the equilibrium state is reached. It turns out that Monte Carlo algorithms can be modified to also explore kinetics of diffusion and chemical reactions.^{65,66} The so-called *kinetic Monte Carlo method* has offered a way to simulate chemical systems at large timescales, even hours; it has been applied to several systems of interest in catalysis with noteworthy results (see Chapter 6 and references within). While intelligent sampling stays at its core, states and transitions in kinetic Monte Carlo algorithms are redefined into the framework of potential energy surfaces and transition state theory. As known from quantum chemistry,⁶⁷ it is possible to describe the potential exerted on nuclei by electrons separating degrees of freedom adiabatically; this separation generates a function of the nuclear coordinates, the so-called *potential energy surface*. This surface is a scalar field in the nuclear coordinates; if its gradient is posed equal to zero, its critical points are isolated. There are many types

⁶⁴ Frenkel, D.; Smit, B. *Understanding Molecular Simulations*; Academic Press; San Diego (CA); 1996

⁶⁵ Jansen, A. P. J. *An Introduction to Kinetic Monte Carlo Simulations of Surface Reactions*; Springer-Verlag; Berlin; 2012

⁶⁶ Darby, M. T.; Piccinin, S.; Stamatakis, M. "First Principles-Based Kinetic Monte Carlo Simulation in Catalysis" in *Physics of Surface, Interface and Cluster Catalysis* (eds. H. Kasal and M. C. S. Escano); IOP Publishing; Bristol; 2016

⁶⁷ Levine, I. N. *Quantum Chemistry*, 5th Edition; Prentice Hall; Upper Saddle River (NJ); 2000

of critical points; among them, the most important are minima and saddle points. It is possible to trace a boundary along minima encapsulating the minimum and its basin of attraction. These domains, belonging to different minima, are connected via saddle points, also known as *transition states*.⁶⁸ Now, let us assume that, when following a path connecting two minima, the flux is kept unidirectional (this means that there is no means to come back to the starting minima if transition states are reached); with the additional hypothesis that equilibrium has been achieved between reacting minima and transition states, rate constants are written into an exceedingly simple form

$$k = \frac{k_B T}{h} e^{-\frac{\Delta H^\ddagger}{k_B T}} \quad (13)$$

The symbol h in equation 13 stands for Planck's constant, while ΔH^\ddagger is the energy difference between the transition state and the reactant state. A kinetic partition sum of the form

$$Q_K = \sum_i e^{-\frac{\Delta H_i^\ddagger}{k_B T}} \quad (14)$$

is then defined as in equation 14, where i labels all the possible transitions. The probability of a given transition is thus given by the ratio $p_i = e^{-\frac{\Delta H_i^\ddagger}{k_B T}} / Q_K$; the Monte Carlo routine can then be performed in the same way as described above. An event from the list of possible transitions is randomly picked at each time step along with a random number in the $[0, 1]$ interval; the transition is accepted if the event's probability is greater than that.

A first skirmish with Chapter 6: machine learning and adsorbed layers

The role of the lattice of points in our Monte Carlo recipes could not be overestimated. In a previous section, the lattice was just tentatively described as a collection of points V associated to a set of variables S and constants M . However, a lattice of points would just be an abstract construct without an embedding in the three dimensions of real space. To produce such an embedding, let $P = \{\vec{p}_1, \vec{p}_2, \dots, \vec{p}_{|V|}\}$ be the set of Cartesian coordinates of the lattice points, with an application $\omega_3 : V \rightarrow P$ ensuring the needed correspondences. Now, the set P stands for a lattice representation of a more realistic process – a non-parametrized, full *ab initio* simulation, or even an experiment taking place in the laboratory. Now, the standard approach to lattice model generation of real process is to arbitrarily abstract molecular entities, such as the modes of acetylene on a Pd surface, removing all the features due to the molecular lattice gas deformation. This arbitrariness comes at the price of model precision and accuracy; in fact, it would be desirable to automate lattice model construction so as to remove any human bias. This is essentially the most original contribution of Chapter 6, where the automatic generation of cluster expansion Hamiltonians is established on a rigorous basis.

The main idea of Chapter 6 is to take the overlap of molecular projections onto the surface and lattice points. Molecular projections are domains in the surface plane enclosed into a circular boundary of radius W . Denoting with r_j the domain of projection j and with d_i the domain of lattice site i , i.e. its Wigner-Seitz cell, the overlap between the two domains is written as in equation 15

$$c_{ij} = \frac{\int d_i(\{l_D\}_i, \vec{x}) \cdot r_j(\{l_R\}_j, \vec{x}) \cdot d\vec{x}}{\int r_j(\{l_R\}_j, \vec{x}) \cdot d\vec{x}} \quad (15)$$

Equation 15 also contains cell parameters (the terms in curly brackets). In the case of periodic lattices, this reduces to the normalized convolution between the two functions. In Chapter 6, molecular domains were built resorting to their atomic spheres; different sets of parameters were tested, such as van der Waals radii, covalent radii, and combinations of the two.

⁶⁸ Hänggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Phys.* **1990**, 62 (2), 251-341

Overlaps are an indication of site recognition as operated by adsorbed molecules but, things standing as they are, the problem would be left undecided; to make an effective decision, a recognition criterion must be adopted. Following machine learning theory,⁶⁹ a decision function was selected, that is, a Heaviside function of the overlap with origin in 0.5; whenever the overlap is greater or equal than 0.5, lattice point represents the molecule on the surface. Lattice configurations are then automatically calculated along with cluster functions, which is enough to perform the Connolly-Williams inversion.

⁶⁹ Bishop, C. *Pattern Recognition and Machine Learning*; Springer; Singapore; 2006

Chapter 3. C₂H₂-Induced Surface Restructuring of Pd-Ag Catalysts: Insights from Theoretical Modelling

Published on the Journal of Physical Chemistry C 2016, 120 (46), 26320

DOI: [10.1021/acs.jpcc.6b08524](https://doi.org/10.1021/acs.jpcc.6b08524)

Publication Date (Web): October 31, 2016

Copyright © American Chemical Society

Chapter 4. A Monte Carlo Study of Acetylene Adsorption on Pd-Ag Alloys

Emanuele Vignola,^{†‡} Stephan N. Steinmann,[†] Ahmad Al Farra,[‡] Bart D. Vandegehuchte,[§] Daniel Curulla[§] and Philippe Sautet^{†&*}

[†]Univ Lyon, ENS de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UM R 5182, F-69342, Lyon, France

[‡]Total Research and Technology Gonfreville, BP 27, F-76700 Harfleur, France

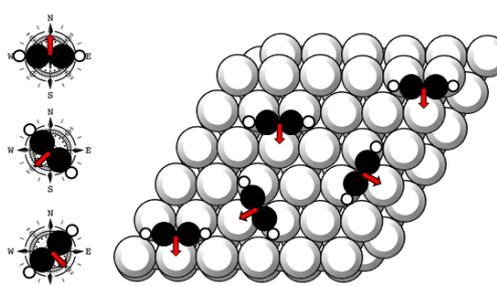
[§]Total Research and Technology Feluy, Zone Industrielle Feluy C, Seneffe, Belgium

[&]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, CA 90095, United States

*e-mail: sautet@ucla.edu; tel.: 1-310-825-8485

To be submitted to the Journal of Physical Chemistry C

ABSTRACT: Restructuring of alloy surfaces induced by strongly bound adsorbates is nowadays a well-established phenomenon occurring in catalysis and membrane science. Its relevance in catalytic processes is crucial, for it alters the ensemble distribution and cause unwanted side reactions depressing reaction yields and inactivating catalysts. This work assesses the restructuring of Pd-Ag alloys induced by adsorption of acetylene in the framework of the ensemble formalism. A detailed model of the (111) surface plane energetics is acquired at the Density Functional level of theory, which is then simulated with a Monte Carlo approach to evaluate averages as functions of thermodynamic constraints.



Introduction

Achieving novel properties by mixing of different substances is certainly not a revolutionary concept in chemistry. Solid solutions are particularly attractive from this point of view; the hardening of metals due to impurities has great technological impact – think, for instance, of the effect carbon and second metals have on steel¹ or the entropic hardening in highly-disordered materials.^{2,3} The thermodynamic description of alloys, as described in phase portraits, is crucial in this respect; predicting the evolution of a two-component solid system given a set of macroscopic constraints allows for forecast of failure in materials, assessment of its thermal stability, prevention of components segregation.

Control of such features is particularly important at the surface of the materials, where catalysis and membrane science occur. In this respect, building reliable models of the system's energetics is an essential aspect, which is best discussed by examples. Let us consider, for instance, the study of Pt-Pd nano-alloys by De Clercq and Mottet.⁴ Here, a tight-binding, semi-empirical potential was fitted to DFT data and provided as input of a Monte Carlo algorithm. Along with

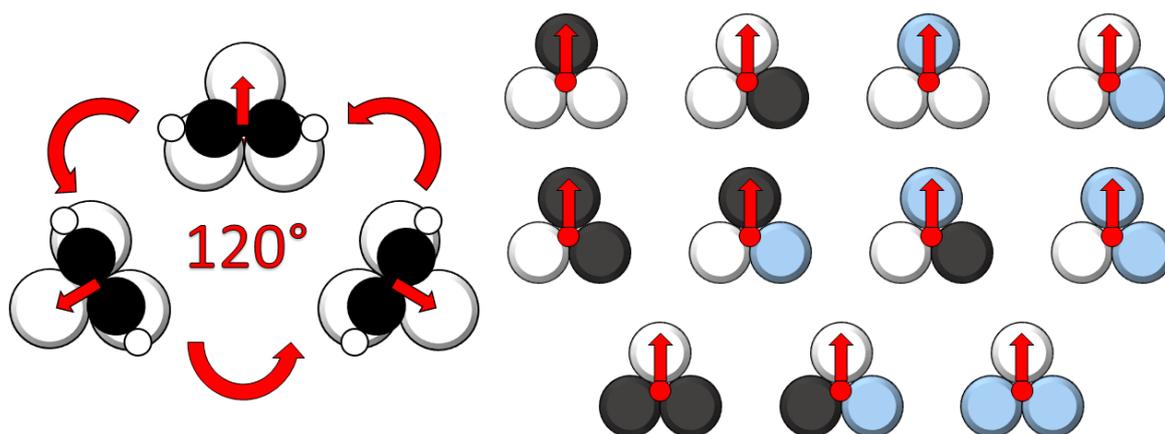


Figure 1. The effect of acetylene adsorption on pair interactions at the surface layer. Grey atoms represent Pd atoms, Ag atoms are coloured in light blue. White atoms stand for sites not included in the cluster

prediction of Pd segregation and Pt subsurface segregation, decoration of onion-shell structures were explored as functions of particle size and of the temperature. The latter was addressed as to promote formation of elaborated structures when decreased. Apart from intrinsic theoretical interest, such prediction are of value in the context of catalysis, where patterns in alloys control reactivity.⁵ A full comprehension of such control can only be achieved by implementing adsorbate-catalyst interactions into the model energetics. It is nowadays clear that these interactions play a major role in catalytic systems; environmental factors may significantly alter the compositional profile of the alloy topmost layers, especially if they are in contact with a gaseous reservoir.⁶ For instance, when atoms and molecules adsorb preferentially one component, the segregation profile can be reverted. This is the case of CO adsorption on the (100) plane of Pd-Au alloys, where competition between metal-adsorbate adsorption energy and adsorbate-adsorbate lateral generate stripes of Pd at room temperature; an analogous example derives from the computational study of Chen *et al.*,⁷ which highlighted the effect of oxygen binding on Au-Ag on pattern formation. In both cases, the energy was modelled by Ising-type Hamiltonians fitted to DFT calculations. Another remarkable case in this sense is the restructuring of Pd-Ag alloys employed to selectively hydrogenate acetylene to ethylene. Theory and experiments show indeed that the presence of adsorbates on the catalytic surface revert the composition at the gas-metal interface, favouring mixed configurations whereas the least catalytically active Ag should be predominant. This phenomenon was also investigated along the two directions parallel to the surface. The topmost layer of the alloy (111) plane is well described in terms of its triangular *ensembles*⁸ arrangements of three metal atoms. In high vacuum conditions, STM measurements found that the ensembles are randomly distributed for a wide span of conditions;⁹ this was attributed to the small difference between interactions for pairs of same and different metals. Our recent study¹⁰ added that adsorption of acetylene modifies the ensemble distributions. In particular, lateral segregation was predicted for layers close to chemical purity, i.e. with an Ag fraction either close to 0 or 1, while mixed ensembles prevail for equimolar ratios. These results were explained by the layer energy/entropy balance; the entropic drive enhancing random distributions is overcompensated by the adsorption energy of acetylene at the compositional tails. Anyway, in view of the system's complexity, we relied on rather strong approximations, among which a mean-field treatment of lateral interactions and a local interpretation of the acetylene-ensemble bonding; on top of that, interactions between molecules were assumed to be independent of the layer composition. A more refined approach to Pd-Ag ensemble distribution in acetylene would be highly beneficial in this respect.

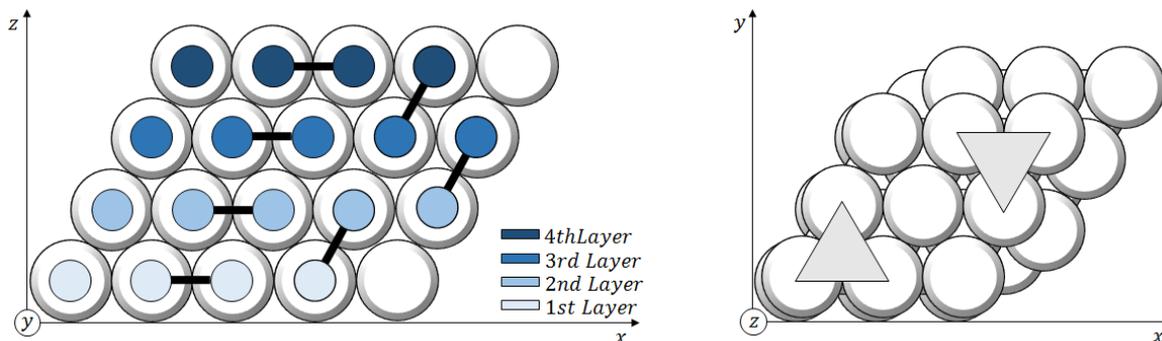


Figure 2. Catalyst model and selected interactions. Triangles in the top view (right part of the picture) denotes acetylene adsorption sites

This paper addresses the surface restructuring of Pd-Ag surface alloys on a more detailed basis. An interaction Hamiltonian is proposed which fits DFT calculations and is provided as input to a Monte Carlo code. Ensemble distributions are then estimated as thermodynamic constraints, i.e. the temperature and the surface composition, are changed.

Computational Details

All configurations were optimized at the Density Functional level of theory in the PAW formalism.^{11,12} All computations were performed with VASP 5.3.3.^{13,14} The functional of Perdew, Burke, and Ernzerhof¹⁵ was used, with the dispersion correction of Steinmann and Corminboeuf.¹⁶ The (111) surface was modeled by a $p(3 \times 3)$ unit cell with 4 metallic layers, 2 of which held fixed to simulate bulk properties. A vacuum layer of 15 Å was used. The plane waves basis set was chosen to have a cutoff energy of 400 eV. Brillouin zone integration was performed by a $3 \times 3 \times 1$ Monkhorst-Pack¹⁷ k-points grid and a Methfessel and Paxton¹⁸ smearing of 0.2 eV. The wavefunction and geometric gradient were converged to 10^{-6} eV and 5×10^{-2} eV/Å, respectively.

Numerical simulations were performed with a FORTRAN90 Monte Carlo code of our invention. The code is based on repetitions of a 4-layers, $p(2 \times 2)$ cell through real space; the number of repetitions is left as an input parameter. The user is at will to choose the simulation temperature and the acetylene reservoir pressure, along with the Ag molar fraction of the 4 layers and the number of iterations. The system is then initialized to a random configuration. The configuration sampling is based on the Metropolis algorithm.¹⁹ The output provides a calculation of the reservoir chemical potential in the ideal gas approximation. The chemical potential is calculated from statistical mechanics,²⁰ where the partition sum includes translational and rotational contributions only (vibrational and electronic degrees of freedom are neglected). The output files also include ensemble distributions, their coverage, and the selected geometries.

Results and Discussions

Ab initio construction of the model and its validation. It was decided to model the system with an Ising-type Hamiltonian represented on a cluster basis.²¹ A model of such kind requires a lattice of points distributed through the three dimensions of space. Each point in the lattice may be occupied by a Pd or an Ag atom; a specific chemical realization of the lattice is called a *configuration* of the system. Functions of configurations can also be defined; to our purposes, the most important function of the system's configuration is the *configurational energy*. In the present work, the configurational energy was expanded in terms of single-body, two-body ... N -body cluster contributions relative to subsets of atoms. Ising-type Hamiltonians of alloys are analogous to the spin Hamiltonians employed to model magnetic phenomena, but with one severe difference. As it is known, spins not only interact on each other, but also with an external field to which they can align or not. Prediction of alloy phase

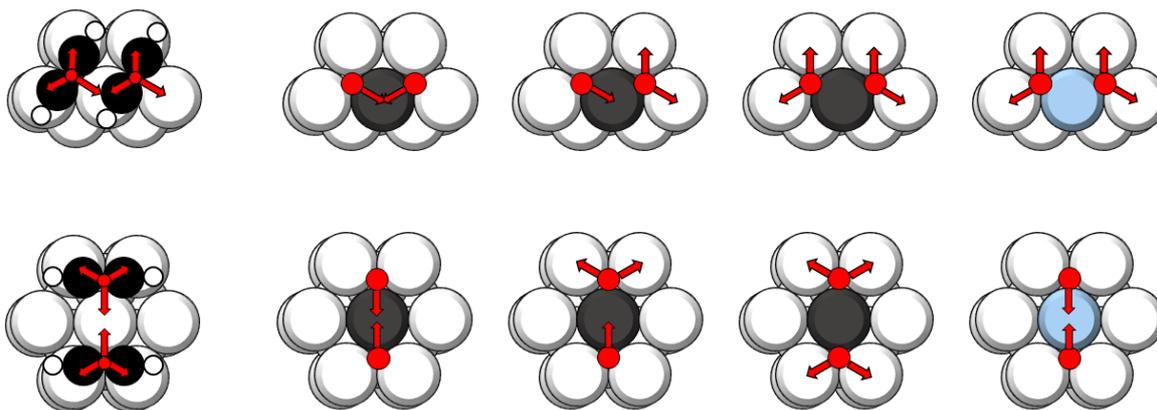


Figure 3. Lateral interactions and their effect on the energy of a single atom in the surface layer. Grey atoms represent Pd atoms, Ag atoms are coloured in light blue. White atoms stand for sites not included in the cluster

diagrams does not require any contact with external fields; anyway, a mathematical equivalence between external fields and gaseous reservoirs can be demonstrated.²² This equivalence is the essence of the present context, although it must undergo some rewriting, due to the heterogeneity of acetylene adsorption; that is, the acetylenic “field” will be described by an additional gas-lattice 2D layer superimposed to the lattice of metallic atoms. The local orientation of the acetylenic field relative to an adsorption site will be depicted with an arrow, as shown in Figure 1.

An aspect common to alloy systems is the energy of a particular metallic species depends on its position along the normal to the catalyst surface;²³ as a result, one of the species is expected to segregate at the top of the surface whenever the temperature is low enough. Distinct energy coefficients were therefore assigned to Pd and Ag atoms for each of the positions in the 4 layers. This is pictured in Figure 2, where interactions between different positions are also displayed. Along with single-atom terms, pair interactions were included in the model, while third-order and higher coefficients were neglected based on previous knowledge of the system.⁹ Notice that both intra- and inter-layer contributions were considered in agreement with the symmetry loss along the normal axis (z axis of Figure 2). The number of metallic atom interaction terms to be included into the model amounts therefore to 32.

To assess the effect of acetylene on the (111) surface plane of Pd-Ag alloys, the geometry of adsorption on palladium must be considered in the first place. It is known²⁴ that acetylene adsorbs on palladium (111) at hollow sites, triangular arrangements of catalytic atoms (see Figure 1 and 2). At low pressure and temperature, a molecule of acetylene lies in the middle of such sites, strongly bound to two of the site’s atoms and weakly to the other one, in a mode of adsorption known as the di- σ mode. The molecule seldom jumps to another site via a diffusional transition state where it interacts with two of the site’s metallic atoms – the so-called “bridge” mode; an additional, local minimum is the π mode, where acetylene sits on only one metallic atom. Since no more than 3 metallic atoms arranged in a triangular fashion are required to capture acetylene adsorption in full, single-adsorbate corrections to the overall energetics should be coherently stated in terms of triangular ensembles. This criterion with the triangular clusters reported in Figure 1, which also shows an important characteristic of the di- σ mode; the site accommodates the molecule in three different orientations, generated by rotations of 120° . For the case of pure metal ensembles, i.e. Pd₃ and Ag₃, the three rotations are degenerate in energy because of symmetry considerations. On this basis, each compositionally homogeneous site would just require two single-body adsorbate-induced correction (one for the weakly interacting and the other one for the strongly interacting metal atom) and two two-body corrections; nevertheless, the presence of impurities in sites of mixed composition, Pd₂Ag

and PdAg₂, lowers the site symmetry, thereby imposing the addition of seven more mixed two-body corrections. The legend of Figure 1 should be clear by now: acetylene exhibits different orientations, as represented by an arrow normal to the C-C axis, pointing at the weakly interacting metal atom; the two kinds of atom (weakly and strongly interacting) may be either palladium or silver. This makes 11 correction coefficients to be furnished to the model Hamiltonian.

The 43 coefficients considered up to this point describe the adsorption process of an ideal lattice-gas adsorbed layer. As we showed previously,¹⁰ acetylene molecules adsorbed on Pd-Ag catalyst exert rather strong interactions on each other at short contacts; lateral interaction coefficients must therefore be furnished to the model as well. Lateral interactions of nearest neighbour pairs were the first to be appraised. Drawing starting geometries for these interactions was particularly hard and only possible when molecules exhibited same orientations; optimizations resulted either into much strongly interacting pairs on pure sites (with values around 1.20 eV over the ideal gas approximation) or into the ground-state of adsorbed cyclobutadiene on mixed sites. In view of this, nearest neighbour lateral interactions were not included in the model. On the other hand, second and third nearest neighbour interactions were found for a large span of sites compositions. As displayed in Figure 3, they contribute for a further set of 8 single-body terms.

The cluster expansion was fitted to a sample of 107 DFT calculations, which were generated so as to comprehend all the listed interactions for all possible sites compositions. Nevertheless, optimizations were not successful for some of the configurations including more than one molecule adsorbed; these configurations were therefore rejected out of the sample. A subset of 74 calculations relaxed to the envisaged minimum and the 51 coefficients pictured in Figures 1, 2, and 3 were isolated. The coefficient numerical values were interpolated by a least-squares fit, by Moore-Penrose pseudoinversion²⁵ of the following equation

$$\mathbf{E}_{DFT} = \mathbf{A} \cdot \boldsymbol{\epsilon} \quad (1)$$

Here, \mathbf{A} is a 74×51 matrix whose elements are the occurrences of cluster in a given configuration, \mathbf{E}_{DFT} the 1×59 column vector of the configuration total energies, and $\boldsymbol{\epsilon}$ is the 1×74 column vector of the cluster coefficients. The results, shown in Table 1, are rather satisfactory from the standpoint of chemistry. Coefficients associated to adsorption on Pd all favour the process, while interaction of acetylene with Ag is of little importance. The weak/strong relation between acetylene and metallic atoms is captured for Pd, with an energy difference between the two coefficients of 0.16 eV; in an

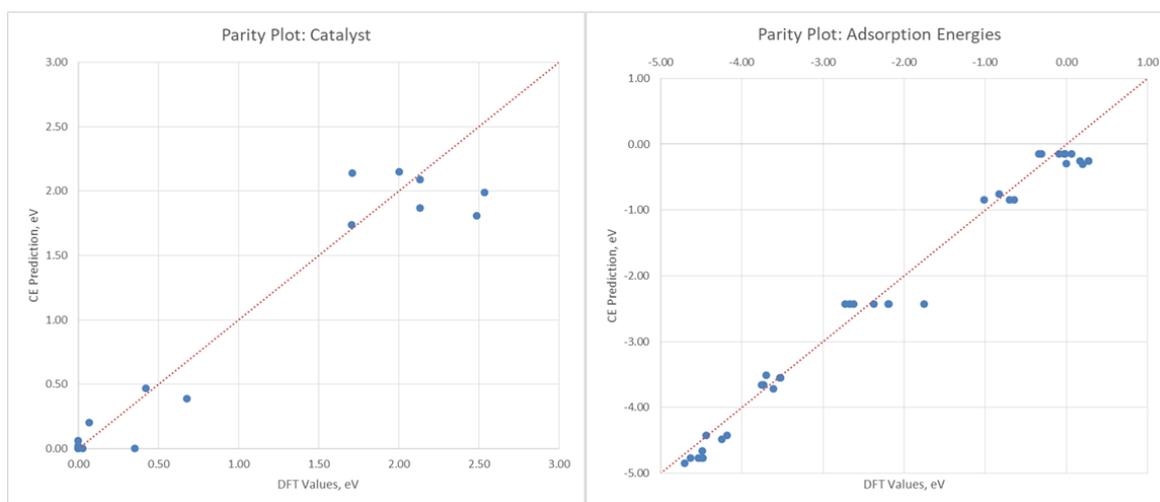


Figure 4. Parity plots of bare surface models (left) and adsorption energies (right)

analogous way, the coefficient of a weakly interacting Ag atom is 0.14 eV higher in energy than the strongly bonded one. Lateral interactions are generally repulsive or of no significance. Second nearest neighbour interactions discourage orientations for which at least one of the molecules interacts weakly with a Pd atom; the trend is less neat for third nearest neighbour pairs, which seems to privilege alternated orientations. The situation for an Ag atom is different, which exhibits negligible interactions.

It is known that least-squares interpolations are subjected to overfitting, i.e. not only they model relevant information, but noise as well. For this reason, leave-one-out cross validation[#] was performed on the system of equation 1 to ascertain the model Hamiltonian predictive power. This consisted in the exclusion of one configuration from the computational pool to pseudoinvert the remaining system of 73 equations; the excluded configurational energy was then predicted and compared to its ab initio calculated value. After 74 repetition of this routine, a cross-validation score was obtained taking the average deviation of the leave-one-out interpolations – all the root mean squares are reported in Table 2. The score of 0.3 eV was considered safe enough to guarantee a reasonably precise model.

Leave-one-out cross validation allows one to validate the model to the extent of the available sample. It is anyway instructing to check the model inspecting its predictive power towards atomic swaps and adsorption events, which are at the core of the numerical results presented later on. Parity plots of predicted atomic swaps in structures of constant composition (left) and adsorption energies (right) are shown in Figure 4. It can be seen that the sample is a bit clustered for what concerns the catalyst, while adsorption energies were captured consistently. Adsorption energies of -2.42 eV on Pd₃ sites, -1.48 or -1.40 eV on Pd₂Ag sites (depending on the orientation), -0.76 or -0.66 eV on PdAg₂ sites, and -0.15 eV on Ag₃ sites were found. It should be noted that these values do not depend on the catalytic subsurface layers, for the model is limited to interactions of acetylene and the topmost layer only; this limitation is reflected into the small deviations in the right part of Figure 4, which are mostly due to a second-layer effect. This effect must be summed to the effect lattice deformations, which are not stated in the model. The dependence of the lattice constant on the catalyst composition is the main responsible for the deviations displayed in the left part of Figure 4, indeed.

Numerical Simulations. The model Hamiltonian was provided as input to the Monte Carlo code. Numerical experiments were designed with two goals in mind, namely 1) to predict the equilibrium

distribution of Ag through the model, with special focus on the topmost layer, and 2) to assess the effect of the temperature on this distribution. To this end, the global molar fraction of Ag and the temperature were used as control parameters, keeping the simulated pressure of acetylene at the constant value of 1 bar. The fraction of Ag was scanned with steps of 0.1, generating a set of eleven simulations per pressure-temperature couple. The temperature was increased by 100 K after completion of the precedent set, starting with the catalytically relevant temperature of 300 K up to 1300 K; this amounts to 121 simulations distributed into 11 sets, scanning the 300 K – 1300 K range. The simulated systems consisted of 400 atoms distributed through the full 4-layers model, with 200 sites available to adsorption. Convergence of all simulations was achieved long before 100000 steps.

Figure 5 shows the distributions of triangular ensembles in the topmost layer and their covered fraction as the global Ag fraction in the full 4-layers model is increased. Since the picture is almost constant in the range 300 K – 500 K and in the range 600 K – 1200 K, only results obtained at 300 K and 700 K are displayed, so that most important features can be highlighted (all graphs are available in the Supporting Information). The striking feature at 300 K is the domination of Pd₃ ensembles in the wide range 0.0 – 0.7 of Ag fraction, accompanied by a high coverage of the surface. As soon as a composition of 0.8 in Ag is reached, a dramatic fall in the Pd₃ ensemble population is observed; Ag₃ abruptly represents the layer, along with a small fraction of Pd₃ sites, which eventually disappear at 1.0. The same picture seems to hold at 700 K as well, but with one difference; after the 0.8 composition is reached, Pd₃ ensembles disappear almost completely. This aspect is striking as the populations of Pd₃ and PdAg₂ trade places with the increase of temperature: palladium seems to show a preference for mixed sites after 600 K.

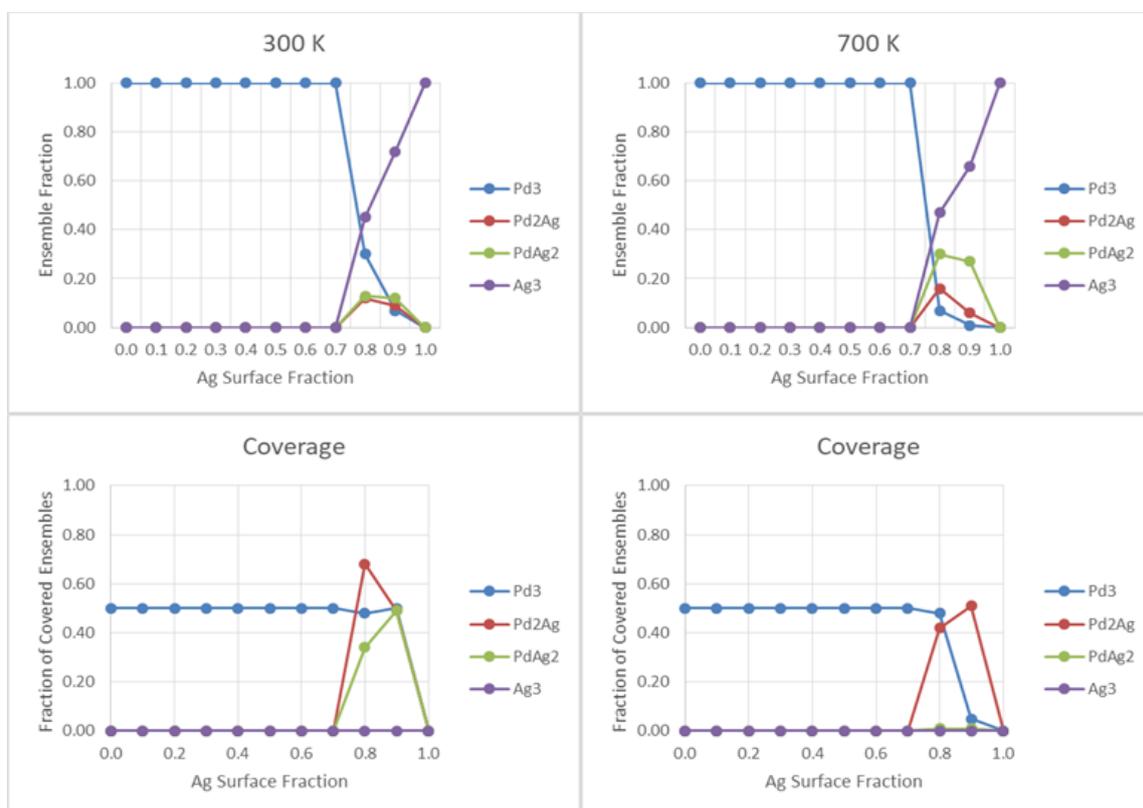


Figure 5. Results of numerical simulations at two significant temperatures with a simulated acetylene pressure of 1 bar

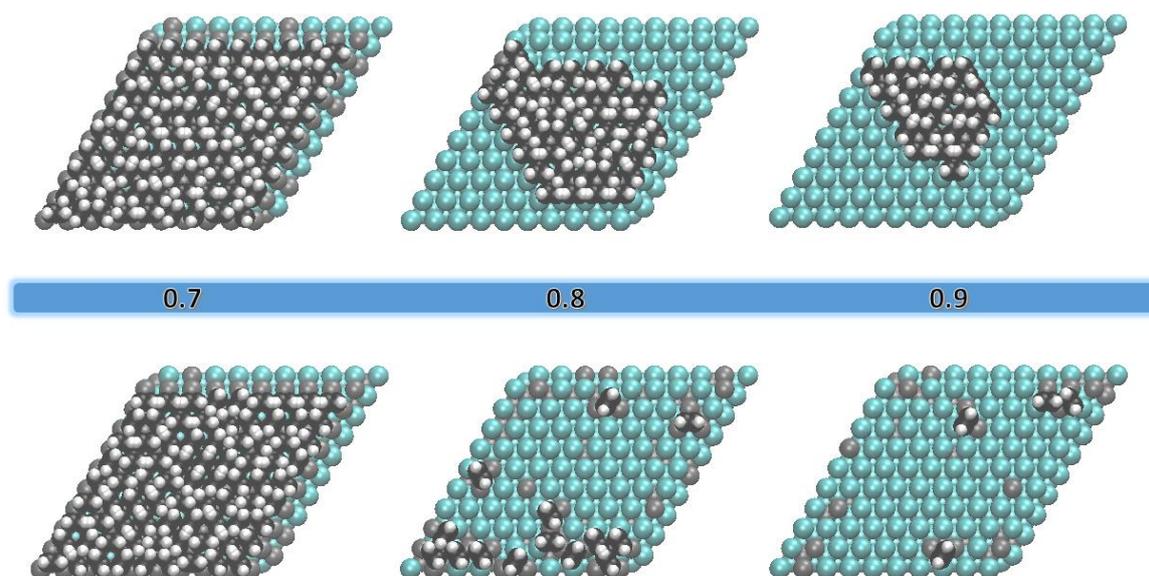


Figure 6. Comparison of the transition at 300 K (top) and 700 K (bottom) as Ag gets constrained to populate the toppermost layer. From left to right, the most stable configurations at 0.7, 0.8, and 0.9 Ag fraction.

The chemical picture of this change is best seen in Figure 6. Here, snapshots of the most stable model configurations at 0.7, 0.8, and 0.9 Ag fraction are displayed. At 300 K and 0.7, a full layer of Pd is covered by acetylene molecules, while at 0.8 and 0.9 an island of Pd (still covered) is segregated. At 700 K, the picture at 0.8 and 0.9 is somewhat different. The isolated particle gets dispersed through the layer; all the catalytic activity is discharged upon the now covered Pd₂Ag sites, while PdAg₂ sites are left uncovered. Notice that this trend was also predicted by Vignola *et al.*¹⁰ with their approximate model; the only difference with present results lies in the small boundary of Pd₂Ag and PdAg₂ sites originating by the finiteness of the model here considered.

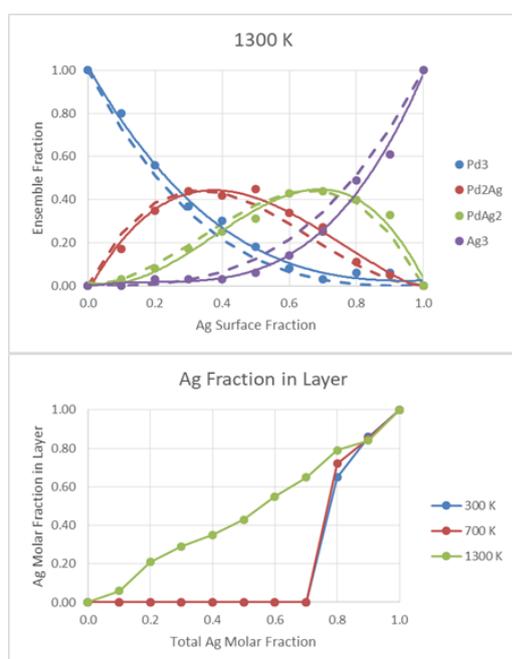


Figure 7. High-temperature limit for sites distributions. Dashed curves represent random distributions, full lines are polynomial interpolations of data. Evolution of the layer content of Ag is also plotted for three temperatures of interest

The nature of the 0.7 – 0.8 transition is best appreciated by a sustained increase in the temperature. Figure 7 shows the fractional amount of Ag in the toppermost layer as a function of the total Ag fraction as plotted for three relevant temperatures in the 300 K – 1300 K range. Numerical results of ensemble distributions at 1300 K are also plotted and superimposed to random distributions, in order to make the reader sure of their randomized character; this also shows that the 0.7 – 0.8 transition is suppressed at 1300 K. As long as the temperature is low enough, the profile of the Ag layer fraction is discontinuous in the 0.7 – 0.8 range; this discontinuity gets smoothed out at 1300 K. This reminds the behaviour of order parameters in equilibrium phase transitions, such as the induced magnetization in ferromagnetic systems or order-disorder phase transitions in bulk alloys. This transition also allows for a nice chemical interpretation. As mentioned in the Introduction, adsorbates induce strong segregation of Pd, pushing atoms to the toppermost

layer; when the adsorbed state still gains free energy over the gas phase state, molecules stick to the surface and promote segregation. At high temperatures, the chemical potential of the gas phase is too low to push molecules on the catalytic surface, which lacks the drive to organize the topmost layer. It is significant that the abrupt change takes place when the global content of Pd is close to 1/4: this is exactly the amount of Pd required to fill a whole layer.

Conclusions

This work ascertained that, at realistic pressures, acetylene adsorption alters the decoration of Pd-Ag topmost layers significantly in a wide range of operational conditions (300 K – 1300 K). At low concentrations of Ag, reverse segregation pushes Pd atoms to populate the catalyst entirely. As soon as 0.8 Ag is reached, a transition is observed. The nature of this transition depends on the temperature; under 600 K, a Pd particle is segregated, while over 600 K Pd gets dispersed through the layer. Such a conclusion was already seen in our previous study, of which the present work is a validation.

Supporting Information available

Embedding energies, results with the PBE exchange correlation functional and all geometries are provided. All programs for numerical simulations are also available.

Corresponding Author

sautet@ucla.edu

Acknowledgements

Ab initio calculations were performed using the local HPC resources of PSMN at ENS-Lyon. *Selective Hydrogenation of Acetylene in an Ethylene-Rich Flow: Insights from Molecular Modeling* is a research project funded by TOTAL Refining & Chemicals and the ANRT.

References

- (1) Callister, W. D.; Rethwisch, D. G. *Materials Science and Engineering. An Introduction*, 7th ed.; Wiley Publishers, 2006.
- (2) Tsai, M.-H.; Yeh, J.-W. High-Entropy Alloys: A Critical Review. *Mater. Res. Lett.* **2014**, 2 (3), 107–123.
- (3) Zou, Y.; Ma, H.; Spolenak, R. Ultrastrong Ductile and Stable High-Entropy Alloys at Small Scales. *Nat. Commun.* **2015**, 6, 7748.
- (4) De Clercq, A.; Giorgio, S.; Mottet, C. Pd Surface and Pt Subsurface Segregation in Pt₁-cPd_c Nanoalloys. *J. Phys. Condens. Matter* **2016**, 28, 64006.
- (5) Clarke, J. K. A. Selectivity in Catalysis by Alloys. *Chem. Rev.* **1975**, 75 (1), 291–305.
- (6) Tao, F.; Salmeron, M. In Situ Studies of Chemistry and Structure of Materials in Reactive Environments. *Science (80-)*. **2011**, 331, 171–174.
- (7) Chen, W.; Schmidt, D.; Schneider, W. F.; Wolverton, C. Ordering and Oxygen Adsorption in AuPt/Pt(111) Surface Alloys. *J. Phys. Chem. C* **2011**, 115, 17915–17924.
- (8) Ponc, V.; Bond, G. C. Catalysis by Metals and Alloys. In *Elsevier*; Ponc, V., Bond, G. C., Eds.; Elsevier B.V.: Amsterdam, The Netherlands, 1995; pp 393–430.

- (9) Engstfeld, A. K.; Hoster, H. E.; Behm, R. J. Formation, Atomic Distribution and Mixing Energy in Two-Dimensional PdAg_{1-x} Surface Alloys on Pd(111). *Phys. Chem. Chem. Phys.* **2012**, *14* (30), 10754.
- (10) Vignola, E.; Steinmann, S. N.; Vandegehuchte, B. D.; Curulla, D.; Sautet, P. C₂H₂-Induced Surface Restructuring of Pd–Ag Catalysts: Insights from Theoretical Modeling. *J. Phys. Chem. C.*
- (11) Blochl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50* (24), 17953–17979.
- (12) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59* (3), 1758–1775.
- (13) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47* (1), 558–561.
- (14) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54* (16), 11169–11186.
- (15) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (16) Steinmann, S. N.; Corminboeuf, C. Comprehensive Benchmarking of a Density-Dependent Dispersion Correction. *J. Chem. Theory Comput.* **2011**, *7*, 3567–3577.
- (17) Pack, J. D.; Monkhorst, H. J. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* **1976**, *13* (12), 5188–5192.
- (18) Methfessel, M.; Paxton, A. T. High-Precision Sampling for Brillouin-Zone Integration in Metals. *Phys. Rev. B* **1989**, *40* (6), 3616–3621.
- (19) Frenkel, D.; Smit, B. *Understanding Molecular Simulations*; Academic Press: San Diego, 1996.
- (20) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications, Ed.; New York, 1960.
- (21) Sanchez, J. M.; Ducastelle, F.; Gratias, D. Generalized Cluster Description of Multicomponent Systems. *Phys. A Stat. Mech. its Appl.* **1984**, *128* (1–2), 334–350.
- (22) Lee, T. D.; Yang, C. N. Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model. *Phys. Rev.* **1952**, *87* (3), 410–419.
- (23) Ruban, A. V.; Abrikosov, I. A. Configurational Thermodynamics of Alloys from First Principles: Effective Cluster Interactions. *Reports Prog. Phys.* **2008**, *71* (4), 46501.
- (24) Dunphy, J.; Rose, M.; Behler, S.; Ogletree, D.; Salmeron, M.; Sautet, P. Acetylene Structure and Dynamics on Pd(111). *Phys. Rev. B* **1998**, *57* (20), R12705–R12708.
- (25) Barata, J. C. A.; Hussein, M. S. The Moore-Penrose Pseudoinverse. A Tutorial Review of the Theory. **2011**.

Chapter 5. Evaluating the Risk of C-C Bond Formation during Selective Hydrogenation of Acetylene on Palladium

Emanuele Vignola,^{†‡} Stephan N. Steinmann,^{†*} Ahmad Al Farra,[‡] Bart D. Vandegehuchte,[§] Daniel Curulla,[§] and Philippe Sautet^{†&¶*}

[†]Univ Lyon, ENS de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France

[‡]Total Research and Technology Gonfreville, BP 27, F-76700 Harfleur, France

[§]Total Research and Technology Feluy, Zone Industrielle Feluy C, Seneffe, Belgium

[&]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, CA 90095, United States

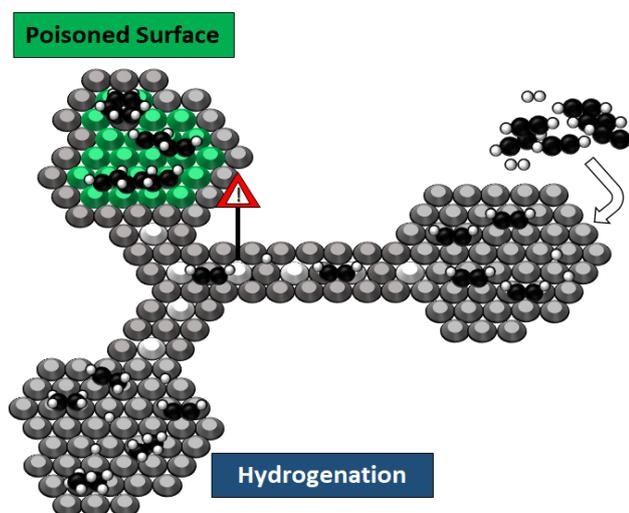
[¶]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, United States

*e-mail: stephan.steinmann@ens-lyon.fr; tel.: 0033-4-72-72-81-55

*e-mail: sautet@ucla.edu; tel.: 1-310-825-8485

Submitted to ACS Catalysis

ABSTRACT: Palladium-based catalysts are known to promote the selective hydrogenation of acetylene to ethylene. Unfortunately, coupling reactions of the numerous surface intermediates generated in this process can happen alongside. These side reactions are undesired, generating the so-called “green oil”, i.e., C₄+ hydrocarbons that poison the active sites of the catalyst. The current work assesses the energetic and kinetic aspects of C₄ side products formation from the standpoint of computational chemistry. Our results demonstrate that the C-C coupling of common surface species, in particular acetylene and vinyl, are competitive with selective hydrogenation. Furthermore, the thus formed oligomers tend to be hydrogenated more easily, consuming hydrogen normally spent on acetylene hydrogenation. The analysis of site requirement suggests that isolated Pd₂ ensembles are sufficient for selective hydrogenation and would suppress oligomerization.



Introduction

Carbon-carbon coupling reactions are among the most useful transformations in organic chemistry.¹ They allow chemists to build complex carbon frameworks from their reagents, thereby

achieving products of greater complexity, spreading benefits from daily laboratory practice to the pharmaceutical, the biochemical and the polymer industry. Palladium is at the core of this chemistry for fine chemicals, since it is the elected metallic catalyst promoting carbon-carbon coupling.² Though the majority of catalysts are palladium complexes to be used in homogeneous phase reactions, supported heterogeneous palladium catalysts start to attract interest as a promising alternative. Today, Pd-based heterogeneous catalysts are particularly important in the industrial production of polyethylene. Ethylene streams from oil refineries contain about 1% acetylene. Leaving acetylene in the ethylene feedstock for polyethylene would lead to deteriorated material properties and decreased catalyst lifespans. Therefore, the ethylene stream is subjected to a partial hydrogenation process,³ where acetylene is selectively hydrogenated to ethylene, while avoiding over-hydrogenation towards ethane. This selective hydrogenation is carried out over Pd alloy catalysts, with Pd surface atoms constituting the active sites. .

Unfortunately, during this process Pd can also catalyse oligomerization of hydrogenation intermediates, reducing ethylene throughput; indeed, Pd acts as a C-C coupling catalyst, producing complex mixtures of oligomeric compounds which reduces the atom efficiency.⁴ The mixture obtained from C-C coupling contains three main fractions, a light end fraction which remains in the gas phase, a liquid portion that is entrained by the gas in the form of fine droplets (denoted as “green oil”), and a heavy fraction, the “sticky green oil”. The sticky green oil blocks active sites and therefore slowly deactivates the catalyst. The catalyst needs to be regenerated periodically by oxidation followed by re-activation through reduction. Suppressing green-oil formation would improve catalyst life span and increase atom efficiency.

The characterization of these coupling derivatives is not a trivial task. Few groups have tried to analyse the carbonaceous deposits, and their results are not always in agreement. While it is commonly agreed that the gaseous fraction is prevalently composed of C₄ and C₆ hydrocarbons,⁴ discrepancies are more obvious regarding the composition of green oil. Yayun *et al.*⁵ reported that the liquid heavy polymers have an average composition of C_nH_{(1.8-1.9)n} with a number of carbon atoms between 14 and 17, and an olefin/paraffin ratio between 0.1 and 0.4. Most remarkably, no aromatics were detected although palladium is renowned to boost acetylene trimerization.⁶ Gandman *et al.*⁷ have presented a different green oil composition, with a content of aromatic compounds as high as 74.5% in weight, 23.0% of olefins, and a small fraction of diolefins. The origins of the oligomers is also controversial. Most authors agree that the oligomers are exclusively derived from acetylene, and that they are produced only in the presence of hydrogen. Therefore, the reaction is often referred to as acetylene hydro-oligomerization. Nevertheless, according to the results obtained by Yayun *et al.*⁵, ethylene participates in the process as well.

In addition to C-C coupling, C-C decomposition products are expected. In particular, carbon atoms penetrate the surface layer and preferentially occupy the octahedral vacancies of the lattice to form an interstitial alloy.^{8,9} This Pd/C phase crucially alters the adsorption energetics of the reagents,¹⁰ thereby lowering acetylene conversion but improving the selectivity towards ethylene as it weakens ethylene adsorption and hinders the kinetics of subsurface hydrogen diffusion.

The complete reaction network encompasses Horiuti-Polanyi additions of hydrogen atoms to acetylene¹¹, as well as oligomerization by C-C bond formation and decomposition of intermediates into carbon and hydrogen. Consecutive hydrogenation and dehydrogenation lead to ethylidyne (CCH₃)

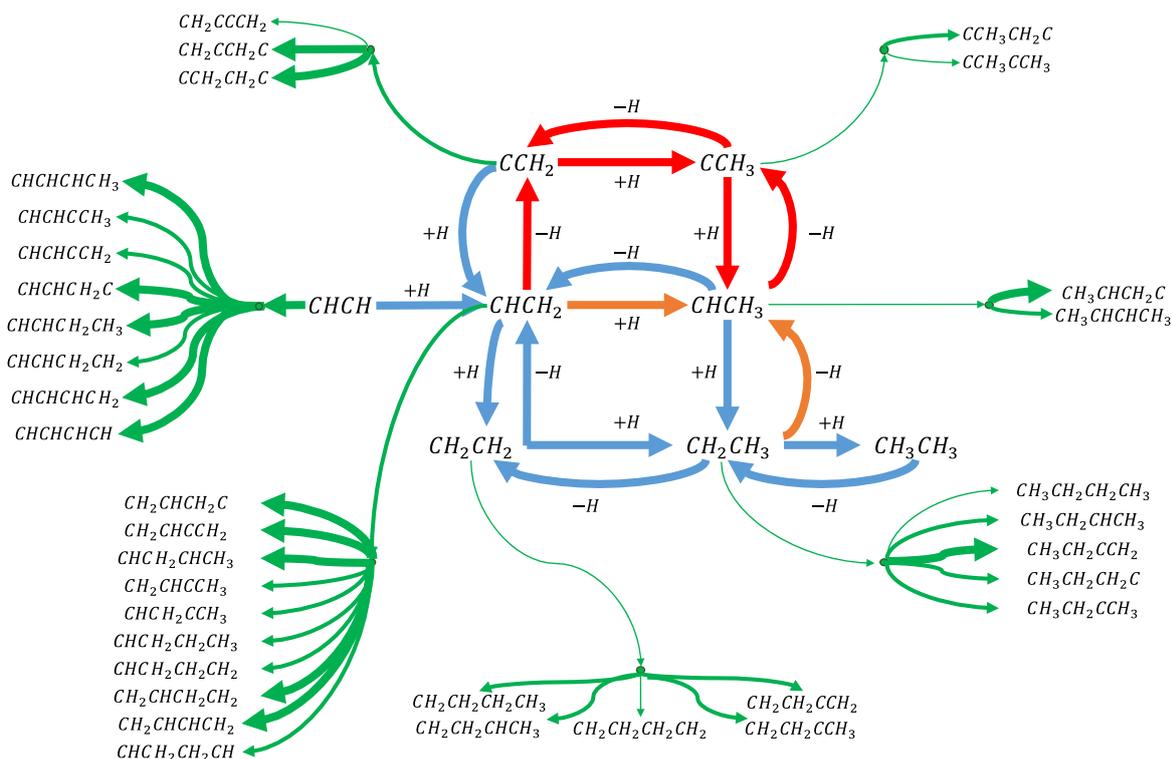


Figure 1. Reaction Scheme considered. The green arrows indicate unwanted oligomerization, the blue ones indicate the “direct” hydrogenation of acetylene and the red ones refer to the isomerization process to dissymmetric C₂ species. Arrows in thin, medium and bold lines represent activation energies < 1.5 eV, between 1.5 and 2.0 eV and > 2 eV respectively. For the various hydrogenation elementary reactions of C₄ species considered, the reader is referred to Figure 5.

as an dissymmetric side-intermediate, either via vinyl (CHCH₂) or ethylidene (CHCH₃). The ethylidyne route, computationally investigated in 2002,¹² and later refined by Moskaleva *et al.*¹³ was finally probed by IR measurements.¹⁴ Ethylidene formation seems to be kinetically preferred when subsurface hydrogen is allowed to take part to vinyl hydrogenation.¹⁵ As already noted, decomposition products (C₁ species, subsurface carbon and hydrogen) can play a role in the mechanism, as investigated by Mei *et al.*¹⁶ Herein, however, we consider only C₂ and C₄ species, in order to focus on the competition between oligomerization and selective hydrogenation of acetylene. The elementary steps considered herein are shown in Figure 1, where the hydrogenation reactions converting C₄ species into each other are excluded for sake of clarity (see Figure 5 for the C₄ species).

If decomposition steps were amply considered next to acetylene hydrogenation, coupling mechanisms are even less explored, surprisingly considering that they lead to noticeable side products. Based on the accepted view that butadiene is involved in the formation of the heavy hydrocarbons, Lopez and Vargas-Fuentes proposed¹⁷ an acetylene-acetylene cis coupling step to C₄H₄ requiring an ensemble of at least 4 active metal atoms. Yang *et al.*¹⁸ proceeded with the investigation by comparing three paths to 1,3-butadiene on plain and stepped surfaces in the presence of subsurface carbon and alloyed silver atoms; it appeared that, on a clean Pd(111) surface, acetylene-acetylene coupling followed by hydrogenation is the preferred road to butadiene, while the Pd(211) stepped surface favours vinyl-vinyl coupling. The presence of subsurface carbon or small amounts of alloy metals reduces adsorption energies for all intermediates and slightly (~0.1 eV) raises activation energies. Though these contributions are valuable first steps, Figure 1 illustrates that a large spectrum of carbon-carbon couplings are theoretically possible and that only a detailed investigation of them can identify the most relevant pathways.

This paper addresses the aspects of acetylene hydrogenation and oligomerization on palladium from first-principle calculations. Particular attention has been given to the roles of ethylene and of the ethylidyne-related fragments, which are analysed extensively. The presented detailed mechanistic study emphasizes the role of early hydrogenation intermediates, both responsible for the selectivity towards ethylene, rather than ethane, and of the undesired oligomerization reactions. Finally, implications for catalysis on Pd-Ag alloy catalysts are discussed based on the site and ensemble-requirements for the selectivity-controlling steps.

Computational Details

Total energies were calculated at the Density Functional level of theory in the PAW formalism,^{19,20} with the Vienna Ab Initio Simulation Package.^{21,22} The functional of Perdew, Burke, and Erzhennoff²³ was employed, corrected for intermolecular forces following the scheme of Steinmann and Corminboeuf.²⁴ A plane wave basis set was used, with a cut-off energy of 400 eV. Surface models were cut out of palladium's f.c.c. structure along the 111 crystal direction. A p(3x3) unit cell with 4 metallic layers was used for the surface model, keeping the bottom-most two layers fixed, in order to simulate bulk properties. 15 Å of vacuum were inserted between periodic images along the axis normal to the surface, which proved to be sufficient to avoid spurious interactions between repeated images of the cell. Sampling of the Brillouin zone was performed with a Monkhorst-Pack²⁵ generated 3x3x1 k-points grid. The Methfessel-Paxton²⁶ smearing scheme was employed in all calculations, with a broadening value of 0.2 eV. The transition states (TS) have been located as follows: the co-adsorbed state corresponding to the two reacting fragments has been constructed based on the adsorption mode of the individual fragments, moved to neighbouring positions. Then, a nudged-elastic band (NEB)²⁷ computation with 8 images between the initial and final state was performed, seeded by interpolations between the two states obtained by the Opt'n Path code²⁸ which uses a combination of internal and Cartesian coordinates. After about 50 cycles of NEB, the improved guess for the transition state was fully optimized to a first order saddle point by the dimer method²⁹ and the normal modes checked to contain only one imaginary frequency corresponding to the bond formation process.

Chemical potentials of gaseous species were calculated from statistical mechanics,³⁰ ignoring the vibrational contributions to partition functions, at 300 K and 1 bar.

Results and Discussions

Hydrogenation of C₂ species. We first present the thermodynamics and kinetics for the hydrogenation/dehydrogenation steps of C₂ intermediates derived from acetylene.

Industrially, the reaction is carried out on Ag-Pd based catalysts at around 350 K for and in a large excess of ethylene. However, for this model study on Pd(111), we have chosen standard conditions to reduce the complexity, i.e., 300 K and 1 bar for all gases. Figure 2 shows the free energy profile for the C₂ hydrogenation reaction network subset of Figure 1. A molecule of acetylene is adsorbed on the catalyst exhibiting a free energy loss of 1.95 eV. Addition of a surface hydrogen (referenced as ½ H₂), leads to vinyl (CHCH₂) corresponding to a barrier of 1.08 eV. At this point, the system may evolve along two different pathways (red and blue in Figure 2). On the one hand, an additional atom of hydrogen results in ethylene following an activation energy of 0.96 eV. Either ethylene desorbs quickly (the process is non-activated, endergonic by 0.7 eV), or hydrogenates further to ethyl with an activation energy of 1.00 eV. On the other hand, vinyl may dissociate relatively fast

(barrier of 0.58 eV) to atomic hydrogen and vinylidene, CCH_2 . Vinylidene may be hydrogenated to ethynidyne, CCH_3 with an activation energy of 1.01 eV (the reverse barrier is 1.26 eV). Ethynidyne plus hydrogen is a particularly interesting intermediate, being both the global free energy minimum and the kinetically most inert point. The activation energy of CCH_3 to $CHCH_3$ is 1.20 eV, while the reverse barrier is only 0.17 eV). In any event, if $CHCH_3$ is produced, its hydrogenation to CH_2CH_3 can proceed with an activation energy of 0.95 eV. Ethyl can be finally hydrogenated to ethane with a barrier of only 0.75 eV, and the weakly bound ethane product quickly desorbs. An additional interconnection between the “direct” (in blue) and the “dissymmetric” (in red) pathways comes from the dissymmetric hydrogenation of vinyl. Here, vinyl goes directly into $CHCH_3$ by an activation energy of 0.95 eV.

From these free energy pathways, the experimentally observed CCH_3 can either be formed from CCH_2 (barrier of 1.01 eV) or by the dehydrogenation of $CHCH_3$ ($\Delta E^\ddagger=0.17$ eV), which converts an ethane precursor into an inert spectator. The hydrogenation of $CHCH_2$ is found to be the most critical step for ethylene selectivity, since the barriers towards ethylene and towards the unwanted $CHCH_3$ are found to be equivalent (0.96 eV). Once ethylene is formed, its desorption is easier than its further hydrogenation implying that no ethane should be formed this way. Based on the stability of intermediates and the energy barriers, we expect that the surface would be predominantly covered by hydrogen, C_2H_2 and CCH_3 . This predicted state of the surface constitutes a first approximation to a more detailed state to be captured introducing lateral interactions into the model.

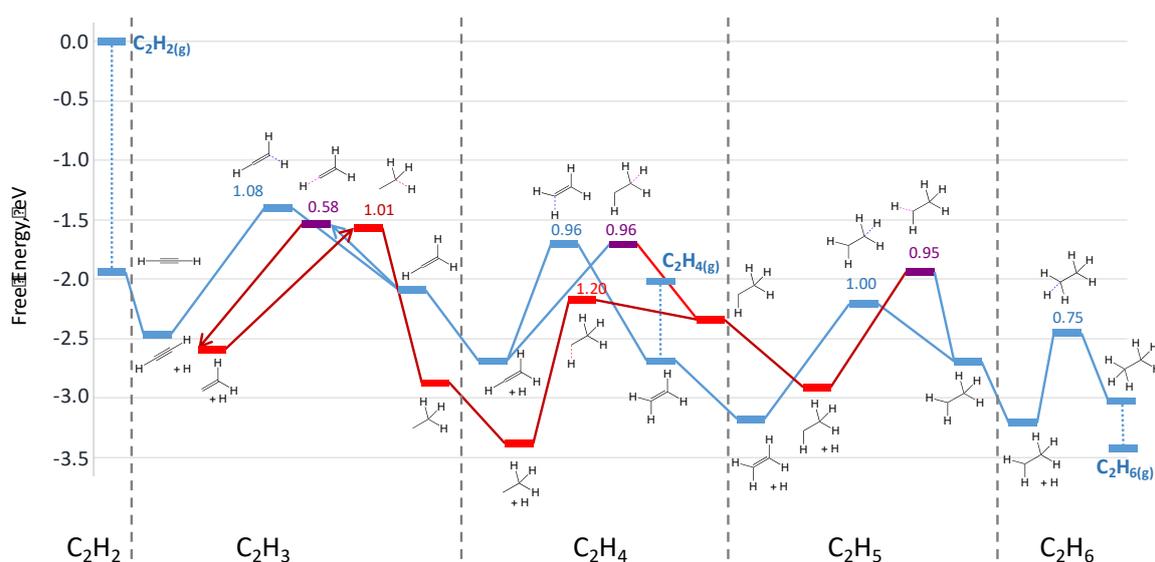


Figure 2. C_2 hydrogenation network free energy (eV) on Pd(111) at standard conditions. The path in blue is the “direct” hydrogenation path to ethylene and ethane. The path in red involves a rearrangement of hydrogen atoms, leading to the dissymmetric ethynidyne intermediate. In purple are the transition states that connect the two pathways. All depicted intermediates are chemisorbed on the surface. Numbers indicate the activation energies of the elementary steps.

C₄ derivatives formation. Figure 2 only includes reactions leading to the hydrogenation and isomerization of C₂ species. We include now the elementary steps involved in oligomerization. All C₂ coupling reactions were considered, since an *a priori* identification of thermodynamic and kinetic restrictions is difficult. Most of the C₂ intermediates are dissymmetric and therefore have two non-equivalent carbons that can form a new C-C bond. A convenient way to classify these chemical functionalities is to count the number of hydrogens bound to the carbon atom. If S stands for the catalytic surface, C for the carbon atom involved in the coupling and R for the second carbon fragment of the C₂ intermediate, the chemical functions available for C-C coupling are the carbyne group, R-C≡S, the methylidyne group, R-CH=S, and the methylene group, R-CH₂-S. Following this logic, acetylene has only one functional group, S=CH-CH=S, while vinyl has two, i.e., R-CH=S and S-CH₂-R. Applying these rules leads to a total of 58 distinct C₄ intermediates. On top of that, E/Z isomers may exist and the activation energy might depend on the type of surface site (top, hollow etc). We explored all the possible intermediates and, for the stable ones, located the E or Z transition state, considering three possibilities for the newly formed C-C bond: top, bridge and hollow. With this approach, we were able to identify 50 reaction steps as reported in Table 2 in the Annex.

Figure 3 illustrates four distinct C-C couplings. Figure 3a shows the *trans* acetylene-vinyl coupling on a surface top site, which activation energy amounts to a relatively low value of 1.03 eV. It

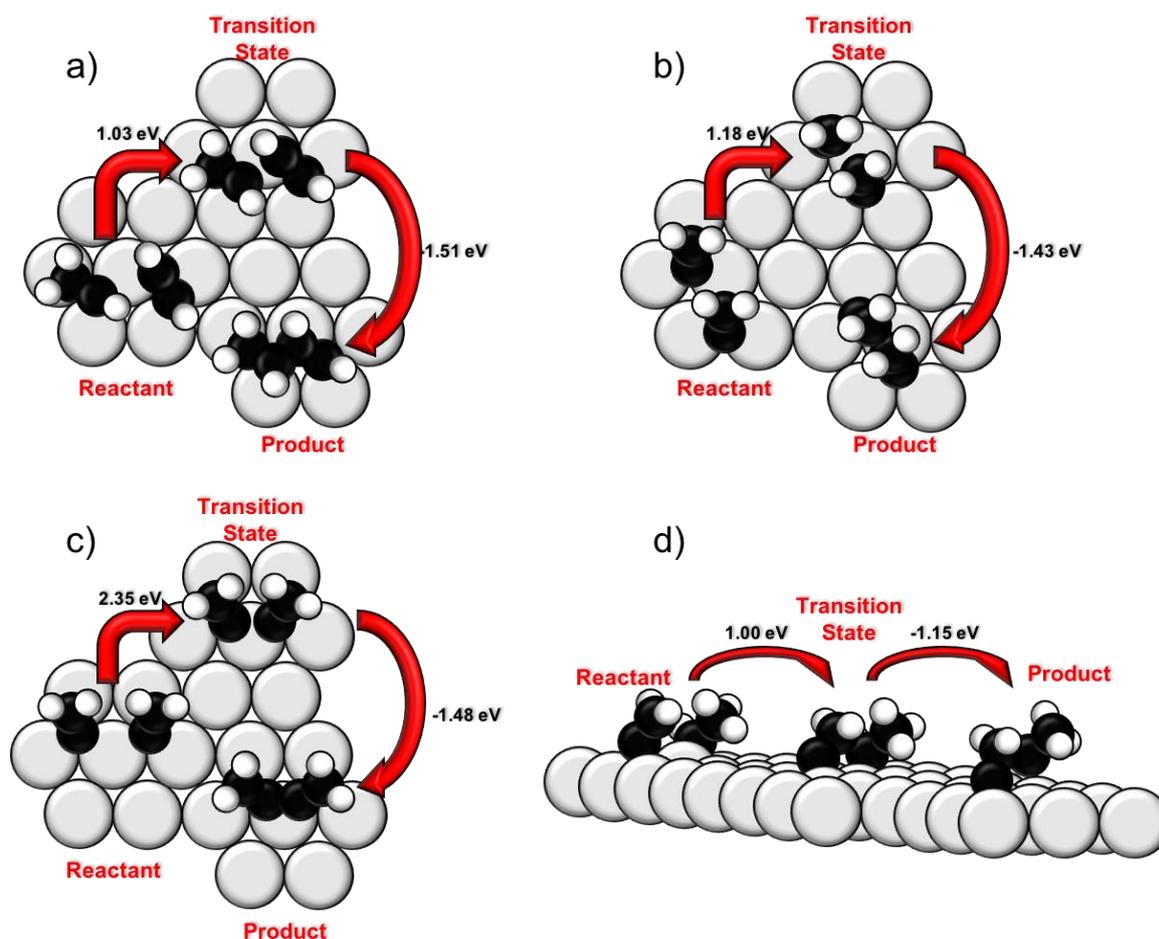


Figure 3. Selected, typical C-C coupling steps between C₂ surface species. a) CHCH+CHCH₂ to give CHCHCHCH₂; b) and CCH₂+CCH₂ to give CCH₂CH₂C; c) CCH₂+CCH₂ to give CH₂CCCH₂; d) CHCH₃+CCH₂ to give CH₃CHCH₂C.

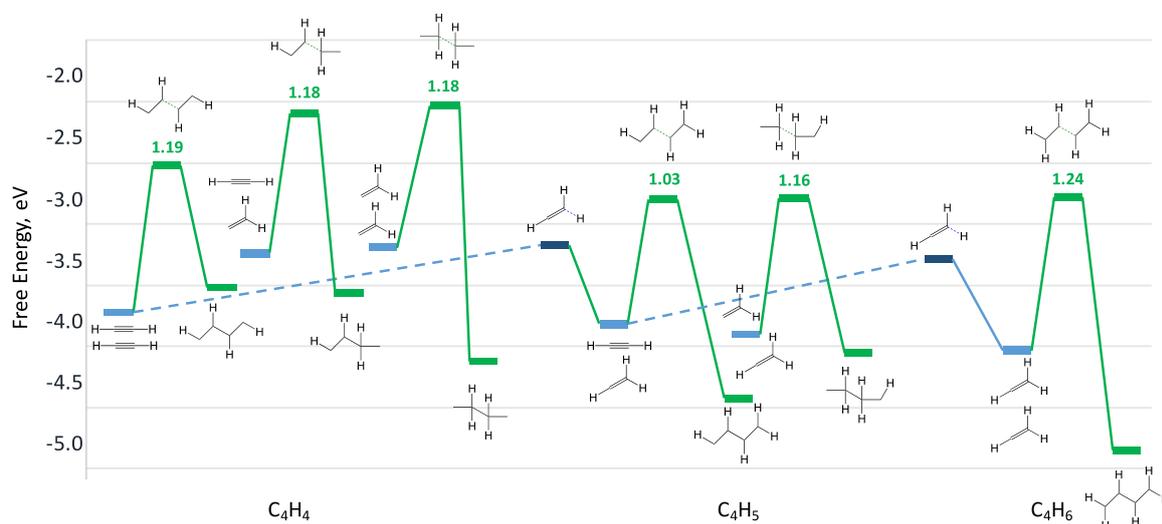


Figure 4. Key oligomerisation steps forming C_4 intermediates by coupling of C_2 fragments. All energies are referenced to acetylene and hydrogen in the gas-phase under standard conditions. For the paths between C_2 intermediates of the same stoichiometry, see Figure 2.

can be seen how closely the reactant positions and especially the transition states resemble the final product structure. The acetylene carbon-carbon axis stays nearly parallel to the Pd-Pd bond and the methylidyne group in vinyl remains completely unperturbed on top of one Pd atom. Only the methylene group is displaced during the transformation. This mode of C-C coupling is typical for fast reactions during which the π systems remain parallel to the surface.

Coupling carbon atoms showing little or no interaction with the catalytic surface is a process with a moderate barrier; In contrast, those steps requiring strongly bound carbons to be desorbed from the surface are disfavoured (see Table 2). This is demonstrated with the coupling of two CCH_2 species. An activation barrier of 1.18 eV is obtained for the coupling of two vinylidene molecules via their methylene functional groups. The spatial arrangement exhibited by this transformation is shown in Figure 3b. No significant alteration of the adsorbate/surface interaction is predicted in forming CCH_2CH_2C . This sharply contrasts with the alternative coupling to CH_2CCCH_2 . Here, the partial cleavage of the triple bond between the CCH_2 sp carbon atoms and the catalyst is needed twice. As shown in Figure 3c, the molecules are coupled on top of a surface Pd atom, resulting in a high activation energy of 2.35 eV.

The reaction of $CHCH_3$ and CCH_2 to give CH_3CHCH_2C has the lowest activation energy, 1.00 eV. Figure 3d shows the mechanism, in which $CHCH_3$ is lifted from the surface and couples to the spectating $-CH_2-$ group. Since $CHCH_3$ is rather unstable (it dehydrogenates with a barrier <0.2 eV), its partial desorption is rather easy, agreeing with the fact that weakly bound functional groups are activated more easily.

Regarding thermodynamics on the Pd surface, formation of C_4 molecules such as *cis*- and *trans*-butene, butadiene, and butane are among the most exothermic reactions (see Table S2). The balance between formation of alkyl chains and desorption of strongly bound functional groups drives the formation of less common oligomeric precursors. In this perspective, the most exergonic reaction comes from the coupling of ethyl and vinylidene to give $CH_3CH_2CH_2C$; the gain in free energy is -1.05 eV. Close to this value is the free energy gain for 1,3-butadiene formation, -0.86 eV. Reactions involving ethylene are mostly endergonic, spanning an interval between -0.04 eV and 1.05 eV. Together with the relatively large activation energies for ethylene couplings (>1.3 eV), the absence of

C₄ fragment is 0.97 eV, which is competitive with the hydrogenation of acetylene (1.08 eV) and vinyl (0.96 eV). Beyond this average value, we found cases of extremely fast hydrogenations; 22 intermediates are hydrogenated faster than vinyl and six have an activation barrier lying under 0.70 eV. The fastest C₄ hydrogenation is, with $\Delta E^\ddagger=0.31$ eV, the hydrogenation of CHCHCHCH₃ to CHCHCH₂CH₃. The C₄ coupling products starting from vinylidene (CCH₂) most likely go through CHCH₂CH₂C, followed by the hydrogenation of the terminal carbon, leading to CHCH₂CH₂CH, which is then converted to CHCH₂CH₂CH₂. The other important route starts with the most commonly formed oligomer, CHCHCHCH₂, which hydrogenates at the methylene carbon atom. The resulting CHCHCHCH₃ surface intermediate then quickly hydrogenates to CHCHCH₂CH₃.

The observation that C₄ hydrogenations are competitive with C₂ hydrogenations is not only problematic in terms of hydrogen atom efficiency, but also problematic since acetylene and CHCH₂ are involved in the oligomerization reaction. Hence, if the C₂ species are not hydrogenated fast enough to ethylene, they are more likely to oligomerize. The oligomers in turn increase the hydrogen demand for C₄ hydrogenations. Hence, a significant hydrogen partial pressure might be necessary to maintain a sufficient surface hydrogen coverage. The hydrogen coverage might also help to slow down the C-C coupling by isolating C₂ species from each other through adsorbed hydrogen atoms.

A striking feature of the considered activation and reaction energies is the absence of correlation between energy values and chemical functionalities involved. sp carbon atoms are as prone to hydrogenation as methylene groups. In an attempt to define a descriptor to linearly scale the activation energies, other parameters such as the reactants' distance in transition states and the energy of the reactions were screened; in any event, no correlation was found even after removing the most important "outliers". This might be due to the consensus of many causes which could only be addressed by a non-trivial descriptor.

Ensemble Requirements for Key Reaction Steps

A key question for the design of optimal catalysts for hydrogenation of acetylene in order to improve selectivity and avoid unwanted oligomers as green oil is to determine the ensemble of atoms required for each elementary step in the mechanistic network. One way to depict this active site requirement is to determine the surface atoms that are formally occupied by a given intermediate or transition state. As we have established in a previous study³¹ the projection of atomic radii on a graphical lattice including top and 3 fold hollow sites is an accurate method to assign sites to a given adsorbate. Here, we apply this approach to four key species in the mechanism: (a) acetylene, which needs to be adsorbed strongly for hydrogenation (b) CHCH₂, which is a certain intermediate in the selective hydrogenation towards ethylene (c) CHCH₃, which might proceed to unselective ethane formation and cause surface poisoning through CCH₃ accumulation, and (d) the transition state for the C-C bond formation between acetylene and CHCH₂. The latter process has a competitive barrier (1.03 eV) compared to that for the hydrogenation to ethylene (0.96 eV) and represents the path leading to unwanted oligomers. The corresponding site requirements are given in Figure 6.

In agreement with our previous study on the Pd-Ag alloy,³² Figure 6a indicates that there are two top sites (red circles) which are crucial for the binding energy of acetylene. Furthermore, the C-C bond is close to the hollow site, defined by a third top site. The latter is however not directly occupied by the molecule. This implies that an ensemble of 2 Pd atoms will strongly bind acetylene, with a third surface atom modulating the strength of the interaction (Ag for weaker binding, Pd for strong binding).

Three hollow sites are directly occupied (green circles).

Moving to the key intermediate for hydrogenation, CHCH_2 , only one top site is occupied by the adsorbate (see Figure 6b). However, similarly to acetylene, a hollow site is found below the C-C bond. Hence, the neighbouring atoms of this hollow site are expected to have a strong impact on the stability of this intermediate. Since, however, the site requirements of CHCH_2 are rather less (1 instead of 2 top sites) than more stringent compared to acetylene, the first hydrogenation reaction should be able to proceed on the very same ensemble as for acetylene adsorption, implying that surface diffusion is not required for hydrogenation.

Looking at the third intermediate (CHCH_3), leading to total hydrogenation, Figure 6c suggests that it is also stabilized by essentially the same ensemble of sites as the previous intermediates. Hence, it seems difficult to improve the selectivity towards ethylene, versus ethane, just by controlling the ensemble of Pd atoms: one should rely instead on electronic effects weakening the adsorption of ethylene and favouring its desorption versus further hydrogenation. As one might have expected, the oligomerization process (Figure 6d) requires a larger ensemble of Pd atoms: Four top sites (and six hollow sites) are occupied by the transition state. Hence, if one of the top sites is exchanged for a less reactive metal (e.g., Ag), a significant increase in the activation energy can be expected. This explains the decrease of oligomer production when a PdAg alloy is used, instead of Pd, although other alloys of Pd with less reactive atoms with carbon (as PdAu) should lead to the same effect.¹⁷

To conclude, the site requirements for the main C_2 species are all quite similar. Hence, only specific electronic effects from ensemble composition could affect the relative barriers and intermediate energies and hence the selectivity toward ethylene. The C-C coupling, on the other hand, required a markedly larger ensemble and oligomer formation is likely to be strongly slowed down by catalysts composed of exclusively Pd_2 (or maybe Pd_3) ensembles embedded in a less reactive metal. Our analysis has hence identified the configuration of the sites affecting the stability of intermediates and transition states, which could open to more detailed selective catalyst design using transition metal alloys.

Conclusions

This paper explored the vast network of chemical reaction steps leading to coupling by-products during selective hydrogenation of acetylene on a Pd(111) surface. Computational models for the hydrogenation and coupling of C_2 intermediates were developed, and later extended to the hydrogenation of resulting C_4 intermediates.

Hydrogenation of C_2 intermediates is in general faster than coupling; nevertheless, some dimer formation steps are found to compete with hydrogenation, especially those not requiring cleavage of highly stable carbon-surface bonds; these include coupling of early intermediates, such as acetylene ($CHCH$) and vinyl ($CHCH_2$), and of vinylidene (CCH_2) when the strongly bound carbon atom is not involved in the reaction. Ethylene is found to be kinetically inert towards dimers formation. This suggests that its role in the formation of green oil is of little or no importance. Ethylidyne shows exceptional stability and is kinetically inert, which agrees well with previous research.

Hydrogenation of C_4 intermediates is kinetically favoured on average. C_4 intermediates are likely to consume a significant fraction of the surface hydrogen species; this could promote further coupling of early intermediates, which in turn would exhaust the surface from adsorbed hydrogen. This emphasizes that the detrimental effect of “green oil” plays not only on the selectivity, but even on the overall reaction mechanism.

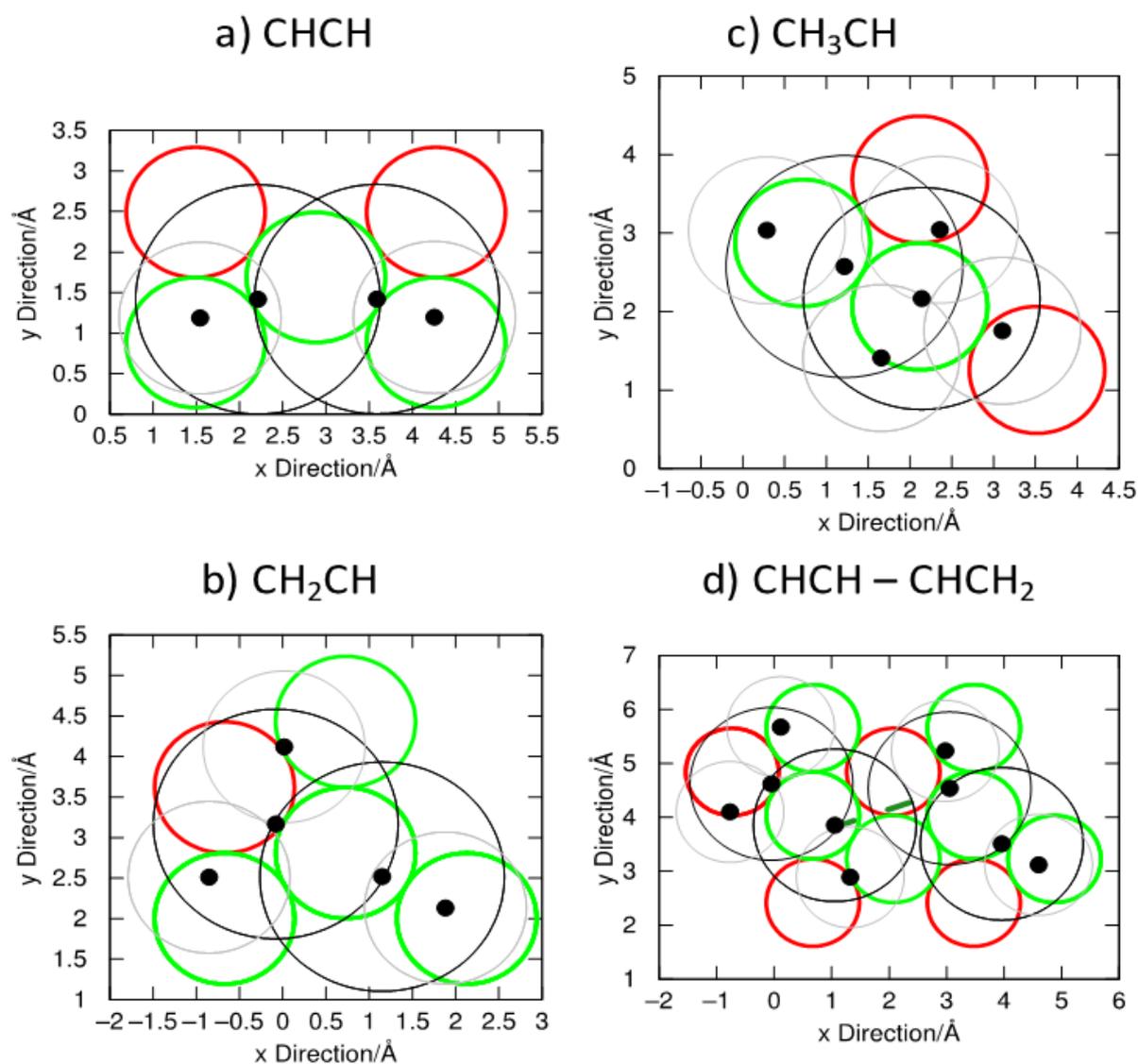


Figure 6. Site requirements for key intermediates on a Pd(111) surface. Occupied top sites (in red) and 3-fold hollow sites (in green) are highlighted. The atomic nuclei of the adsorbate are given as black dots and the size of the atoms indicated as black and grey circles for carbon and hydrogen, respectively. The bond being formed in d) is indicated by a dashed green line.

Coupling steps were also investigated for their ensemble requirements, i.e., the size and geometry of a surface site required for reaction. Calculations showed that an ensemble of 4 sites of Pd is required to permit acetylene oligomerization. This implies that Pd-based alloyed catalyst should really be well-mixed in order to prevent green oil formation. This insight is crucial for the design of novel catalysts for acetylene selective hydrogenation.

Supporting Information available

Reaction and activation energies and all geometries are provided.

Corresponding Author

stephan.steinmann@ens-lyon.fr, sautet@ucla.edu

Acknowledgements

Ab initio calculations were performed using the local HPC resources of PSMN at ENS-Lyon. *Selective Hydrogenation of Acetylene in an Ethylene-Rich Flow: Insights from Molecular Modelling* is a research project funded by TOTAL Refining & Chemicals and the ANRT. This work was granted access to the HPC resources of CINES and IDRIS under the allocation 2014-080609 made by GENCI.

References

- (1) Nicolaou, K. C.; Bulger, P. G.; Sarlah, D. *Angewandte Chemie - International Edition*. 2005.
- (2) Yin, L.; Liebscher, J. *Chem. Rev.* **2007**, *107*, 133–137.
- (3) Borodziński, A.; Bond, G. C. *Catal. Rev.* **2006**, *48* (2), 91–144.
- (4) Bos, A. N. R.; Westerterp, K. R. *Chem. Eng. Process.* **1993**, *32*, 1–7.
- (5) Yayun, L.; Jing, Z.; Xueru, M. *Proc. Jt. Meet. Chem. Eng., Chem. Ind. Eng. Soc. China. AIChE. Beijing* **1982**, *2*, 688–702.
- (6) Gentle, M.; Muettterles E., L. *J. Chem. Phys.* **1983**, *87*, 2469–2472.
- (7) Gandman, Z. E.; Aerov, M. E.; Men'shchikov, V. A.; Getmantsev, V. S. *Int. Chem. Eng.* **1975**, *15* (1), 183–185.
- (8) Sautet, P.; Cinquini, F. *ChemCatChem* **2010**, *2* (6), 636–639.
- (9) Armbrüster, M.; Behrens, M.; Cinquini, F.; Föttinger, K.; Grin, Y.; Haghofer, A.; Klötzer, B.; Knop-Gericke, A.; Lorenz, H.; Ota, A.; Penner, S.; Prinz, J.; Rameshan, C.; Révay, Z.; Rosenthal, D.; Rupprechter, G.; Sautet, P.; Schlögl, R.; Shao, L.; Szentmiklósi, L.; Teschner, D.; Torres, D.; Wagner, R.; Widmer, R.; Wowsnick, G. *ChemCatChem* **2012**, *4* (8), 1048–1063.
- (10) Torres, D.; Cinquini, F.; Sautet, P. *J. Phys. Chem. C* **2013**, *117*, 11059–11065.
- (11) Sheth, P. A.; Neurock, M.; Smith, C. M. *J. Phys. Chem. B* **2003**, *107*, 2009–2017.
- (12) Pallassana, V.; Neurock, M.; Lusvardi, V. S.; Lerou, J. J.; Kragten, D. D.; Santen, R. A. Van. *J. Phys. Chem. B* **2002**, No. 111, 1656–1669.
- (13) Moskaleva, L. V.; Chen, Z. X.; Aleksandrov, H. a.; Mohammed, A. B.; Sun, Q.; Rösch, N. *J. Phys. Chem. C* **2009**, *113* (6), 2512–2520.

- (14) Stacchiola, D.; Tysoe, W. T. *J. Phys. Chem. C* **2009**, *113* (19), 8000–8001.
- (15) García-Mota, M.; Bridier, B.; Pérez-Ramírez, J.; López, N. *J. Catal.* **2010**, *273* (2), 92–102.
- (16) Mei, D.; Sheth, P.; Neurock, M.; Smith, C. *J. Catal.* **2006**, *242* (1), 1–15.
- (17) López, N.; Vargas-Fuentes, C. *Chem. Commun. (Camb)*. **2012**, *48* (10), 1379–1391.
- (18) Yang, B.; Burch, R.; Hardacre, C.; Hu, P.; Hughes, P. *J. Phys. Chem. C* **2014**, *118* (3), 1560–1567.
- (19) Blochl, P. E. *Phys. Rev. B* **1994**, *50* (24), 17953–17979.
- (20) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59* (3), 1758–1775.
- (21) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47* (1), 558–561.
- (22) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54* (16), 11169–11186.
- (23) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (24) Steinmann, S. N.; Corminboeuf, C. *J. Chem. Theory Comput.* **2011**, *7*, 3567–3577.
- (25) Pack, J. D.; Monkhorst, H. J. *Phys. Rev. B* **1976**, *13* (12), 5188–5192.
- (26) Methfessel, M.; Paxton, a. T. *Phys. Rev. B* **1989**, *40* (6), 3616–3621.
- (27) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113* (22), 9978–9985.
- (28) Fleurat-Lessard, P.; Dayal, P. 2005.
- (29) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111* (15), 7010–7022.
- (30) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications, 1960.
- (31) Vignola, E.; Steinmann, S. N.; Vandegehuchte, B. D.; Curulla, D.; Stamatakis, M.; Sautet, P. *J. Chem. Phys.* in press 2017
- (32) Vignola, E.; Steinmann, S. N.; Vandegehuchte, B. D.; Curulla, D.; Sautet, P. *J. Phys. Chem. C* **2016**, *120*, 26320–26327.
- (33) Studt, F.; Abild-Pedersen, F.; Bligaard, T.; Sørensen, R. Z.; Christensen, C. H.; Nørskov, J. K. *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (48), 9299–9302.

Chapter 6. A Machine Learning Approach to Graph-Theoretical Cluster Expansions of the Energy of Adsorbate Layers

Emanuele Vignola,^{†,‡} Stephan N. Steinmann,[†] Bart D. Vandegehuchte,[†] Daniel Curulla,[†] Michail Stamatakis,^{#,*} and Philippe Sautet^{†,&,*}

[†]Univ Lyon, ENS de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France

[‡]Total Research and Technology Gonfreville, BP 27, F-76700 Harfleur, France

[†]Total Research and Technology Feluy, Zone Industrielle Feluy C, Seneffe, Belgium

[#]Department of Chemical Engineering, University College of London, Torrington Place, London WC1E7JE, United Kingdom

[&]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, CA 90095, United States

*e-mail: m.stamatakis@ucl.ac.uk, sautet@ucla.edu

Published on the Journal of Chemical Physics

Doi: <http://dx.doi.org/10.1063/1.4985890>

Publication Date: August 7, 2017

The accurate description of the energy of adsorbate layers is crucial for the understanding of chemistry at interfaces. For heterogeneous catalysis not only the interaction of the adsorbate with the surface, but also adsorbate-adsorbate lateral interactions significantly affect activation energies of reactions. Modeling the interactions of adsorbates with the catalyst surface and with each other can be efficiently achieved in the cluster expansion Hamiltonian formalism, which has recently been implemented in a graph-theoretical kinetic Monte Carlo (kMC) scheme to describe multi-dentate species. Automating the development of the cluster expansion Hamiltonians for catalytic systems is challenging and requires the mapping of adsorbate configurations for extended adsorbates onto a graphical lattice. The current work adopts machine learning methods to reach this goal. Clusters are automatically detected based on formalized, but intuitive chemical concepts. The corresponding energy coefficients for the cluster expansion are calculated by an inversion scheme. The potential of this method is demonstrated for the example of ethylene adsorption on Pd (111), for which we propose several expansions, depending on the graphical lattice. It turns out that for this system the best description is obtained as a combination of single molecule patterns and a few coupling terms accounting for lateral interactions.

I. INTRODUCTION

Many chemical production processes for fuels, plastics, fine chemicals, fertilizers and pharmaceuticals are catalytic in nature.¹ Catalyst development remains crucial to reduce the energy intensive nature of the processes and improve the chemical selectivity of the reaction. Three performance criteria can be identified: activity, selectivity, and stability. While catalyst development is mostly driven by experiments, computational studies provide valuable insights into the nature of

the active sites and the reaction mechanism. They have become reliable over the last decade with respect to predictive power, computational speed and can now actively participate in the catalyst design.²⁻⁷

Single crystal surfaces, approximated as a periodic repetition of a unit cell (typically in the order of 10-20 surface atoms) are commonly applied to modeling heterogeneous catalysts and their energy is often described by Density Functional Theory (DFT).^{8,9} This approach provides a fair trade-off between accuracy and computational speed. Surface structures, adsorption energies and reaction mechanisms can be investigated rather easily at low coverage, e.g., one adsorbate per surface unit cell. At higher coverage, however, lateral interactions between adsorbates develop and cause adsorbate rearrangements on the surface, requiring larger and therefore more complex model supercells.¹⁰ The ordering of adsorbate structures depends on the adopted supercell. For instance, consider the adsorption of 3 molecules on 4 indistinguishable sites.¹¹ Only one (symmetry distinct) arrangement is possible in the smallest unit cell. However, a supercell with 16 sites and 12 adsorbates suggests a much higher number of possible arrangements, making the calculations much more complex especially when chemical kinetics are involved. As this task becomes insurmountable for explicit first principles models, approximate methods are recommended for this endeavor.

Recent years have witnessed a surge of lattice based (kinetic) Monte Carlo (kMC) methods which rely on model Hamiltonians parameterized by DFT.¹²⁻¹⁴ The advantage of model Hamiltonians is that they yield results much faster than the underlying DFT computations, while they are able to properly assess the configurational average and evolution of a surface under reactive conditions. In principle, the model Hamiltonian needs to account for three energy contributions: i) The energy of the surface, which includes deformation energies, ii) the interaction of adsorbates with the surface and iii) the interaction between adsorbates (also called lateral interactions). In this work, we focus on ii) and iii), with the surface distortion energy being implicitly incorporated in the interaction energies. Such a model is appropriate for surfaces that do not significantly restructure under reactive conditions. The *interaction model* for adsorbate-surface systems aimed at in this work is rather general and could be applied to surface reconstruction as well.

A two-dimensional Ising-type model is typically selected as the basic model Hamiltonian.^{15,16} An Ising-type model is defined as a set of lattice points in a given spatial arrangement which are mapped onto a set of spin-like values. The energy is derived from an effective Hamiltonian, which incorporates the effect of an external field and the coupling between different spins, e.g. given a positive coupling constant, if a site has positive spin, the neighboring site “prefers” a negative spin. The energy thus depends on the configuration of spins within the lattice and on the intensity of an external field. Such Ising-type models (and the closely related lattice-gas Hamiltonians) are very effective in the description of phase transitions of multicomponent materials.^{17,18} They have also been employed to represent statistical models of adsorbed layers.¹⁹ In general, the partition function of Ising-type models is not known analytically except in special cases.²⁰ Yet, Monte Carlo algorithms have been found to be very effective in converging to the ground state, or in evaluating ensemble averages. The parameters of the model Hamiltonian have to be known beforehand and are either determined from truly *ab initio* energy evaluations or, in some systems, fitted to reproduce experiments.

Model Hamiltonians with good accuracy can be established for many problems, especially bulk materials. They are also applicable to systems larger than the scope of direct DFT computations. The well-known ATAT suite of programs by Ceder and co-workers automatically computes phase diagrams of alloys, based on model Hamiltonians that are fitted to DFT data.^{21,22,23} The ATAT tool pertains to equilibrium phenomena, not to the kinetics of a system. However, for catalysis, both the surface structure, i.e., coverage under realistic conditions, and the kinetics of the corresponding processes are

of great significance. kMC has proven its potential in this respect having been successfully applied to a large variety of catalytic systems.^{24–37} A recent kMC framework is the graph-theoretical *Zacros* code of Stamatakis *et al.*^{38,39} which is particularly suited for applications in catalysis, since it is able to deal with multidentate species, complex elementary steps (e.g. involving more than two sites in specific geometric arrangements), and reaction barriers changing in the presence of spectator species that exert lateral interactions. As is common in lattice-based kMC, the evolution of the system in time is represented as a set of transformations of a graph. The structural parameters of a catalytic surface are mapped to the graph’s vertices, with occupation variables representing the presence of adsorbates.

Establishing the model Hamiltonian of such a complex system is more difficult than for a bulk alloy because the number of possible configurations is enormous (even though finite). Molecular species bond specifically to a surface layer and can adopt many geometries. Furthermore, the link between the adsorbate geometry and its pattern on the graphical lattice (synonymously called “cluster” herein) is not trivial. In the past, users had to construct a model Hamiltonian by carefully analyzing the computed configurations, tracking the system’s relevant features, and mapping them to a graphical lattice. Herein, we instead rely on a *machine learning*⁴⁰ approach which is based on pattern recognition. Once the recognition selection process has been defined, the state of the lattice is inferred automatically without significant knowledge on the system. Our contribution first sketches the theoretical background and then proposes a chemically meaningful, operative framework to automatically determine model Hamiltonians in terms of a cluster expansion, given a set of configurations, their adsorption energies and a graphical lattice. We apply our approach to the adsorption of ethylene on Pd(111) as a particularly important model system in heterogeneous catalysis.

II. THE CORRESPONDENCE BETWEEN REAL AND GRAPHICAL INTERACTIONS

Consider a system of M atoms. Graph theory implies that the total energy of the system can be expanded in a series of atomic interaction energy terms between n atoms, with $n \leq M$,⁴¹ taking the form of the following equation

$$\begin{aligned}
 E_{tot}^{atm}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M) &= \sum_i W_i(\mathbf{r}_i) + \sum_{i>j} W_{ij}(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i>j>k} W_{ijk}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) \\
 &+ \sum_{i>j>k>l} W_{ijkl}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k, \mathbf{r}_l) + \dots
 \end{aligned} \tag{1}$$

where the indexes i, j, k, l, \dots go from 1 to M , $W_i(\mathbf{r}_i)$ is the (electronic or potential) energy of a given atom at location \mathbf{r}_i , and $W_{ijkl\dots}$ is the energy contribution of the i, j, k, l, \dots group of atoms which is a function of their location in (continuous) space, $\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k, \mathbf{r}_l, \dots$. These groups of atoms are also called

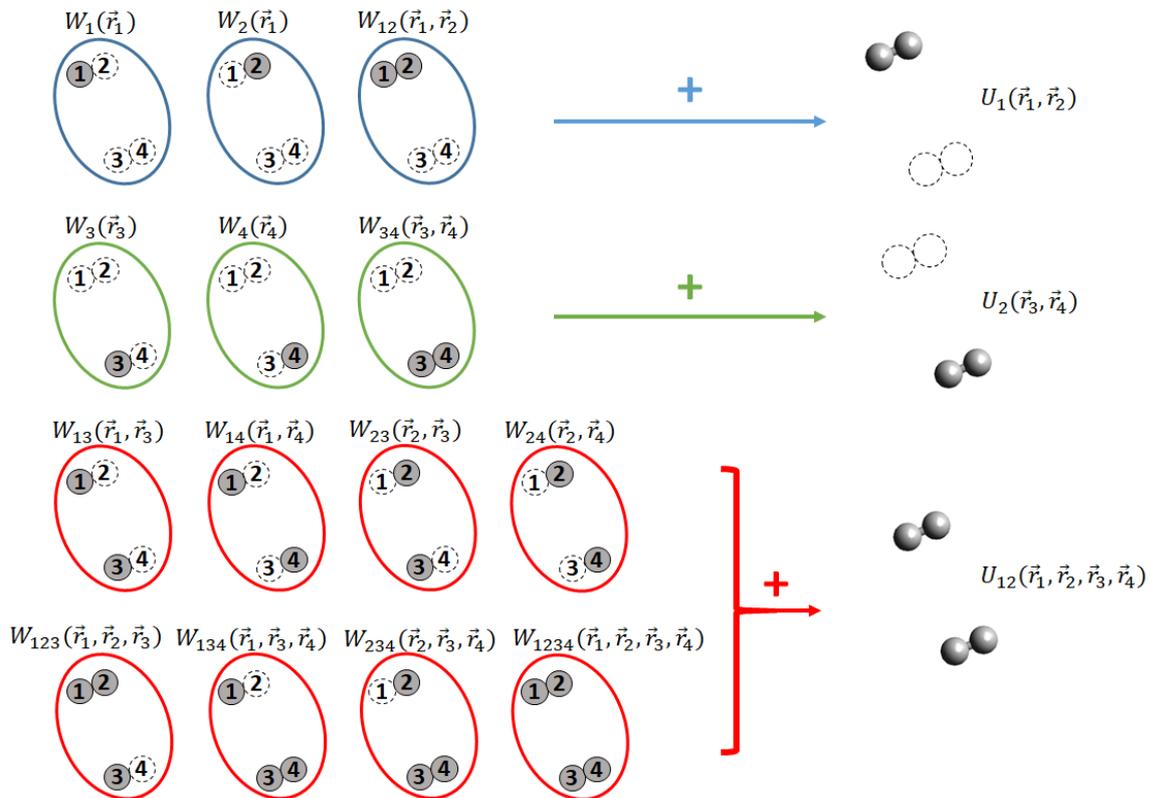


Figure 1. From atomic to molecular cluster expansions. Two top rows: 1-body clusters, bottom two rows: 2-body clusters, where “body” refers to a molecule.

“clusters”.

Though this expansion is exact, it is not well suited to describe molecular systems. It can, however, be transformed from an atomic to a molecular expression. As an example, consider the interaction of two molecules (Figure 1 depicts the case of two diatomics). The energy U_α of a molecule or ‘body’ α is obtained as the sum of all terms involving atoms from that body:

$$U_\alpha(\{\mathbf{r}\}_\alpha) = \sum_{i \in \alpha} W_i + \sum_{i \in \alpha, j \in \alpha} W_{ij} + \dots; \quad (2)$$

where we use curly brackets to indicate the set of atomic coordinate (\mathbf{r}) belonging to molecule α and we dropped the argument (r_i) for sake of clarity. The interaction energy between two molecules α and β is similarly given by:

$$U_{\alpha\beta}(\{\mathbf{r}\}_\alpha, \{\mathbf{r}\}_\beta) = \sum_{i \in \alpha, j \in \beta} W_{ij} + \sum_{i \in \alpha, j \in \alpha, k \in \beta} W_{ijk} + \sum_{i \in \alpha, j \in \beta, k \in \beta} W_{ijk} + \dots \quad (3)$$

This allows us to express the total energy of a system in terms of molecular cluster energies, i.e., in this case a “cluster” refers to a group of molecules:

$$E_{tot}^{mol}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M) = \sum_{\alpha} U_{\alpha}(\{\mathbf{r}\}_{\alpha}) + \sum_{\beta>\alpha} U_{\alpha\beta}(\{\mathbf{r}\}_{\alpha}, \{\mathbf{r}\}_{\beta}) + \dots \quad (4)$$

where the first term accounts for the energy of each molecule α in the particular geometry $\{\mathbf{r}\}_{\alpha}$, the second for the interaction energy between all pairs of molecules, the following term for triplets, and so on. This formulation was also deduced in dedicated books.⁴²

The present approach is now specifically applied to adsorption energies. We divide atoms into two sets: the ones of adsorbates (ads, N atoms) and those of the catalyst surface (cat, $M-N$ atoms). The adsorption energy is then given by

$$\Delta E_{ads} = E_{tot}(ads@cat) - E_{tot}(ads) - E_{tot}(cat) \quad (5)$$

For adsorption on relatively rigid surfaces, we assume the coordinates of the catalyst as fixed during the adsorption event. This assumption is standard in the theory of lattice statistics.⁴³ It is not compulsory and surface reconstructions can be described even with lattice based model Hamiltonians.⁴⁴ Furthermore, we account for the deformation energy of both the catalyst and the molecules in their respective interaction energy, as these deformations are rather local. Doing so, we can focus on the adsorption energy only:

$$\Delta E_{ads}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \approx E(\{\mathbf{r}\}_{\alpha}, \dots, \{\mathbf{r}\}_{\omega}) \quad (6)$$

where $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ are the coordinates of the atoms within adsorbed molecules labeled by $\alpha \dots \omega$. In this case, the interaction energy is given by formally two or more body terms:

$$E(\{\mathbf{r}\}_{\alpha}, \dots, \{\mathbf{r}\}_{\omega}) = \sum_{\alpha} [U_{\alpha\tau}^{int}(\{\mathbf{r}\}_{\alpha})] + \sum_{\beta>\alpha} [U_{\alpha\beta\tau}^{int}(\{\mathbf{r}\}_{\alpha}, \{\mathbf{r}\}_{\beta})] + \dots \quad (7)$$

where τ stands for the catalyst. In this equation the first term accounts for the adsorption energy of “isolated” molecules and the second term for lateral interactions.

In the following, we describe how equation 7 is transformed into a cluster expansion on a fixed lattice and how its energy coefficients are determined. The first essential step is mapping Equation 7 from real space onto a lattice in order to build Ising-type (or lattice gas) Hamiltonians. This coarse-graining is performed by using a two-dimensional lattice of points in real space. In the framework of the Zacros code, strictly positive occupation variables are used instead of pseudo-spins to identify what kind of species occupy which lattice point, so that molecular configurations have a lattice representation. Let S be the set of occupation variables σ assigned to each molecular adsorption mode, i.e., σ and π -bound ethylene have a distinct occupation variable. Then, a mapping \mathcal{M} from the lattice coordinates to the set of occupations will generate the lattice configuration. Let s_v be the occupation variable relative to the lattice point at positions \mathbf{p}_v , where v runs over all lattice points; then, we can write the energy of each configuration as:

$$\hat{E}(s_1, s_2, \dots, s_V) \approx \sum_{\alpha} [\hat{U}_{\alpha\tau}^{int}(\mathcal{M}(\{\mathbf{r}\}_{\alpha}))] + \sum_{\beta>\alpha} [\hat{U}_{\alpha\beta\tau}^{int}(\mathcal{M}(\{\mathbf{r}\}_{\alpha}, \{\mathbf{r}\}_{\beta}))] + \dots \quad (8)$$

Where we have used the hat to indicate the transition from continuous (\mathbf{r}) to discrete (\mathbf{p}) space. The approximate equality comes from the “coarse-graining”, i.e., several configurations in real space can be projected into the same \mathbf{s} vector, which results in an average energy for $E(\mathbf{s})$:

$$\hat{E}(\mathbf{s}) = \langle E(\mathcal{M}(\mathbf{r})) \rangle_{\{\mathcal{M}(\mathbf{r})=\mathbf{s}\}} \quad (9)$$

where the brackets indicate the average over all acceptable realizations of a given configuration mapped into the same coarse-grained occupation vector representation, see curly brackets. As a result, the total energy may be written as:

$$\hat{E}(s_1, s_2, \dots, s_V) = \sum_{\mu} \hat{U}_{\mu}^{int}(\mathbf{s}) = \sum_{\mu} \varepsilon_{\mu} \quad (10)$$

Where the index μ stands for a collection of molecular indices $(\alpha, \beta \dots, \alpha\beta \dots)$, i.e., for the clusters that are encoded in \mathbf{s} , and ε_{μ} is the cluster energy coefficient relative to μ .

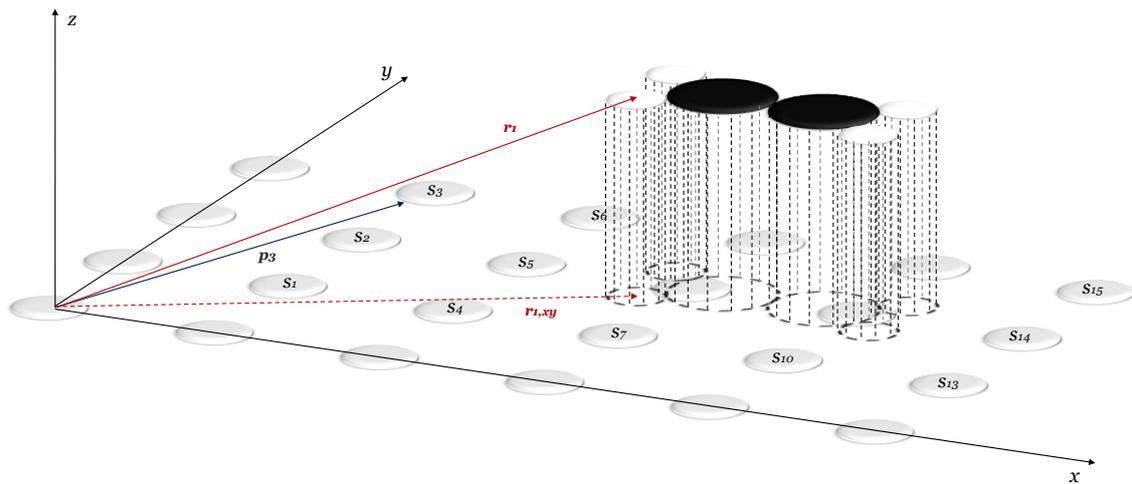


Figure 2. Illustration of \mathcal{M} , the projection of adsorbate coordinates $\{r\}$ on the lattice points at $\{p\}$, applied to the case of ethylene adsorption on a Pd surface.

Notice that redundancies may appear in equation 10, i.e., some of the clusters nested in the sums may be equal. Combining identical terms together, equation 10 is converted into equation 11

$$\hat{E} = \sum_{\kappa} \xi_{\kappa} \varepsilon_{\kappa} \quad (11)$$

Where ξ is the cluster multiplicity and κ runs over non-identical molecular clusters (also called patterns herein); in this scheme, ξ plays the role of a cluster basis function. Thus, κ ranges from 1 to $|P|$, the number of unique clusters detected in the set of configurations analyzed. The next section discusses the identification of the different clusters, with the following section devoted to the determination of the cluster energies (ε) through inversion.

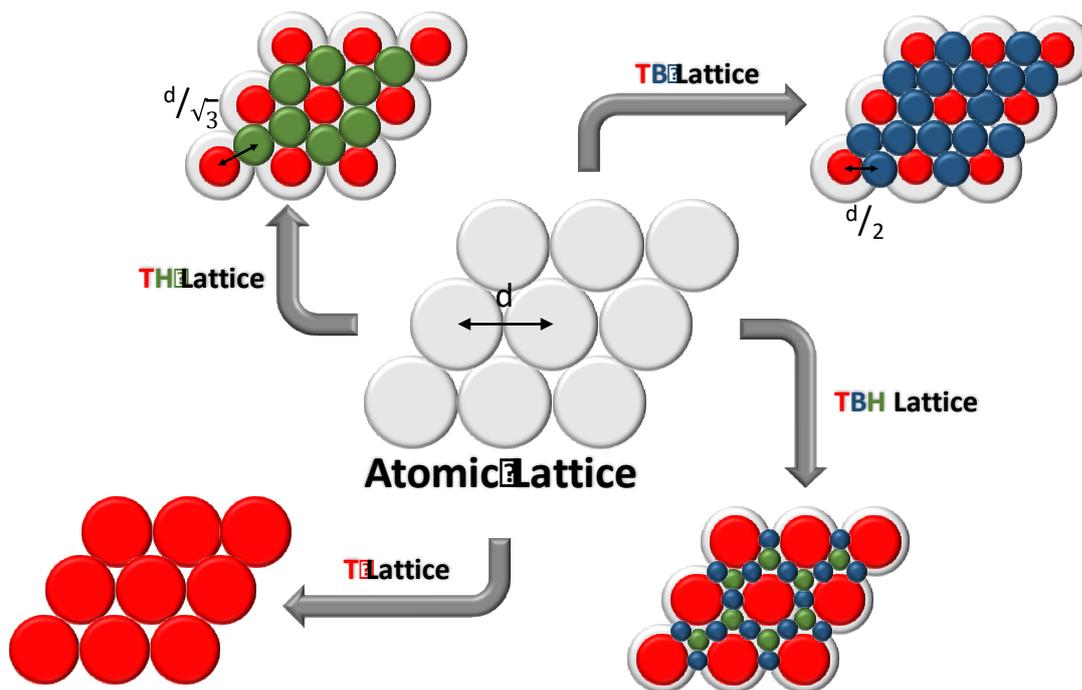


Figure 3. Illustration the construction of the different lattices: the size of the sites are chosen at the largest, non-overlapping circles, i.e., they touch at the midpoints. Therefore, the size of the top sites (in red) are largest in the T lattice and smallest in the TB lattice.

III. A RECOGNITION MEASURE TO MAP MOLECULAR CONFIGURATIONS INTO PATTERNS ON THE GRAPHICAL LATTICE

The projection of molecular coordinates onto the graphical lattice and the associated coarse-graining of the energy require the function φ . In other words, the projection φ allows us to assign the binding mode of molecules without arbitrary choices by the user, simply by applying a set of rules to identify the clusters, to which we also refer as patterns, on the lattice which correspond to a given adsorption mode (see Figure 2). In our case, we treat flat surfaces; hence, each vertex of the lattice is identified in Euclidean space by a 2-dimensional vector $\mathbf{p}_v = (x_v, y_v)$. Then, the coordinates of the atoms are projected on this plane. However, in general the atom will not be projected exactly on top of one of the vertexes, but only in its vicinity. Additionally, depending on the spacing of the lattice, a single atom might be large enough to occupy more than one site. Hence, we need to assign a size to the atoms and the vertexes to reliably identify the occupied vertexes by computing the overlap between the area of the atom and the vertex. Many ways for estimating the size of an atom exist.⁴⁵ Two extreme, but simple, choices are available: Van der Waals radii⁴⁶ (R_{vdw}) provide an upper limit on the size of an atom and therefore impose significant separations between neighboring adsorbates. Covalent radii⁴⁷ (R_{cov}), on the other hand, provide a lower limit of the size of atoms, which ensures that high coverages are accessible without leading to overlapping patterns. The area of the vertex could be chosen as the Wigner-Seitz cell which is obtained by connecting the mid-points between vertexes by straight lines. However, the overlap between the atoms and the Wigner Seitz cell and the atoms would be cumbersome to compute. Hence, we decided to use the largest, non-overlapping circles to approximate the Wigner Seitz cells as illustrated in Figure 3. Together with the circular atoms, the overlap $A_{i,v}$ of atom i with a given site v can be evaluated analytically:

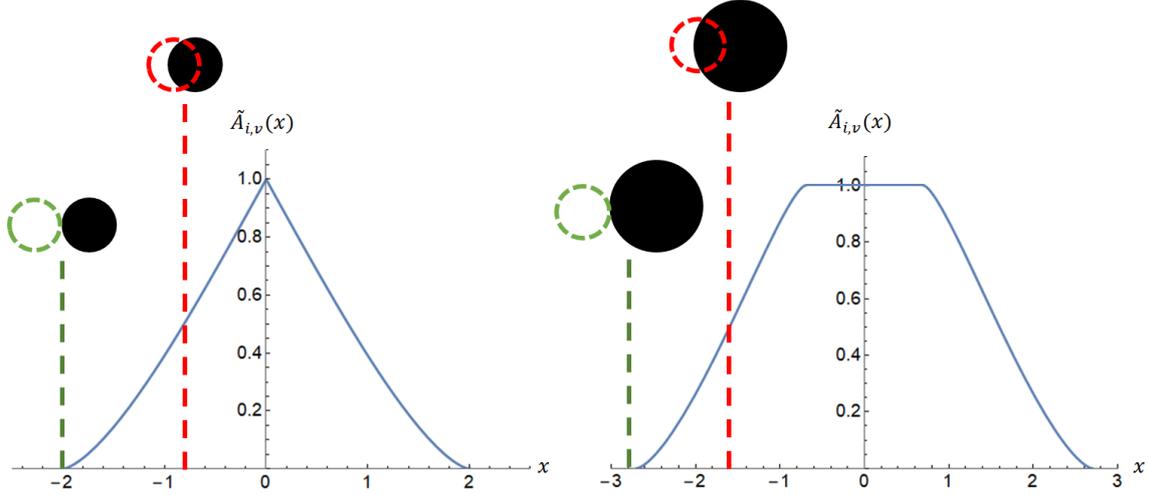


Figure 4. Normalized overlap as a function of relative position, starting at circles that touch each other, indicated by the green line. The red line indicates the 0.5 threshold. The scheme on the left refers to two circles with a radius of unity, the one on the right to one circle with a unity radius and one of 1.7 Å (vdW radius of carbon).

$$\begin{aligned}
A_{i,v} &= A(\mathbf{r}_i, \mathbf{p}_v) \\
&= \rho_i^2 \cos^{-1} \left(\frac{|\mathbf{r}_i - \mathbf{p}_v|^2 + \rho_i^2 - \rho_v^2}{2d\rho_i} \right) + \rho_v^2 \cos^{-1} \left(\frac{|\mathbf{r}_i - \mathbf{p}_v|^2 - \rho_i^2 + \rho_v^2}{2d\rho_v} \right) \\
&\quad - \frac{1}{2} \sqrt{(\rho_i - |\mathbf{r}_i - \mathbf{p}_v| + \rho_v)(|\mathbf{r}_i - \mathbf{p}_v| - \rho_i + \rho_v)(\rho_i + |\mathbf{r}_i - \mathbf{p}_v| - \rho_v)(\rho_i + |\mathbf{r}_i - \mathbf{p}_v| + \rho_v)}
\end{aligned} \tag{12}$$

where ρ_i and ρ_v are the radii of the atom and the vertex, respectively. To obtain a transferable measure, the overlap of each atom with vertexes is expressed as the fraction of the vertex that is covered by the atom:

$$\tilde{A}_{i,v} = \frac{A_{i,v}}{A_v} \tag{13}$$

where A_v is the area of the circle associated to vertex v , and $A_{i,v}$ the overlap with atom i . Note, that this framework only addresses the case of direct interaction with the catalyst surface and is not adapted for multi-layer adsorptions (e.g., adsorption on a H pre-covered surface), except if the definition of the catalyst lattice includes the information of, for example, an add-layer of hydrogen.

With a measure at hand, a recognition criterion can be finally cast on a quantitative basis. The $[0, 1]$ range of fractional overlap must be divided into classes: occupied and unoccupied. The criterion is stated as follows: a molecule \mathbb{M} recognizes and therefore occupies a site v if at least one of its atoms generates an overlap percentage greater than 50%, i.e. a vertex is assigned to a given molecule if at least one atom of that molecule produces a $\tilde{A}_{i,v} > 0.5$. Figure 4 shows the evolution of the overlap for the recognition of a site (circle in broken line) of radius 1.0 Å by an atom of identical size (full line circle) and by a carbon atom of van der Waals radius of 1.7 Å (left and right part of the scheme, respectively). The x-axis describes the distance between the centers of the site and the atom, with negative and positive values indicating that the site is on the “left” and “right” of the atom, respectively. The role of the radius of recognition is clearly captured, as the carbon atom shows an earlier recognition threshold (center separated by 1.6 Å; indicated by the red line) than the smaller site (centers separated by 0.8 Å). Of course, the size of the atoms and the vertexes need to be compatible, i.e., if $\rho_i < \sqrt{2}/2 \rho_v$ then the vertex cannot be recognized.

The recognition process can be summarized as a sum over the atomic coordinates of Heaviside functions of the overlap-threshold difference, $\theta(\tilde{A}_{i,v} - 0.5)$. Note that the geometrical threshold of

Multiplying equation 16 by the right-hand side of equation 17 and simplifying terms, the following solution for the pattern coefficients is obtained in the form of equation 18

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^+ \cdot \boldsymbol{E} \quad (18)$$

The solution of equation 18 gives a set of coefficients to represent Hamiltonians of adsorbed layers.

The cardinality of the set P could be, in principle, very large; this has the effect of requiring an even larger set of configurations to be computed. Cluster expansions are, however, known to converge fast, so that the expansion can be truncated at a finite distance and at low orders of many-body patterns.^{50–53} Additionally, the system of equations is likely to contain linear dependencies: for instance, there could exist a configuration whose pattern count ξ is a linear combination of the counts of two other configurations. This may, in principle, be tackled by avoiding superposition of configurations while sampling. However, in terms of statistics and recalling that we map a continuous space (adsorbates on the surface) into a discrete space (lattice occupations), removing such “redundancies” would preclude to being able to assess the soundness (introduced) error due to this discretization. The other source derives from the truncation of the expansion: removal of long-range pattern may in fact generate identical (short-range) configurations. Since no *a priori* knowledge of the cutoff or of the expansion order is available for an arbitrary catalytic system, this second source of linear dependencies can hardly be avoided. A *singular value decomposition* (SVD) during the Moore-Penrose pseudo-inversion is capable of circumvent the issues related to the linear dependencies.⁵⁴ SVD (with eigenvectors corresponding to an eigenvalue below machine precision being removed) was employed for all of the selected applications discussed in Section V.

While the Moore-Penrose pseudoinverse identifies a unique matrix corresponding to the solution with the least squares of the residuals ($\boldsymbol{E} - \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}$), one would like to find that the parameters have chemically reasonable values, since it would give confidence that the parameters are “transferable” to new configurations. In particular, we expect significant stabilizing single body patterns for chemisorbed molecules and less important, but repulsive two-body patterns. For this reason, we will also report maximum and minimum values for the one- and two-body contributions. In our experience, this expectation is satisfied for systems of equations that have a rank greater or equal to $|P|$.

A first step towards assessing the reliability, or rather the trustworthiness of a given least squares fit can be done “internally”. Since, as mentioned above, the “sampling problem” does not have an *a priori* solution, the importance of each point/configuration for the fit is an important measure. For example, if one parameter of the fit is connected to only one configuration, then this configuration is highly important (in statistics we would call it an outlier) for the fit and the value of this parameter is uncertain. The only way to properly deal with such a situation is to increase the sampling set to include more configurations, allowing us to define this specific parameter. The *Cook’s distance*^{55,56} of a configuration is convenient to measure its influence on the fit. The Cook’s distance can readily be obtained from the Moore-Penrose pseudoinverse. Firstly, the so-called *hat matrix* of the system \boldsymbol{H} , is defined as:

$$\boldsymbol{H} = \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}^+ \quad (19)$$

The diagonal element of the hat matrix H_{ii} , is known as the *leverage* of point i ; it measures how far a configuration stands from the average pattern occurrence. The leverage is used to standardize the differences between the configurational energy predicted by the model and the ones known from raw data; these normalized differences are known as *standardized residuals* e_i and are defined by the following equation

$$e_i = \frac{E_i - \sum_j \xi_{ij} p_j}{\sqrt{1 + H_{ii}}} \quad (20)$$

The Cook's distance D_i , of a given configuration is obtained as:

$$D_i = \frac{H_{ii} e_i^2}{|P| \sigma^2 (1 + H_{ii})^2} \quad (21)$$

Where σ^2 is the mean squared prediction/raw data residual. The Cook's distance is a tool to highlight the need for a more thorough acquisition of data/sampling. Rigorously, one should perform an F-test to identify the influential points and then analyze these points in order to decide whether or not they are to be included in the final analysis. While some authors suggest that $4/n$ (with n being the number of samples) should be the maximum Cook's distance,⁵⁷ others suggest to simply remove samples with Cook's distances above one.⁵⁸ In the following, we adopt the later cutoff. Since our data points are obtained from optimized DFT geometries, high Cook's distances cannot be equated to regular outliers: they still are physically meaningful. However, they can be considered outliers with respect to the training set. Hence the user has the choice to either increase the sampling around these points or to remove them, if he considers them irrelevant for his specific application.

Cook's distances measure the sparseness of the available sample. A more frequently resorted tool to analyse interpolation reliability is cross-validation.⁴⁰ Cross validation schemes partition the sample into a training and a validation set; regression is then performed using data contained into the former, while the latter is employed to obtain a residual (error estimate). The scheme is iterated as to cover all the possible training/validation partitions of the sample and residuals are summed as to get validation scores for each of the model's coefficients. Although no cross validation score was calculated in the present work, Cook's distances and cross-validation scores are complementary to address the robustness of models; assuming that clusters with high uncertainties and/or close to zero mean values during cross-validation should be either refined or removed, this would, indeed, help to "compress" the cluster expansions and making them more robust/transferable. Furthermore, this could be exploited to identify the configurations that require higher body patterns to be well described.

Pattern detection and determination of the cluster energies as describe above is implemented in an in-house FORTRAN90 code interfaced with LAPACK,⁵⁹ and will be made public in the near future.

V. APPLICATION TO THE ETHYLENE-PALLADIUM SYSTEM

In the previous sections, a complete framework to recognize patterns and parameterize a cluster expansion for adsorbate layers was proposed. In this section, this scheme is applied to ethylene adsorption on the Pd (111) surface.

Palladium is well-known in the catalysis community for its hydrogenation capacity, e.g., for fatty acid hydrogenation.⁶⁰ The hydrogenation of ethylene over Pd is a widely studied model system.⁶¹⁻⁶³ Previous theoretical studies have identified two single-molecule adsorption modes: The most stable mode is the di- σ mode in which the molecule is bonded to two palladium atoms, distorting its geometry quite strongly according to the Dewar-Chatt-Duncanson model^{64,65}. The latter model accounts for hydrocarbon-metal interactions in which a fraction of metal electron occupies the anti-bonding orbitals of the molecule.⁶⁶ The carbon-carbon bond distance extends by 0.12 Å (9%) and the hydrogen atoms are tilted by 9 degrees instead of being co-planar. The other adsorption mode is the π mode. This mode induces no significant change in molecular structure. The molecule merely interacts with a single palladium atom around its centre of mass.

V.1 THE TRAINING SET

Out of these two basic modes, we have constructed a diverse training set of 32 configurations on a $p(4\times 4)$ slab supercell of Pd(111), that span different coverages (1/4 monolayer, ML to 1/16 ML) and different relative orientations. The most important ones are represented on Figure 5. We started with the low-coverage (1/16 ML) situation, where the di- σ and π configurations are isolated in the surface unit cell (see configuration 7 and 29 in Figure 5). Then, we also included three configurations representative of the high coverage situation (1/4 ML): 4 π bound, 4 di- σ bound, and 2 π together with 2 di- σ bound molecules on the $p(4\times 4)$ cell. 31 configurations representative of intermediate coverage (1/8 ML) were constructed based on various combinations of di- σ and π coordination and neighbouring adsorption sites. Additionally, 10 configurations (3/16 ML) were constructed based on the 1/8 ML configurations by adding one molecule wherever possible. Note that the construction of this training set is merely exemplary and is neither complete nor ideal, but just serves to illustrate the application of the framework outlined above. In other words, some of the configurations included might rarely occur in reality while others, in particular high coverage structures, might be missing. We consider the question of the construction and validation of the training set as a separate topic, which will be addressed in the future. All configurations were optimized at the DFT level exploiting the PAW formalism.^{67,68} All computations were performed with VASP 5.3.3.^{69,70} The functional of Perdew, Burke, and Ernzerhof (PBE)⁷¹ was used, with the dispersion correction of Steinmann and Corminboeuf (dDsC).^{72,73} The (111) surface was modeled by a $p(4\times 4)$ unit cell with 6 metallic layers, 2 of which held fixed to simulate bulk properties. A vacuum layer of 15 Å was used. The plane waves basis set was chosen to have a cut-off energy of 400 eV. Brillouin zone integration was performed by a $3\times 3\times 1$ Monkhorst-Pack⁷⁴ k-points grid and a Methfessel-Paxton smearing⁷⁵ of 0.2 eV. The wavefunction and geometric gradient were converged to 10^{-6} eV and 5×10^{-2} eV/Å, respectively. As for our computational tools, van der Waals radii were taken from the work of Bondi,⁴⁶ while covalent radii came from Pyykkö and Atsumi.⁴⁷

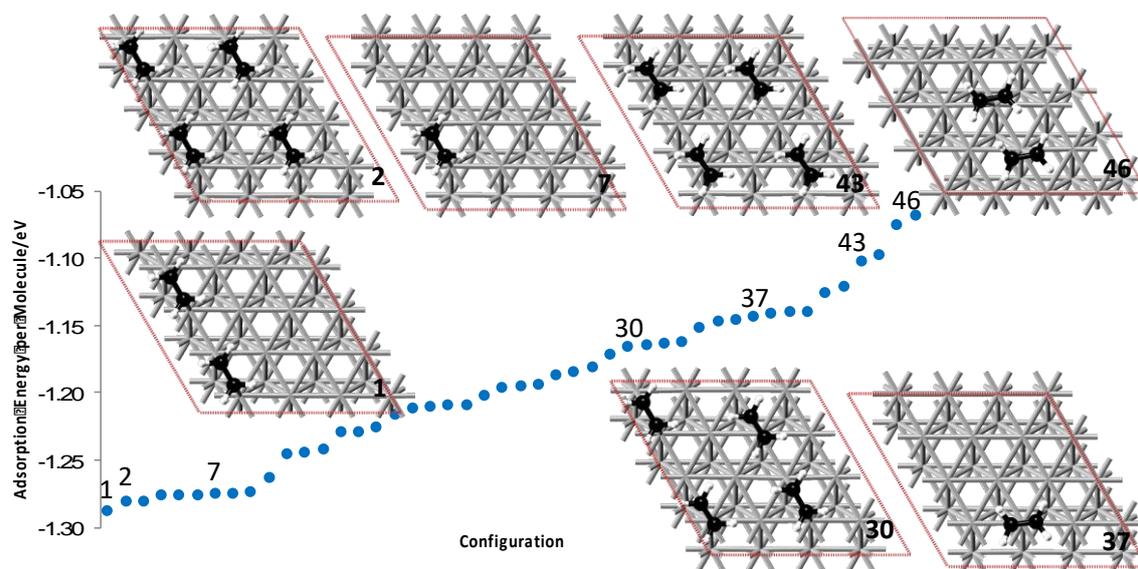


Figure 5. Adsorption energies per molecule in eV for all configurations considered. Geometries are given for the low (7 and 37) and high (2,30,43) coverage limits and the most and least stable binding structure (1,46). The unit cell is given by red dotted lines, Pd (in grey) atoms of the first two layers are shown. Carbon is depicted in black and hydrogen in white. The structural representation has been produced by the MAPS software⁷⁶

Figure 5 presents the adsorption energy per ethylene molecule for all investigated configurations ordered by increasing energy (configurations 1 and 32 are most and least stable, respectively) per ethylene molecule. The structures of the high (1/4 ML) and low (1/16 ML) coverage configurations are depicted as well. As expected, the di- σ mode (configuration 7; $E_{\text{ads}}=-1.28$ eV) is significantly more stable than the π mode (configuration 29; $E_{\text{ads}}=-1.14$ eV). In general, the through-surface and through-space lateral interactions between co-adsorbed molecules, should have a destabilizing effect either due to Pauli repulsion or by limiting electron donation to or from the surface. We therefore expect that the di- σ mode at low coverage (1/16 ML) exhibits the lowest adsorption energy per molecule. Configuration 7 in Figure 5 corresponding to that mode is, however, surpassed by six other modes lower in energy. Considering configurations 1 and 2 that correspond to the high coverage (1/4 ML) di- σ mode, we recognized slightly attractive lateral interactions as being responsible (in the order of 0.01 eV per molecule). The small stabilization originates from a subtle balance between repulsive and attractive London dispersion interactions. Indeed, we have checked that the results at the PBE level, i.e., without the explicit inclusion of dispersion interactions, are consistent with the expectation of purely repulsive lateral interactions. In other words, since the adsorbates in the high coverage regime not only experience repulsion, but are also stabilized by dispersion interactions, predicting the stability of high coverage structures is sensitive to the use of a dispersion correction.

V.2 PATTERN RECOGNITION

Pattern recognition is first carried out for the two low coverage (1/16 ML) adsorption modes (di- σ and π in Figure 6). Clusters are shown as a function of the size of the atoms, expressed as the barycenter function of R_{cov} and R_{vdw} ($x \cdot R_{\text{vdw}} + (1-x) \cdot R_{\text{cov}}$), and the different sites distinguished by the graphical lattice. As mentioned in III, considering H adsorption as a limiting case, we can derive a minimal fraction of the vdW radius to be included, so that H is large enough to occupy the smallest site, as well as a maximal fraction, assuming that a H atom at a given site should not be recognized as occupying a neighboring site as well. Therefore, we show the results for the x_{min} and the x_{max} in Figure 6. The first line is a lattice only accounting for top sites and is denoted by T. Adding the 3-fold hollow sites (H) or two-fold bridge sites (B) represents a more detailed lattice (TH and TB lattice, respectively). Finally, combining all three sites together yields the TBH lattice in the bottom line. In these pictures, the position of the C and H atomic nuclei are given by black dots, the size of the atom by a black circle with the corresponding atomic radius and the occupied sites (overlap > 50%) are given by colored circles (T, B and H are red, blue and green, respectively). For the coarsest T lattice, the two adsorption modes (di- σ and π) are readily distinguished as occupying two and one site, respectively, see Figure 6, top row. This accounts well for the expected bi- and mono-dentate adsorption mode and the dependence thereof on the chosen atomic radii is negligible.

This significantly differs from the more detailed TH lattice: At x_{\min} (30% R_{vdw}) a “top-top” mode is obtained for the di- σ configuration, which intuitively corresponds to an ethylene π bond being transformed into two C-Pd bonds. Accordingly, for the π configuration a top mode is obtained. At x_{\max} (70% R_{vdw}), however, the strongly overlapping sites, characteristic for the bond and represented by the top sites previously recognized, are now surrounded by a crown of (hollow) sites. Recognizing

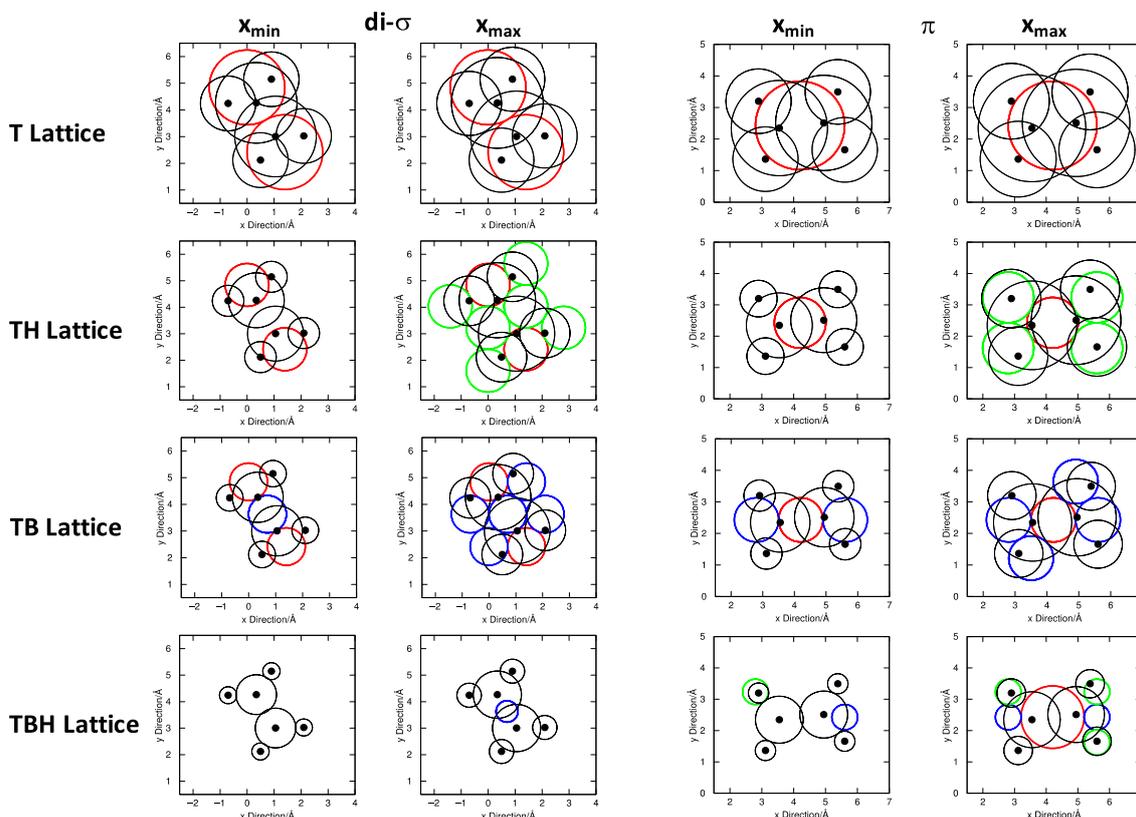


Figure 6. Recognition of the di- σ (left, configuration 7) and π (right, configuration 37) as a function of the atomic radii ($x \cdot R_{\text{vdw}} + (1-x) \cdot R_{\text{cov}}$) and the four different lattices. C and H atomic nuclei and radii are indicated as black dots and circles, respectively. Recognized top, bridge and hollow sites are given by red, blue and green circles, respectively. x_{\min} and x_{\max} are 0.8 and 1.0 for the T lattice, 0.3 and 0.7 for the TH lattice, 0.2 and 0.5 for the TB lattice and 0 and 0.15 for the TBH lattice, respectively.

that these sites are occupied by ethylene is important, in the sense that it excludes other molecules to adsorb on these sites. In the case of x_{\max} , this excludes, for instance, that a second di- σ mode adsorbs in the immediate neighborhood without distortion (the top site in the empty corner delimited by the green hollow sites cannot be occupied without an overlap with the already occupied H sites). The TB lattice, which in general is less relevant than the TH lattice, as several small adsorbates such as H or CO strongly interact with the hollow sites, overall provides a similar description, with patterns that account well for the elongated shape of the molecule. In contrast to the TH lattice, at x_{\min} (20% R_{vdw}), the bridge sites already play a role and are recognized as being occupied by the molecule, which makes, of course, perfect sense, since they are in the middle of the C-C and between the H atoms of the CH_2 group for the di- σ and π configuration, respectively. The TBH lattice turns out to be least useful for our purpose, mostly because of the large imbalance of the size of different sites: since the

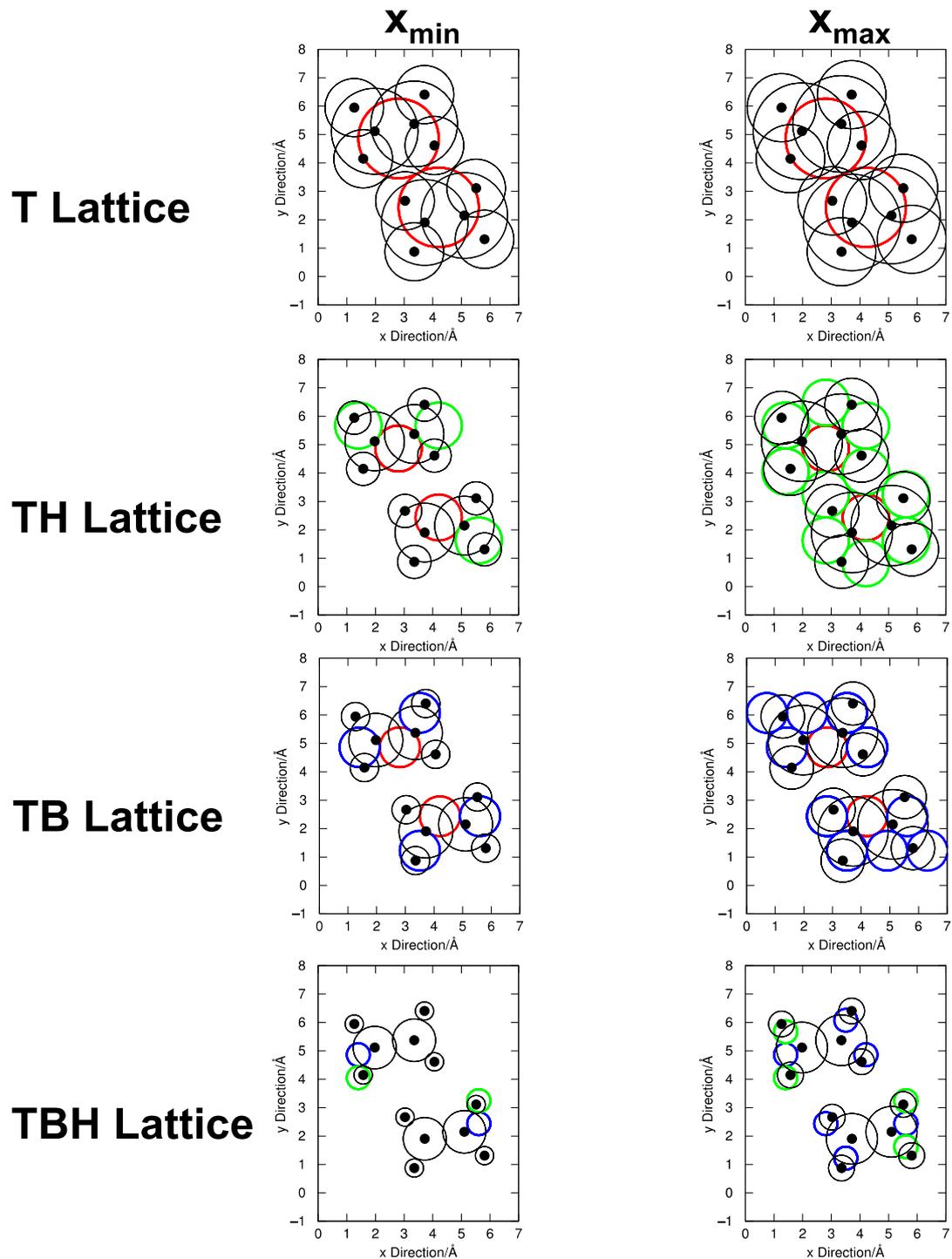


Figure 7. Illustration of pattern recognition mode (configuration 46) as a function of the atomic radii ($x \cdot R_{vdw} + (1-x) \cdot R_{cov}$) and the four different lattices. Atomic nuclei and radii are indicated as black dots and circles, respectively. Recognized top, bridge and hollow sites are given by red, blue and green circles, respectively. x_{min} and x_{max} are 0.8 and 1.0 for the T lattice, 0.3 and 0.7 for the TH lattice, 0.2 and 0.5 for the TB lattice and 0 and 0.15 for the TBH lattice, respectively.

smallest sites have a radius of only 0.4 \AA , x_{min} and x_{max} are 0 and 0.15, respectively, which does not allow for a reliable recognition of the different adsorption modes. Therefore, we will not discuss results for this lattice in later sections. In conclusion, the chemically most intuitive patterns are obtained on the T lattice, independent on the atomic radii. On the TH lattice, only with the larger radii

(70% R_{vdw}) reasonable patterns are obtained that reflect well the space occupied by ethylene and the role hollow sites could play in this case. However, in this case adsorption on sites close by is not possible without distortion, a topic on which we will come back in the next section. On the TB lattice, already at x_{min} the picture is “reasonable”, although just like on the top lattice the different adsorbates can in principle approach each other very closely. Finally, the sites in the TBH lattice are so small, that the recognition is very sensitive to the atomic radii and no particularly suitable patterns are obtained.

Having discussed the low coverage patterns, we now address the following question: Do lateral interactions lead to detectable distortions? To do so, Figure 7 addresses the extreme case of configuration 46, which is the least bound configuration at the per ethylene molecule basis and is therefore subject to the most significant lateral interactions. While the T lattice remains insensitive to the atomic radii and simply identifies two π modes, the TH and TB lattices clearly illustrate the distortion of the π mode due to lateral interactions (compared to the two left columns of Figure 6): instead of simply a top site or a linear bridge-top-bridge pattern obtained at low coverage for x_{min} , the lateral interactions induce a distortion which leads to bent, dissymmetric structures. In fact, the distortions due to the lateral interactions are actually seen both by the loss of some sites and the gain of others. In other words, distortion arises only when molecules are arranged in such a way that their van der Waals contours overlap; consequently their graphical representation is reshaped. This implies that the distorted single-body patterns should not occur “alone”, but only in the vicinity of other patterns. Hence, we can expect that there are some two-molecule patterns that should be treated as a single entity, as opposed to two independent molecules, simply because the distorted patterns cannot exist “alone”. This is also the only way of dealing with situations where neighbouring molecules start to actually overlap, since this is unacceptable for the graph theoretical representation unless they are merged into a single entity.

Overall, we find that rather small atomic radii are most promising, although they can lead to adsorptions on the graphical lattice which are “too close”, i.e., leading to excessive lateral interactions. These situations can, however, be taken care of separately by introducing the corresponding prohibitive penalty terms.

V.3 CLUSTER EXPANSION OF THE INTERACTION ENERGY

We have parameterized a variety of cluster expansions (CE) as a function of the lattice type and of the atomic radii and nearest neighbour interactions. For each value of the atomic radii, and each 46 configurations, we first detect the patterns for the molecule/surface interaction as explained above in section III and with results discussed in section V.2. 1-Body patterns are associated with a single molecule (or a “super molecule”, composed of molecules which would overlap otherwise) while patterns corresponding to the interaction of a given pair of molecules are called 2-body patterns. The energy contributions for these 1-body or 2-body patterns are then calculated from the energy of the configurations by applying the inversion procedure of section IV.

Table 1. Summary of cluster expansions using all 46 configurations as a function of the graphical lattice, the size of the atoms and the inclusion of interactions with neighbors (2 body patterns: NN: nearest neighbors). The number of unique patterns and their size is reported, and the minimum and maximum of single and two body pattern energies, together with the RMS and maximum error and the maximum Cook’s distance. The value in square brackets indicates the number of linear dependencies, which is roughly equal to the number of patterns that should treat two molecules in one 1B pattern. A star () is used to indicate that linear dependencies between 1B and 2B patterns prohibit the definition of 2B interaction energies.*

Lattice	%Van der Waals	Neighbours included	# of Patterns	# of Sites min,max	Min,max E(1B)	Min,max E(2B)	Error (RMS, MAX) eV	Max Cook’s D
T	80	none	2	1,2	-1.25 -1.11	N.A.	0.07, 0.22	0.22

T	100	none	2	1,2	-1.24 -1.12	N.A.	0.10, 0.33	0.26
TH	30	none	7	1,4	-1.28 -1.05	N.A.	0.03, 0.12	> 1000
TH	70	none	11 [1]	5,8	-1.28 -0.99	N.A.	0.02, 0.05	> 1000
TB	20	none	8	2,4	-1.27 -1.06	N.A.	0.05, 0.19	> 1000
TB	50	none	17 [1]	5,7	-1.28 -0.97	N.A.	0.02, 0.10	> 1000
T	80	NN	14	1,4	-1.28 -1.12	-0.01 0.24	0.03, 0.13	> 1000
TH	30	NN	27 [4]	1,7	-1.28 -1.05	-0.03 0.16	0.01, 0.05	> 1000
TB	20	NN	13 [3]	2,8	-1.27 -0.91	N.A.*	0.04, 0.19	> 1000

Table 1 provides a summary of the parameterizations, which describe all 46 configurations. Depending on the case a rank deficient system of equations is obtained. This can happen when, for instance, two 1-body patterns contained in a 2-body pattern are heavily distorted and, therefore, do not occur in any other configuration. The same can, actually, already occur at the 1-body expansion level. Mathematically, our inversion procedure using SVD still allows solving the system of equations, even though it is effectively rank deficient. However, it is impossible to separately calculate the energy contributions of the individual, linearly dependent patterns, since they will always appear together in any configuration. From a chemical physics point of view, this means that the corresponding arrangement should be treated as a single entity of two molecules: its energy is well defined. Furthermore, in this case the corresponding distorted single-molecule patterns are not supposed to occur alone, but only in the neighbourhood of another (distorted) molecule. Hence, the procedure where “primitive patterns” that are linearly dependent on each other are summed together in larger clusters (single body patterns of two molecules) makes both mathematical and chemical sense. The number of such patterns is indicated in square brackets in Table 1.

We measure the quality of a given cluster expansion mainly by the root mean square error (RMSE) and the maximum error. In addition, Figure 8 shows selected parity plots. As mentioned above, we also report Cook’s distances, which indicate the presence (or not) of influential configurations. As it can be seen, there are many cluster expansions for which some configurations have Cook’s distances above 1, which is indicative of highly influential points. Indeed, patterns that are defined by only one configuration have automatically an infinite Cook’s distance, which simply reflects the fact that a given cluster is not present more than once in the training set. Since, however, all 46 configurations correspond to local minima on the DFT surface, these cluster energies are still physically meaningful. Nevertheless, we have also established a second family of cluster expansions (see Table 2), where we have adopted the common practice⁵⁸ to exclude configurations that lead to Cook’s distances greater than 1.0, as these points clearly are highly influential for the least squares fit. Of course, it would be preferable to increase the training set further, in order to probe similar configurations. But as the optimal construction of a training and validation set is out of the scope of the present work and is likely to depend on the target (e.g., low vs. high coverage), we do not follow this computationally much more expensive strategy. Exclusion of configurations necessarily means that the training set is less diverse and that some of the configurations can no longer be described. The parity plots (Figure 8), however, demonstrate that we still span a similar range of adsorption energies per molecule, by

excluding between 1 and 20 configurations. In Figure 8 we also see that the R^2 values are barely affected by this procedure, suggesting that our training set is reasonably diverse.

Table 2. Summary of cluster expansions using a subset of configurations as a function of the graphical lattice, the size of the atoms and the inclusion of interactions with neighbors (2 body patterns: NN: nearest neighbors). The number of unique patterns and their size is reported, and the minimum and maximum of single and two body pattern energies as determined by the least squares fit, together with the RMS and maximum error. Configurations leading to Cook's distances larger than 1 have been excluded and the resulting number of configurations used in the fit is given in the last column

Lattice	%Van der Waals	Neighbours included	# of Patterns	# of Sites min,max	Min,max E(1B)	Min,max E(2B)	Error (RMS, MAX) eV	# Configurations
TH	30	none	6	1,3	-1.28 -1.06	N.A.	0.03, 0.12	45
TB	20	none	4	2,3	-1.27 -1.06	N.A.	0.03, 0.07	38
T	80	NN	10	1,4	-1.28 -1.12	-0.01 0.24	0.03, 0.13	41
TH	30	NN	9 [1]	1,7	-1.28 -1.05	-0.03 0.16	0.01, 0.02	29
TB	20	NN	5	2,7	-1.27 -1.14	0.01 0.08	0.02, 0.05	27

Starting with the simplest approximation, the single body expansions on a T lattice, where only di- σ and π modes are recognized we find large maximum errors (>0.2 eV), given the range of interaction energies. Without much surprise, the simple picture with only two modes and no lateral interactions is insufficient. Reassuringly, however, the di- σ and π mode have an energy of -1.25 and -1.11 eV, respectively, which compares very well to -1.28 and -1.14 eV for the low coverage (1/16 ML) configurations. This observation actually holds for all CEs presented, i.e., the minimal (π) and maximal (di- σ) energy coefficients obtained for the single body patterns (6th column of Table 1) are consistent with the corresponding low-coverage adsorption energies. Increasing the atomic radii to 100% R_{vdw} can lead to “miss-assignments” on the T lattice (e.g., a π becomes a di- σ), which explains the larger RMSE. Adding the 2-fold bridge or 3-fold hollow sites in the pattern recognition procedure increases the number of unique 1-body patterns and allows to improve the quality of the cluster expansion: these topological features allow for a detailed description of single modes even at small intermolecular distances, accounting for the deformation energy induced by the lateral interactions (without explicitly considering these interaction in the expansion). Indeed, the TH and TB lattices are characterized by 7 and 8 single body patterns at x_{min} . This more detailed description slightly reduces the RMSE from 0.07 to 0.03 - 0.05 eV and the maximum error lies now around 0.1 eV. However, without much surprise, some of the heavily distorted modes only occur in one (or few) configurations. Hence, Cook's distance reaches very high levels. Folding the lateral interactions into 1-body energy terms for slightly deformed adsorption modes is, apparently, a successful way to improve the fit with 1-body patterns exclusively. It is, however, unclear whether this would lead to a good energy prediction of configurations outside the training set; this question will be addressed in future work.

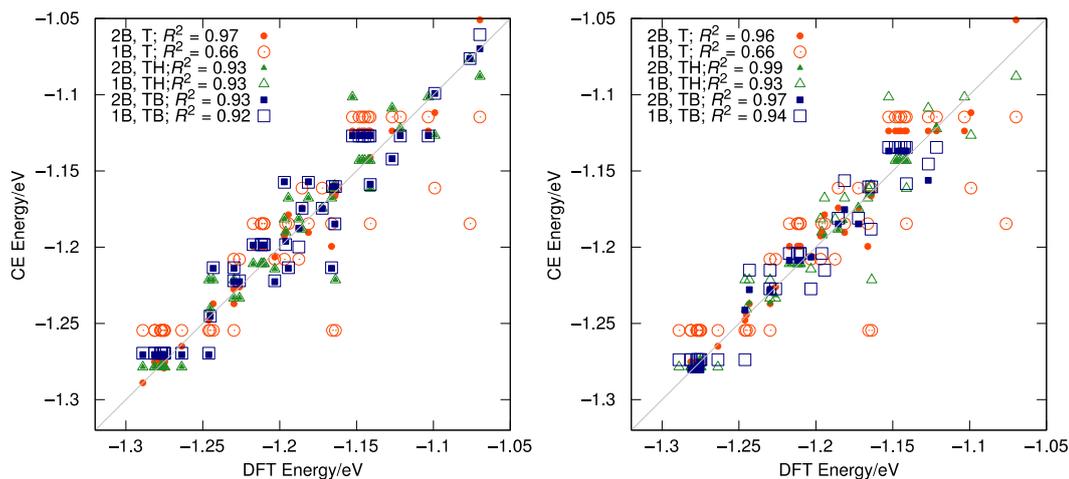


Figure 8. Parity plots of the interaction energy per molecule given by the cluster expansion (CE) compared to DFT values for four typical models of Table 1 (left), which describe all 46 configurations and of Table 2 (right) which account only for a subset of configurations. Single body (1B) expansions and 2-body (2B) expansions that take nearest neighbors (NN) into account are distinguished. The percentage of vdW radii and the R^2 is given before and after the semi-colon, respectively. The bisector is indicated by a thin grey line.

Including nearest neighbour interactions (2-body patterns) reduces the RMS and maximum errors from 0.07 and 0.2 eV to 0.03 and 0.1 eV, respectively, for the coarse T lattice. These two body patterns are repulsive by up to 0.2 eV and visibly increase the accuracy of the cluster expansion significantly, which is also seen in the parity plot of Figure 8 (left hand side). On the other hand, the Cook's distances identify highly influential configurations. For the more detailed TH and TB lattice, on the other hand, the inclusion of nearest neighbour interactions is of minor importance, although small improvements can still be seen in the overall performance. This slightly increased accuracy comes, however, at the cost of linear dependencies, which means that combinations of two single body patterns into one single body pattern describing two ethylene molecules would be not only more efficient, but also necessary from a mathematical point of view. Excluding highly influential configurations changes the overall picture only moderately (see Table 2 and Figure 8, right): for the 1-body expansions up to 8 out of the 46 configurations need to be excluded, which does not significantly affect the maximum and minimum cluster energies. The situation for the 2-body expansions is somewhat different, since, except for the T lattice, more configurations have to be excluded in order to have a well-balanced training set. Nevertheless, even in these cases we obtain CEs which span a reasonable range of configurations and adsorption energies.

In summary, we find that a 2-body expansion (with about 10 clusters as displayed in Figure 9) on the T lattice is probably the most convenient parametrization for ethylene adsorption on Pd(111). However, in hydrogenation reactions, for instance, hollow sites are necessarily included in order to describe the adsorbed hydrogen atoms. Hence, expansions on the TH lattice also need to be considered. In this case, we find that in the case of ethylene on Pd(111) lateral interactions tend to be small enough to be incorporated in slightly deformed single body patterns or can be accounted for explicitly. The ultimate comparison of the two possibilities in order to decide which approach is more effective in the determination of the ground-state geometry, would require the application of the CEs to larger unit cells and validation of the obtained results against DFT.

VI. CONCLUSIONS

This paper has proposed a set of machine-learning tools to produce unbiased cluster expansions of atomic and molecular adsorbed layers on a graphical lattice. A measure of site recognition was proposed by associating circles of maximum, non-overlapping radii to lattice points and projecting the molecular surface based on atomic radii onto the lattice plane. A molecule is assigned to a graphical site from an overlap of at least 50% on between the atoms of a molecule and a site. Mapping the recognized patterns into the configurational energy of the system finally retrieves the expanded Hamiltonian, to be calculated by pseudo-inversion.

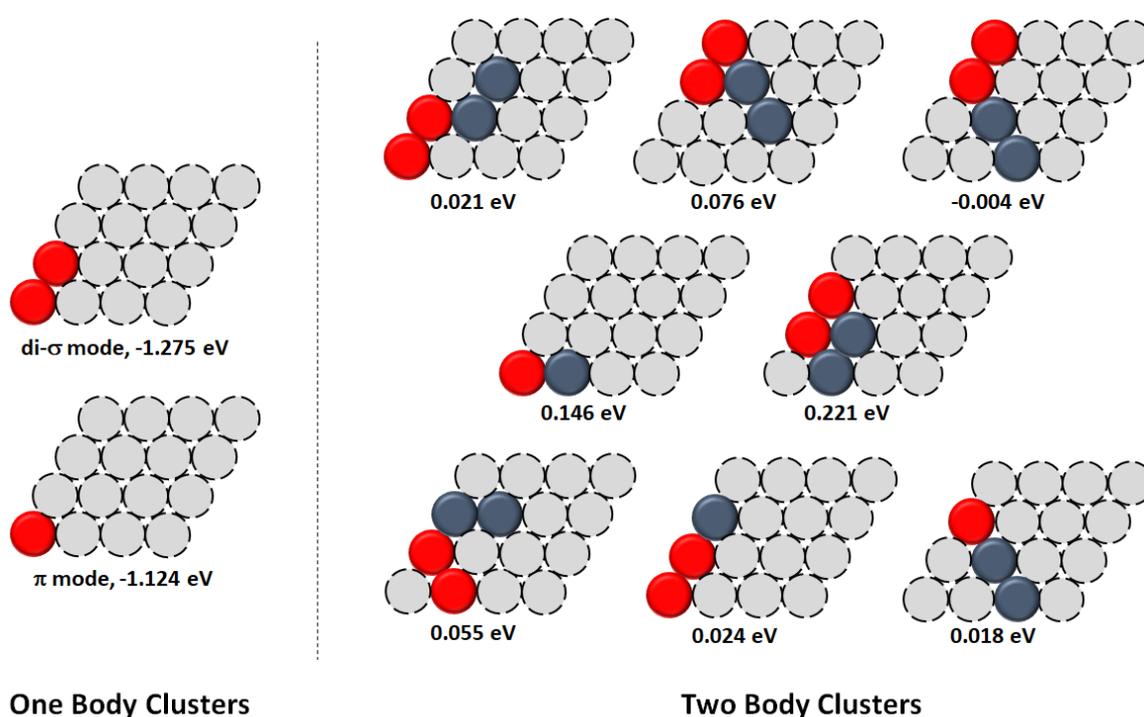


Figure 9. The cluster expansion of ethylene on Pd (111) as recognized with a criterion of 80% van der Waals radii

The proposed scheme has been applied to the adsorption of ethylene on Pd(111). The cluster expansions reproduced the known adsorption features in an accurate way, with both di- σ and π modes being clearly identified. Lattices, which are more detailed than just atop sites, were found capable of resolving important parts of interactions at short contacts, since the repulsion induces measurable distortions at the DFT level. 2-body patterns, on the other hand, are also an efficient approach to increase the accuracy of the cluster expansions. This contribution has not addressed the optimal construction of a training set and validation procedure, although we consider that these tasks are highly important and human-time intensive. However, for a given set of DFT configurations we have proposed a clear protocol to establish cluster expansions based on a set of configurations and their energy without any user input and without ambiguity. The later point is the more relevant since adsorption modes of somewhat flexible molecules such as ethylene are not trivially deduced from DFT computations at short contacts, which are typical for high-coverage situations. We suggest that around 30 configurations are sufficient to have a statistically relevant sampling of the different adsorption modes and nearest neighbour interactions. Having this tool in hand allows one to construct cluster expansions for arbitrary systems in a black-box manner based on the corresponding DFT configurations. Hence, a lattice based kinetic Monte Carlo model for selective hydrogenation on Pd(111) of acetylene including one intermediate could, by now, be constructed with about 300 DFT

computations and minimal time for the determination of the cluster expansions. It also opens up the possibility to construct cluster expansions for more complex surfaces such as alloys, where the identification of the different adsorption modes is even more challenging.

VII. SUPPLEMENTARY MATERIAL

All *ab-initio* computed energies and geometries are provided in the supplementary material.

REFERENCES

- ¹ J.M. Thomas and J.W. Thomas, *Principles and Practice of Heterogeneous Catalysis*, 1st ed. (Wiley-VCH, Weinheim, 1996).
- ² R.A. van Santen and P. Sautet, editors, *Computational Methods in Catalysis and Materials Science* (Wiley-VCH, Weinheim, 2009).
- ³ J.K. Nørskov, T. Bligaard, J. Rossmeisl, and C.H. Christensen, *Nat. Chem.* **1**, 37 (2009).
- ⁴ W. Guo, M. Stamatakis, and D.G. Vlachos, *ACS Catal.* **3**, 2248 (2013).
- ⁵ K.J. Andersson, F. Calle-Vallejo, J. Rossmeisl, and I. Chorkendorff, *J. Am. Chem. Soc.* **131**, 2404 (2009).
- ⁶ F. Calle-Vallejo, J. Jakub Tymoczko, V. Colic, Q. Huy Vu, M.D. Pohl, K. Morgenstern, D. Loffreda, P. Sautet, W. Schuhmann, and A.S. Bandarenka, *Science* **350**, 185 (2015).
- ⁷ O. Sinanoğlu and K.S. Pitzer, *J. Chem. Phys.* **32**, 1279 (1960).
- ⁸ R.G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
- ⁹ M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, and J.D. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- ¹⁰ M.A. van Hove and S.Y. Tong, *Surface Crystallography by LEED* (Springer-Verlag, Berlin-Heidelberg, 1979).
- ¹¹ A.P. Van Bavel, D. Curulla Ferré, and J.W. Niemantsverdriet, *Chem. Phys. Lett.* **407**, 227 (2005).
- ¹² K. Reuter, in *Modeling and Simulation of Heterogeneous Catalysis Reactions: From Molecular Process to the Technical System*, edited by O. Deutschmann (Wiley-VCH, Weinheim, 2011), pp. 370–392.
- ¹³ A.P.J. Jansen, *An Introduction to Kinetic Monte Carlo Simulations of Surface Reactions* (Springer-Verlag, Berlin, 2012).
- ¹⁴ M. Stamatakis and D.G. Vlachos, *ACS Catal.* **2**, 2648 (2012).
- ¹⁵ R.J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).
- ¹⁶ G. Mussardo, *Statistical Field Theory. An Introduction to Exactly Solved Models in Statistical Physics* (Oxford University Press, New York, 2010).
- ¹⁷ D. de Fontaine, in *Solid State Physics Vol. 47*, edited by H. Ehrenreich and D. Turnbull (Academic Press, New York, 1994), pp. 33–176.
- ¹⁸ G. Inden, in *Materials Science and Technology* (Wiley-VCH, Weinheim, 2013), pp. 519–581.
- ¹⁹ T.D. Lee and C.N. Yang, *Phys. Rev.* **87**, 410 (1952).
- ²⁰ S.M. Bhattacharjee and A. Khare, *Curr. Sci.* **69**, 816 (1995).
- ²¹ A. van de Walle and G. Ceder, *J. Phase Equilibria* **23**, 348 (2002).

- ²² A. van de Walle, M. Asta, and G. Ceder, *CALPHAD* **26**, 539 (2002).
- ²³ J.W.D. Connolly and A. R. Williams, *Phys. Rev. B* **27**, 5169 (1983).
- ²⁴ M. Stamatakis, Y. Chen, and D.G. Vlachos, *J. Phys. Chem. C* **115**, 24750 (2011).
- ²⁵ M. Stamatakis, M.A. Christiansen, D.G. Vlachos, and G. Mpourmpakis, *Nano Lett.* **12**, 3621 (2012).
- ²⁶ M.D. Marcinkowski, A.D. Jewell, M. Stamatakis, M.B. Boucher, E.A. Lewis, C.J. Murphy, G. Kyriakou, E. Charles, and H. Sykes, *Nat. Mater.* **12**, 523 (2013).
- ²⁷ S. Piccinin and M. Stamatakis, *ACS Catal.* **4**, 2143 (2014).
- ²⁸ M. Stamatakis and S. Piccinin, *ACS Catal.* **6**, 2105 (2016).
- ²⁹ K. Reuter and M. Scheffler, *Phys. Rev. B* **73**, 45433 (2006).
- ³⁰ D.J. Liu and J.W. Evans, *Prog. Surf. Sci.* **88**, 393 (2013).
- ³¹ A. Bajpai, K. Frey, and W.F. Schneider, *J. Phys. Chem. C* **121**, 7344–7354 (2017).
- ³² S.D. Miller and J.R. Kitchin, *Mol. Simul.* **35**, 920 (2009).
- ³³ K. Frey, D.J. Schmidt, C. Wolverton, and W.F. Schneider, *Catal. Sci. Technol.* **4**, 4356 (2014).
- ³⁴ D.J. Schmidt, W. Chen, C. Wolverton, and W.F. Schneider, *J. Chem. Theory Comput.* **8**, 264 (2012).
- ³⁵ C. Wu, D.J. Schmidt, C. Wolverton, and W.F. Schneider, *J. Catal.* **286**, 88 (2012).
- ³⁶ M. Borg, C. Stampfl, A. Mikkelsen, J. Gustafson, E. Lundgren, M. Scheffler, and J.N. Andersen, *ChemPhysChem* **6**, 1923 (2005).
- ³⁷ C. Stampfl, *Phase Transitions* **80**, 311 (2007).
- ³⁸ M. Stamatakis and D.G. Vlachos, *J. Chem. Phys.* **134**, 214115 (2011).
- ³⁹ J. Nielsen, M. D’Avezac, J. Hetherington, and M. Stamatakis, *J. Chem. Phys.* **139**, 224706 (2013).
- ⁴⁰ C.M. Bishop, *Pattern Recognition and Machine Learning* (Springer-Verlag, New York, 2006).
- ⁴¹ N.L. Biggs, *Interaction Models* (Cambridge University Press, New York, 1976).
- ⁴² A.J. Stone, *The Theory of Intermolecular Forces*, 2nd ed. (Oxford University Press, Oxford (United Kingdom), 2013).
- ⁴³ T.L. Hill, *An Introduction to Statistical Thermodynamics* (Dover Publications, New York, 1960).
- ⁴⁴ V.P. Zhdanov, *J. Chem. Phys.* **110**, 8748 (1999).
- ⁴⁵ J.F. Gonthier, S.N. Steinmann, M.D. Wodrich, and C. Corminboeuf, *Chem. Soc. Rev.* **41**, 4671 (2012).
- ⁴⁶ A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).
- ⁴⁷ P. Pyykkö and M. Atsumi, *Chem. - A Eur. J.* **15**, 186 (2009).
- ⁴⁸ R. Penrose, *Math. Proc. Cambridge Philos. Soc.* **51**, 406 (1955).
- ⁴⁹ J.C.A. Barata and M.S. Hussein, *Brazilian J. Phys.* **42**, 146 (2012).
- ⁵⁰ G. Ceder and G.D. Garbulsky, *Phys. Rev. B* **51**, 57 (1995).
- ⁵¹ N.A. Zarkevich and D.D. Johnson, *Phys. Rev. Lett.* **92**, 255702 (2004).

- ⁵² A. Seko and I. Tanaka, *Phys. Rev. B* **83**, 224111 (2011).
- ⁵³ A. Seko, K. Shitara, and I. Tanaka, *Phys. Rev. B* **90**, 174104 (2014).
- ⁵⁴ G. Golub and W. Kahan, *J. Soc. Ind. Appl. Math. Ser. B, Numer. Anal.* **2**, 205 (1965).
- ⁵⁵ R.D. Cook, *Technometrics* **19**, 15 (1977).
- ⁵⁶ R.D. Cook, *J. Am. Stat. Assoc.* **74365**, 169 (1979).
- ⁵⁷ K.A. Bollen and R.W. Jackman, in *Mod. Methods Data Anal.* (Sage, Newbury Park, 1990), pp. 257–291.
- ⁵⁸ S. Chatterjee and A.S. Hadi, *Regression Analysis by Example*, 4th ed. (Wiley Interscience, Hoboken, 2006).
- ⁵⁹ E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (SIAM, Philadelphia (PA), 1999).
- ⁶⁰ S.J. Eun, Y.J. Mun, and D.B. Min, *Compr. Rev. Food Sci. Food Saf.* **1**, 22 (2005).
- ⁶¹ D. Stacchiola, S. Azad, L. Burkholder, and W.T. Tysoe, *J. Phys. Chem. B* **105**, 11233 (2001).
- ⁶² M. Neurock and R.A. Van Santen, *J. Phys. Chem. B* **104**, 11127 (2000).
- ⁶³ E.W. Hansen and M. Neurock, *J. Catal.* **196**, 241 (2000).
- ⁶⁴ Y.-T. Wong and R. Hoffmann, *J. Phys. Chem.* **95**, 859 (1991).
- ⁶⁵ P. Sautet and J.F. Paul, *Catal. Letters* **9**, 245 (1991).
- ⁶⁶ D.M.P. Mingos, *J. Organomet. Chem.* **635**, 1 (2001).
- ⁶⁷ P.E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
- ⁶⁸ G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
- ⁶⁹ G. Kresse and J. Hafner, *Phys. Rev. B* **47**, 558 (1993).
- ⁷⁰ G. Kresse and J. Furthmüller, *Phys. Rev. B* **56**, 11169 (1996).
- ⁷¹ J.P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- ⁷² S.N. Steinmann and C. Corminboeuf, *J. Chem. Theory Comput.* **7**, 3567 (2011).
- ⁷³ S. Gautier, S. Steinmann, C. Michel, P. Fleurat-Lessard, and P. Sautet, *Phys. Chem. Chem. Phys.* **17**, 28921 (2015).
- ⁷⁴ J.D. Pack and H.J. Monkhorst, *Phys. Rev. B* **13**, 5188 (1977).
- ⁷⁵ M. Methfessel and A. T. Paxton, *Phys. Rev. B* **40**, 3616 (1989).
- ⁷⁶ F. Tao and M. Salmeron, *Science* **331**, 171 (2011).

Conclusions and Outlook

Selective hydrogenation of acetylene in ethylene-rich flows is a fundamental process in the petrochemical industry; indeed, it is the elected method to purify polymer-grade ethylene sold by petroleum companies. The reaction is catalyzed by Pd, which enhances ethylene formation over the total hydrogenation product, ethane, at industrially acceptable yields. In any event, pure Pd catalysts show emergence of additional byproducts, mostly of oligomeric nature; those side reactions leading to such byproducts are undesired, because they inactivate the catalyst in the long run. Formation of the oligomeric mixture, known as "green oil" in the literature, is among the most problematic aspects from the industrial point of view, and still to be addressed today.

It is known in refining technology that alloying Pd with noble metals leads to catalysts of greater lifespan, i.e. producing less green oil and after longer periods of time. Two main effects have been proposed as explanatory for this phenomenon. The first effect, called *electronic effect*, addresses the suppression of oligomers to a modification of the electronic structure of the catalyst; in this perspective, molecules are less prone to adsorption on Pd due to the modified band structure so that the surface looks less crowded. The other effect is the *ensemble effect*. The ensemble effect assumes that the change in the catalyst electronic structure after alloying noble metals is not significant and the adsorption process on Pd is left unaltered; what really changes is the distribution of inactive sites – that is, site not prone to adsorption such as noble metal atoms – preventing collisions of molecules on the surface. While of scarce importance to the electronic effect, it is clear that the ensemble effect is sensitive to the surface topology: if large domains of Pd are formed on the catalytic surface of the alloy, production of oligomers is expected anyway. **The main goal of the present work has been to establish the extent of the ensemble and of the electronic effect to green oil formation with the aim to provide our industrial partner with physical and chemical insight on the selectivity of the reaction.**

Pure Pd catalysts are known to undergo severe modifications when in contact with an atmosphere of acetylene and hydrogen. In particular, formation of hydride and carbide phases is observed by several experimental techniques. Theory also shows that the carbide phase prevents unselective hydrogenation both by desorbing ethylene faster than pure Pd, and by slowing the diffusion of unselective hydrogen from the subsurface. Since the role of physical modification of pure Pd catalysts was found crucial to ethylene selectivity, acetylene's ability to rearrange the surface of Pd-Ag alloys could have exposed the physical premises of oligomerization and catalyst inactivation. An approximate scheme was therefore used to screen this possibility in the quickest way possible. Exploiting the framework of the triangular ensemble description of (111) alloyed surface planes, a statistical-mechanical, mean-field model was proposed to predict the surface topology of the topmost layer in the form of phase diagrams. The model predicted lateral segregation of Pd domains in the impurity range of the catalyst (0.0-0.25 and 0.80-0.90 of Ag in the layer) at 300 K and 1 bar of acetylenic pressure; intermediate compositions were found to favor mixing of metals. Increasing the temperature over 600 K destabilizes formation of Pd islands, which disappear at 900 K; here, the layer is absolutely random. Thermodynamics allows for a nice interpretation of this prediction. Island formation, driven by the adsorption of acetylene (energetic term), has to compete with the tendency of the layer to increase its disorder (entropic term). When the Ag fraction is a small impurity dissolved in Pd (or when Pd is a small impurity in Ag), the entropic component of the free energy is not great. The energy term provided by adsorption of acetylene on Pd is thus too large to get compensated by the entropic counterpart and the layer gets organized into Pd islands. On the contrary, the closer the Ag/Pd ratio gets to 50/50, the more likely is to randomize; the entropy reaches

a maximum at such composition. The entropic term is even more sustained at higher temperatures because of the desorption process. This is to say that thermodynamics fights against the ensemble effect at relevant conditions; acetylene rearranges the shape of the catalyst in such a way as to form large domains of Pd.

The simplified model discussed above only describes a single layer in contact with heat and matter reservoirs. Moreover, it is approximated; triangular ensembles are treated as cells in chemical equilibrium onto which acetylene adsorbs in the averaged field thereby generated. To ascertain the role of order/disorder phase transitions at the surface level of Pd-Ag catalysts via a detailed cluster expansion, a Hamiltonian of a 4-layers surface model was built. This model Hamiltonian was fitted to 107 configurations and statistically validated by leave-one-out cross validation, then provided as input to a on purpose programmed Monte Carlo code. Numerical simulations were run spanning the whole compositional range and a wide set of temperatures (300 K -1300 K) at 1 bar of acetylenic pressure. Again, formation of islands of Pd was retrieved in the range 300 – 500 K, even at 0.5 Ag content. It was also possible to address the interplay between segregation and layer organization: acetylene segregates Pd vertically, pushing Ag to the subsurface layers of the model. Randomization of the surface only happens at very high Ag content or sustained temperatures (around 1300 K, a purely theoretical temperature); this randomization conserves its phase-transition character, as shown by the discontinuity of order parameters. These computational evidence again seems to show that large ensembles are formed in the long run of the chemical transformation.

Having provided evidence that large ensembles of Pd may appear on the surface at some point in the process, modeling of couplings between molecules on Pd was started. This problem still raises a big challenge for theoreticians, due to the high complexity of the reaction network. It must be noted in this sense that the main mechanism for acetylene hydrogenation on Pd is the Horiuti-Polanyi mechanism, i.e. subsequent additions of atomic hydrogen to the molecule until its total hydrogenation to ethane; these addition may proceed in a symmetric or in an dissymmetric way, leading to different intermediates. In the impossibility to a priori exclude any coupling, all of the possible C₂-C₂ couplings were calculated from their reactant state to their product states, passing through the transition state. Though high barriers for C-C couplings were found in general, a few dimerization steps were isolated which compete with the hydrogenation process. A noxious aspect of these steps is they involve hydrogenation early intermediates, such as CHCH and CHCH₂, which get subtracted from their route to ethylene. At the same time, the important hydrogen-storing capacity of oligomers removes hydrogen from the ethylene route, which favours coupling; this suggests that green oil formation is somehow autocatalytic. While the role of ethylene in oligomerization seems of no importance, there is room for benzene to form, even if in small amounts. As for the ensemble effect, sites as small as 4 atoms parallelograms are enough to promote hydrogenation early-intermediates coupling.

Up to this point, the study of the ensemble effect was only study to the level of mean-field kinetics. The mixture of molecules appearing in the hydrogenation process is indeed much complex and unexpected behaviour of the reaction due to lateral interactions cannot be ruled out. A model Hamiltonian as detailed as to capture all the interaction features would be greatly desirable in this sense. It is clear that a manual approach to generate such a Hamiltonian is to be abandoned (it would require human inspection of hundreds of calculations) and that a computational tool is necessary to scan a big load of optimized geometries. Such a tool was developed and tested in our laboratory. The program automatically detects adsorbed modes (also called patterns) by testing geometric criteria and applying a rule of inference; then, builds a pattern occurrence matrix and interpolates the model Hamiltonian solving its associated system of equations. The program successfully detected and

interpolated the simple case of ethylene adsorption on Pd (111); it was also addressed short-contact features, which improves model precision and sets a new standard of model Hamiltonian generation.

Our work tackled fundamental problems in acetylene selective hydrogenations but actually opens for more research. For sure, prediction of Pd segregation (both lateral and vertical) in the alloy challenges experimentalists for a verification. Monitoring the evolution of a Pd-Ag surface model (more than a nanoparticle) in realistic conditions would be greatly beneficial to establish the fitness of our predictions. Theoretical questions are of course not completely exhausted. In particular, the interaction model of the complex mixture of intermediates is yet to be constructed; this requires a huge load of computations and the application of the program above mentioned. Moreover, this is the last step missing before kinetic Monte Carlo simulations are performed; a database of kinetic parameters has been already made available in the present work. In any event, the tool does not take the third dimension of space into account; work in this direction would put computational scientists into the position of run kMC simulations of entire nanoparticles, an aspect not captured in this work, but certainly much relevant. Along with that, formation of Pd-Ag/C and Pd-Ag/H phases was not addressed; considering the boost of selectivity towards ethylene such phases are responsible for in pure Pd catalyst, this line of research should not be overlooked in future work.

Annex

Annex of Chapter 4

Table 1. Cluster coefficients. Two-body corrections due to adsorption are classified based on the orientation of the metallic pair axis, M-M, to the orientation of acetylene C-C axis; lateral interaction contributions are classified by the distance of the pair (2nd nearest neighbour or 3rd nearest neighbour) and the interaction of the metal atom with the two adsorbed molecules (W=weakly interacting, S=strongly interacting).

Cluster	Energy, eV	Cluster	Energy, eV
Layer 1: Pd	-1.93	Layer 2: Ag; Layer 3: Pd	-1.44
Layer 1: Ag	1.44	Layer 2: Ag; Layer 3: Ag	-1.01
Layer 2: Pd	0.32	Layer 3: Pd; Layer 4: Pd	-0.90
Layer 2: Ag	-0.82	Layer 3: Pd; Layer 4: Ag	-0.27
Layer 3: Pd	-0.39	Layer 3: Ag; Layer 4: Pd	-0.50
Layer 3: Ag	-0.10	Layer 3: Ag; Layer 4: Ag	0.19
Layer 4: Pd	-0.47	Weakly Bound Pd	-0.32
Layer 4: Ag	-0.03	Strongly Bound Pd	-0.48
Layer 1: Pd; Layer 1: Pd	0.11	Weakly Bound Ag	0.02
Layer 1: Pd; Layer 1: Ag	-0.71	Strongly Bound Ag	-0.12
Layer 1: Ag; Layer 1: Ag	-1.48	Perpendicular Pd-Pd	-0.46
Layer 2: Pd; Layer 2: Pd	-1.21	Parallel Pd-Pd	-0.23
Layer 2: Pd; Layer 2: Ag	-0.66	Perpendicular Ag-Ag	0.06
Layer 2: Ag; Layer 2: Ag	-0.20	Parallel Ag-Ag	-0.05
Layer 3: Pd; Layer 3: Pd	-0.67	Perpendicular Pd-Ag	-0.08
Layer 3: Pd; Layer 3: Ag	-0.65	Perpendicular Ag-Pd	-0.12
Layer 3: Ag; Layer 3: Ag	-0.76	Parallel Pd-Ag	-0.03
Layer 4: Pd; Layer 4: Pd	-0.66	2 nd WW Pd	0.42
Layer 4: Pd; Layer 4: Ag	-0.71	2 nd SS Pd	0.08
Layer 4: Ag; Layer 4: Ag	-0.70	2 nd WS Pd	0.19
Layer 1: Pd; Layer 2: Pd	-2.03	2 nd SS Ag	0.04
Layer 1: Pd; Layer 2: Ag	-0.57	3 rd WW Pd	0.42
Layer 1: Ag; Layer 2: Pd	-0.19	3 rd SS Pd	0.23
Layer 1: Ag; Layer 2: Ag	1.33	3 rd WS Pd	0.23
Layer 2: Pd; Layer 3: Pd	0.27	3 rd WW Ag	-0.01
Layer 2: Pd; Layer 3: Ag	0.70		

Table 2. Cross Validation. All values are in eV. The score is the square of the residual as calculated with a model not taking the configuration into account; the root mean square gives the overall cross-validation score of the fit.

Total Energy	Prediction	Score	Total Energy	Prediction	Score
-186.16	-185.93	0.32	-101.33	-101.37	0.09
-101.53	-101.07	0.63	-139.38	-139.51	0.21
-166.29	-166.49	0.30	-139.86	-140.20	0.49
-164.60	-164.35	0.40	-140.12	-140.05	0.17
-164.61	-164.75	0.20	-139.29	-139.36	0.17
-166.32	-166.47	0.25	-101.36	-101.32	0.05
-143.09	-143.38	0.42	-190.34	-190.35	0.64
-145.63	-145.31	0.45	-145.28	-145.37	0.23
-143.62	-143.22	0.58	-190.79	-190.70	0.11
-143.14	-143.56	0.61	-190.66	-190.70	0.05
-121.93	-122.03	0.15	-139.83	-139.51	0.57

-138.34	-138.34	0.01	-142.11	-142.08	0.07
-143.31	-143.37	0.17	-142.12	-142.03	0.18
-139.69	-139.66	0.05	-101.25	-101.32	0.09
-144.13	-144.13	0.13	-187.60	-187.63	0.06
-188.78	-188.35	0.53	-187.81	-187.74	0.09
-101.46	-101.21	0.30	-187.78	-187.59	0.40
-166.31	-166.64	0.49	-187.61	-187.63	0.04
-166.79	-166.77	0.03	-187.83	-187.74	0.12
-167.28	-167.17	0.15	-153.92	-153.90	0.64
-168.69	-168.89	0.31	-150.11	-150.08	0.66
-143.18	-143.53	0.49	-144.95	-144.98	0.15
-145.66	-145.46	0.29	-131.12	-131.04	0.35
-145.38	-145.64	0.38	-137.91	-137.87	0.09
-145.87	-145.99	0.17	-139.35	-139.34	0.02
-124.13	-124.46	0.47	-190.86	-190.77	0.15
-124.37	-124.27	0.13	-101.53	-101.36	0.29
-124.34	-124.07	0.44	-136.42	-136.51	0.22
-122.08	-122.40	0.44	-104.49	-104.49	0.00
-124.06	-124.12	0.09	-182.89	-182.80	0.22
-123.99	-123.92	0.13	-135.72	-135.67	0.10
-121.93	-122.25	0.47	-103.84	-103.65	0.28
-190.65	-190.59	0.11	-181.88	-181.95	0.14
-190.64	-190.70	0.07	-185.49	-185.67	0.40
-190.69	-190.70	0.00	-140.67	-140.61	0.09
-190.60	-190.36	0.58	-184.08	-184.08	0.01
-101.25	-101.32	0.09	-188.33	-188.57	0.58

Annex of Chapter 5

Table 1. C_2 Species Hydrogenations.

Reaction Step	R_{C-H} , Å	ν , cm^{-1}	ΔE_{act} , eV	ΔE_{rea} , eV
$CHCH + H \rightarrow CHCH_2$	1.62	844.26	0.91	0.19
$CHCH_2 + H \rightarrow CH_2CH_2$	1.71	805.29	0.92	-0.03
$CHCH_2 + H \rightarrow CHCH_3$	1.60	944.26	0.83	0.16
$CH_2CH_2 + H \rightarrow CH_2CH_3$	1.49	945.30	0.93	0.47
$CH_2CH_3 + H \rightarrow CH_3CH_3$	1.56	914.65	0.58	0.00
$CHCH_3 + H \rightarrow CH_2CH_3$	1.63	813.97	0.90	0.18
$CCH_3 + H \rightarrow CHCH_3$	1.30	311.31	1.05	0.87
$CCH_2 + H \rightarrow CHCH_2$	1.45	922.22	1.02	0.43
$CCH_2 + H \rightarrow CCH_3$	1.67	966.23	0.78	-0.47

Table 2. Coupling of C_2 Species to C_4 Species

$2 CHCH \rightarrow C_4H_4$	Groups	R_{C-C} , Å	ν , cm^{-1}	ΔE_{act} , eV	ΔE_{rea} , eV
Cis coupling 1	HC---CH	2.01	502.49	0.47	-0.39
Cis coupling 2	HC---CH	2.18	467.70	0.85	-0.54

Trans coupling	HC---CH	1.97	508.14	0.79	-0.21
<i>CHCH + CHCH₂ → C₄H₅</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis coupling 1	HC---CH	2.09	502.50	0.61	-0.99
Trans coupling 1	HC---CH	2.12	473.68	0.42	-0.89
Trans coupling 2	HC---CH	2.10	436.40	0.43	-0.98
<i>CHCH + CH₂CH₂ → C₄H₆</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis coupling 1	HC---CH ₂	2.06	513.83	1.13	-0.14
Cis coupling 2	HC---CH ₂	2.35	567.35	1.69	0.61
Trans Coupling	HC---CH ₂	2.20	472.64	1.37	0.18
<i>CHCH + CH₂CH₃ → C₄H₇</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis coupling	HC---CH ₂	2.88	458.65	0.62	-0.82
Trans coupling	HC---CH ₂	2.17	549.83	0.87	-0.90
<i>CHCH + CHCH₃ → C₄H₆</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis coupling	HC---CH	2.01	556.55	0.94	0.00
Trans coupling	HC---CH	2.01	556.46	0.94	-0.03
<i>CHCH + CCH₃ → C₄H₄</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	HC---C	1.91	554.50	1.17	0.27
<i>CHCH + CCH₂ → C₄H₅</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling 1	HC---C	1.95	554.50	1.15	-0.38
Coupling 2	HC---CH ₂	2.25	412.73	0.80	-0.14
<i>2 CHCH₂ → C₄H₆</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling 1	HC---CH	2.18	525.78	0.80	-1.30
Coupling 2	H ₂ C---CH ₂	2.24	449.79	1.63	0.00
<i>CHCH₂ + CH₂CH₂ → C₄H₇</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Trans Coupling 1	HC---CH ₂	2.07	441.93	1.16	-0.39
Cis Coupling	H ₂ C---CH ₂	2.37	449.79	1.63	0.00
Trans Coupling 2	H ₂ C---CH ₂	2.08	382.15	1.49	-0.03
<i>CHCH₂ + CH₂CH₃ → C₄H₈</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	H ₂ C---CH ₂	2.28	542.58	1.73	-0.42
<i>CHCH₂ + CHCH₃ → C₄H₇</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis Coupling	H ₂ C---CH	2.43	318.88	1.56	0.53
Trans Coupling	H ₂ C---CH	1.96	451.33	1.06	-0.30
<i>CHCH₂ + CCH₃ → C₄H₆</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling 1	HC---C	1.93	530.91	1.02	-0.47
Coupling 2	H ₂ C---C	3.14	403.68	1.63	0.56
<i>CHCH₂ + CCH₂ → C₄H₅</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling 1	HC---C	2.01	477.01	0.52	-0.81
Coupling 2	HC---CH ₂	2.29	352.36	0.93	-0.33
<i>2 CH₂CH₂ → C₄H₈</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis Coupling	H ₂ C---CH ₂	2.09	514.29	1.83	-0.05
Trans Coupling	H ₂ C---CH ₂	2.18	758.49	2.36	0.26
<i>CH₂CH₂ + CH₂CH₃ → C₄H₉</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Trans Coupling	H ₂ C---CH ₂	2.79	220.06	1.51	-0.07
<i>CH₂CH₂ + CHCH₃ → C₄H₈</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis Coupling 1	H ₂ C---CH	2.29	524.72	1.78	-0.08
Trans Coupling 2	H ₂ C---CH	2.00	488.31	1.35	-0.12
<i>CH₂CH₂ + CCH₃ → C₄H₇</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	C---CH ₂	1.99	471.53	1.80	0.88
<i>CH₂CH₂ + CCH₂ → C₄H₆</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling 1	H ₂ C---C	2.05	461.69	1.28	0.37
<i>2 CH₂CH₃ → C₄H₁₀</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	H ₂ C---CH ₂	2.21	659.31	2.11	-0.64
<i>CH₂CH₃ + CHCH₃ → C₄H₉</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	H ₂ C---CH	2.31	481.05	1.53	-0.39
<i>CH₂CH₃ + CCH₃ → C₄H₈</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	H ₂ C---C	2.17	579.47	1.80	0.29
<i>CH₂CH₃ + CCH₂ → C₄H₇</i>	Groups	R_{C---C}, Å	<i>v</i>, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV

Coupling 1	H ₂ C---C	2.07	408.87	1.00	-0.30
Coupling 2	H ₂ C---CH ₂	2.64	606.54	1.55	-1.04
2 CHCH₃ → C₄H₈	Groups	R_{C---C}, Å	ν, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Cis coupling 1	HC---CH	2.06	574.82	1.00	-1.04
Trans coupling 1	HC---CH	2.01	596.40	1.05	-0.76
CHCH₃ + CCH₂ → C₄H₆	Groups	R_{C---C}, Å	ν, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	HC---CH ₂	2.11	422.54	0.52	-0.63
2 CCH₃ → C₄H₆	Groups	R_{C---C}, Å	ν, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	C---C	1.96	671.51	1.67	0.22
CCH₃ + CCH₂ → C₄H₅	Groups	R_{C---C}, Å	ν, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling	C---CH ₂	1.10	425.23	1.20	0.65
2 CCH₂ → C₄H₄	Groups	R_{C---C}, Å	ν, cm⁻¹	ΔE_{act}, eV	ΔE_{rea}, eV
Coupling 1	C---C	1.89	652.29	1.68	0.21
Coupling 2	C---CH ₂	2.15	432.50	0.67	-0.12
Coupling 3	H ₂ C---CH ₂	3.08	337.78	0.53	-0.92
Coupling 4	H ₂ C---CH ₂	2.58	342.14	0.48	-0.95

Table 3. C₄ Species Hydrogenations

Reaction Step	R _{C---H} , Å	ν, cm ⁻¹	ΔE _{act} , eV	ΔE _{rea} , eV
CHCHCCH ₂ + H → CHCHCCH ₃	1.51	989.17	1.43	1
CHCHCH ₂ C + H → CH ₂ CHCCH ₂	1.71	873.15	1.03	2
CHCHCHCH + H → CHCHCHCH ₂	1.65	904.45	0.85	3
CHCHCHCH ₂ + H → CHCHCH ₂ CH ₂	1.33	614.83	1.28	4
CHCHCH ₂ CH ₂ + H → CHCHCH ₂ CH ₃	1.58	858.73	0.94	5
CHCHCHCH ₃ + H → CHCHCH ₂ CH ₃	1.40	751.03	0.31	6
CHCHCH ₂ C + H → CHCH ₂ CCH ₂	1.61	934.72	1.03	7
CCH ₂ CCH ₂ + H → CHCH ₂ CCH ₂	1.40	718.39	1.13	8
CH ₂ CCCH ₂ + H → CH ₂ CHCCH ₂	1.67	869.05	0.42	9
CHCHCCH ₂ + H → CHCH ₂ CCH ₂	1.65	888.14	1.74	10
CHCHCCH ₂ + H → CH ₂ CHCCH ₂	1.71	806.34	1.64	11
CCH ₂ CCH ₃ + H → CHCH ₂ CCH ₃	1.42	747.29	0.86	12
CH ₂ CHCCH ₂ + H → CH ₂ CHCHCH ₂	1.79	905.45	1.32	14
CH ₂ CHCCH ₂ + H → CH ₂ CHCCH ₃	1.66	803.61	1.03	15
CHCHCCH ₃ + H → CH ₂ CHCCH ₃	1.55	1054.36	1.06	17
CHCHCCH ₃ + H → CHCH ₂ CCH ₃	1.44	910.31	1.06	18
CHCHCHCH ₂ + H → CH ₂ CHCHCH ₂	1.76	761.58	1.18	19
CH ₂ CCH ₂ CH ₂ + H → CH ₂ CHCH ₂ CH ₂	1.67	805.28	0.51	21
CHCHCH ₂ CH ₂ + H → CHCH ₂ CH ₂ CH ₂	1.66	918.38	1.09	22
CH ₂ CHCHCH ₂ + H → CH ₂ CH ₂ CHCH ₂	1.55	964.94	1.22	23
CHCH ₂ CCH ₃ + H → CHCH ₂ CHCH ₃	1.67	931.84	1.19	24
CCH ₂ CHCH ₃ + H → CHCH ₂ CHCH ₃	1.33	480.51	1.01	25
CHCH ₂ CH ₂ CH + H → CHCH ₂ CH ₂ CH ₂	1.66	519.79	0.85	26
CHCHCH ₂ CH ₂ + H → CH ₂ CHCH ₂ CH ₂	1.77	753.47	1.01	27
CCH ₂ CCH ₂ + H → CCH ₂ CHCH ₂	1.79	760.45	0.87	32
CCH ₂ CH ₂ C + H → CHCH ₂ CH ₂ C	1.62	764.74	1.14	35
CH ₂ CCHCH ₂ + H → CH ₂ CCH ₂ CH ₂	1.35	660.80	1.40	36
CH ₂ CCH ₂ CH ₂ + H → CH ₃ CCH ₂ CH ₂	1.53	945.31	0.80	40
CH ₃ CCHCH ₂ + H → CH ₃ CCH ₂ CH ₂	1.57	911.29	1.23	41
CH ₃ CCH ₂ CH + H → CH ₃ CCH ₂ CH ₂	1.68	883.18	1.34	42
CH ₂ CHCH ₂ CH ₂ + H → CH ₂ CHCH ₂ CH ₃	1.42	725.68	1.05	44
CH ₂ CHCH ₂ CH ₂ + H → CH ₂ CH ₂ CH ₂ CH ₂	1.57	966.78	1.20	45
CH ₂ CHCHCH ₃ + H → CH ₂ CH ₂ CHCH ₃	1.58	911.89	0.68	46

$CH_2CHCH_2CH_2 + H \rightarrow CH_2CH_2CH_2CH_2$	1.35	571.43	1.11	47
$CHCH_2CHCH_3 + H \rightarrow CH_2CH_2CHCH_3$	1.64	791.83	0.96	48
$CH_2CH_2CH_2CH_2 + H \rightarrow CH_2CH_2CH_2CH_3$	1.58	741.29	0.91	49
$CH_2CH_2CHCH_3 + H \rightarrow CH_2CH_2CH_2CH_3$	1.50	853.93	0.87	50
$CHCH_2CH_2CH_3 + H \rightarrow CH_2CH_2CH_2CH_3$	1.65	889.82	0.97	52
$CHCH_2CH_2CH_2 + H \rightarrow CH_2CH_2CH_2CH_2$	1.54	960.69	1.08	53
$CH_2CH_2CH_2CH_3 + H \rightarrow CH_3CH_2CH_2CH_3$	1.59	983.26	0.66	54
$CH_3CHCH_2CH_3 + H \rightarrow CH_2CH_2CH_2CH_3$	1.62	719.17	1.08	55
$CH_2CCH_2CH_2 + H \rightarrow CH_2CCH_2CH_3$	1.59	1009.13	0.35	56
$CH_3CCH_2CH_2 + H \rightarrow CH_3CCH_2CH_3$	1.54	933.27	0.85	60
$CH_2CCH_2CH_3 + H \rightarrow CH_3CCH_2CH_3$	1.60	905.29	0.90	61
$CH_3CHCH_2CH_2 + H \rightarrow CH_3CHCH_2CH_3$	1.55	989.17	4.69	63
$CH_2CCH_2CH_3 + H \rightarrow CH_2CHCH_2CH_3$	1.61	811.34	-2.83	64
$CH_3CHCHCH_3 + H \rightarrow CH_3CHCH_2CH_3$	1.55	967.11	1.15	66
$CH_2CCCH_2 + H \rightarrow CH_2CCCH_3$	1.59	1012.68	0.50	67
$CH_2CCCH_2 + H \rightarrow CH_2CCCH_3$	1.62	976.04	0.50	68
$CH_2CCCH_3 + H \rightarrow CH_3CCCH_3$	1.66	869.63	0.25	69
$CH_2CCHCH_2 + H \rightarrow CH_2CCHCH_3$	1.50	949.73	0.69	70
$CCH_2CHCH_2 + H \rightarrow CCH_2CHCH_3$	1.62	399.45	0.93	71
$CH_2CCCH_3 + H \rightarrow CH_2CCHCH_3$	1.61	825.03	0.67	72
$CCH_2CHCH_2 + H \rightarrow CCH_2CHCH_3$	1.63	405.98	0.85	73
$CH_3CCHCH_2 + H \rightarrow CH_3CCHCH_3$	1.52	1024.74	1.11	74
$CH_3CCCH_3 + H \rightarrow CH_3CCHCH_3$	1.63	810.16	1.15	75
$CH_2CHCHCH_3 + H \rightarrow cis - CH_3CHCHCH_3$	1.44	668.44	0.32	77
$CH_2CCHCH + H \rightarrow CH_3CCHCH$	1.52	955.76	1.43	81
$CH_3CHCHCH + H \rightarrow CH_3CHCHCH_2$	1.74	805.56	0.47	82
$CHCHCHCH_2 + H \rightarrow CHCHCHCH_3$	1.50	663.42	1.17	83
$CH_3CCHCH + H \rightarrow CH_3CHCHCH$	1.56	709.47	0.90	84
$CH_3CHCHCH + H \rightarrow CH_3CH_2CHCH$	1.43	739.49	0.45	85