Simulations of chiral ordering of achiral molecules by Density Functional Theory and Monte Carlo

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy
by
Haiping Lin

May 1, 2008

Contents

Abstract

The self-organization of organic molecules on metal surfaces can play a crucial role in various subjects of science such as electrochemistry and heterogeneous catalysis. However, their practical applications have been suffering from lack of atomic scale understanding of the ordering behavior. Theoretically, however, the description of the self-assembly processes has been limited by the inability of microscopic models to account for the interplay of all elementary processes at realistic temperatures and pressures. In this work, we take the succinate molecules and Cu(110) surface as a model system to study the nature of the asymmetric self-assembly behavior of the achiral adsorbates on an achiral surface. In the first step, density-functional theory (DFT) is used to accurately characterize the molecule-surface system on the microscopic level. The obtained energetics is then employed to parameterize a lattice gas Hamiltonian, which subsequently allows to address the mesoscopic ordering behavior at finite temperatures by means of Monte Carlo simulations. Two well-defined ordered structures have appeared in the Monte Carlo simulations. The (5 0, 3 2) structure has been confirmed to be the ground state configuration by direct DFT calculations. This structure has not been reported before and it can be the ordering configuration observed in some STM experiments. By summing up all the calculations, the nature of the self-organization behavior of succinate molecules on Cu(110) is attributed to the indirect lateral interactions of the adsorbates. This result may provide a new understanding of designing two dimensional periodic architectures on metal surfaces.

Acknowledgements

Chapter 1

Introduction

1.1 Heterogeneous catalysis

In the early of 1800's, the remarkable ease of water production from oxygen and hydrogen on Pt surfaces leaded to the definition of the term "catalysis" [1] : a (chemical) substance that accelerates (or slows down) chemical reactions without being changed itself. Due to its immense economical and environmental importance, catalysis has over the last century been subject of numerous investigations and undergone an enormous development. Today, 90 percent of the chemical manufacturing processes in the world use catalysts in some form [2].

Catalysts can be split into two categories: homogeneous and heterogeneous catalysts. Homogeneous catalysts are in the same phase as the reaction mixture. As an illustration, enzymes are dissolved in a liquid phase, e.g. human blood, and so are the reactants, such as blood-sugars and proteins. The main advantage of this kind of catalyst is that they are well dispersed and form a uniform phase with reactants. This makes homogeneous catalysts easily accessible to reactants. But the main disadvantage is that industrial recycling of catalysts from such a homogeneous system is usually very expensive. In this case it is necessary to separate the catalysts from products, because most homogeneous catalysts consist of ligands and metal ions; losing them with the products will not only increase the cost of production, but also pose health hazard to customers and pollute the environment.

By contrast, heterogeneous catalysts are in a separate phase from reactants. Reactions take place at the interfaces between catalysts and reactants. Catalytic oxidation of hydrogen is an example of this kind. On a Pt or Rh surface hydrogen reacts with oxygen at room temperature to produce water. The product, in this case water, can easily be removed from the metal catalysts. In heterogeneous catalysis, catalysts are often solids while reactants and products are usually in a liquid or a gas phase. If the product is of low molecular weight and volatile, e.g. acetaldehyde, the separation can usually be achieved by simple distillation and the industrial applications are therefore able to be designed on a continuous basis (the separation step is integrated in the production process). Because of the

easier separation, handling, recovery and higher stability, heterogeneous catalysis is industrially very attractive.

1.2 Heterogeneous enantioselective catalysis

1.2.1 Important definitions

Stereoisomer

In chemistry, molecules which have the same chemical compositions and identical connectivity of atoms but different from each other in the way that the atoms are oriented in three dimensional space are referred to as stereoisomers. Fig. 1.1 shows the three stereoisomers of tartaric acid, namely, (R,R)-tartaric acid, (S,S)-tartaric acid and meso-tartaric acid.

(R,R)-tartaric acid   (S,S)-tartaric acid   meso-tartaric acid
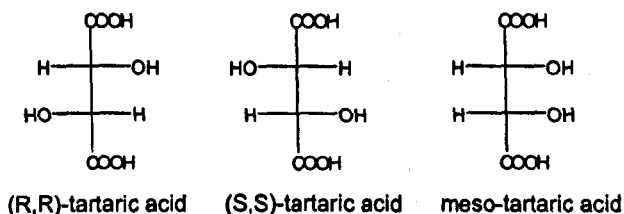
Figure 1.1: *Tartaric acid has three stereoisomers. (R,R)-tartaric acid and (S,S)-tartaric acid molecules are chiral while meso-tartaric acid molecule is achiral.*

Chiral molecules

A close look shows that, (R,R)-tartaric acid and (S,S)-tartaric acid are actually mirror images of each other. Thus, one molecule can never be superimposed onto the other on the same plane, just like a left hand and a right hand (see Fig. 1.2(a)). Molecules of this kind are called chiral molecules. The term "chiral" is derived from the Greek word "cheir", which means "hand". Meso-tartaric acid, however, is identical to its mirror image. It is therefore "achiral" (symmetric). A molecule is chiral if it is not superimposable on its mirror image[3].

Chiral center

Bromochlorofluoromethane is one of the simplest chiral molecules, see Fig 1.2(b). It is chiral because it has a tetrahedral carbon atom which carries four different groups, H, F, Cl and Br. Such a carbon atom (carrying four different substituents) is known as a "chiral center", or a "stereogenic center". It plays an important role in stereochemistry. A molecule which has one chiral center lacks a plane of symmetry and therefore must be chiral. However, molecules that contain more than one chiral centers can be either chiral (e.g. (R,R)-tartaric acid) or achiral (e.g. meso-tartaric acid).
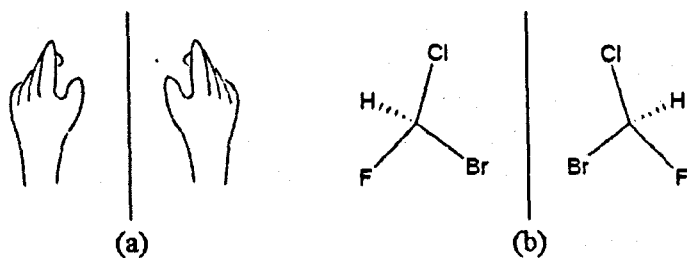
Figure 1.2: *(a) Our hands are chiral, because they are mirror images to each other but the left hand cannot be superimposed on the right one. (b) (S)-bromochlorofluoromethane (left) and (R)-bromochlorofluoromethane (right)*

Enantiomer

Two structures that are not identical but mirror images of one another are called "enantiomers". (R,R)-tartaric acid and (S,S)-tartaric acid are a pair of enantiomers. In achiral environments, two enantiomers are chemically identical. Take (S)-alanine and (R)-alanine as examples. The two enantiomers have the same Nuclear Magnetic Resonance (NMR) and Infrared (IR) spectra, alanine produced in laboratories is always racemic (a mixture of two enantiomers in equal proportions). Their physical properties are also identical, with only one important exception: (S)-alanine rotates a beam of plane-polarized light passing through it to the right while (R)-alanine rotates the light to the left. Racemic alanine, however, lets the light pass without rotations. As a matter of fact, all pure chiral substances rotate the plane-polarized light passing through it to some extent. This property is known as "optical activity" of chiral molecules [4].
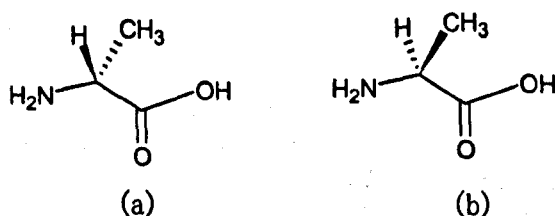
Figure 1.3: *(a) (R)-alanine and (b) (S)-alanine have identical physical properties except for their optical activity*

1.2.2 Labeling an enantimoer

Two methods have been used to label an enantiomer. The first one uses their characteristic optical activities: the enantiomer that rotates plane-polarized light to the right is referred to as the (+)-enantiomer (or the dextrorotatory enantiomer) while the other one which rotates plane-polarized light to the left is referred to

as the (-)-enantiomer (or the laevorotatory enantiomer). The advantage of this labeling method is that it clearly reflects the optical activity of each enantiomer. However, a main drawback is that the configurations of enatiomers cannot be deduced from their names. For instance, in Fig. 1.1, one can hardly say which tartaric acid is the (+)-enantiomer, unless their optical activities have been tested.

The second method describes enantiomers by the configurations of their chiral centers. Firstly, a priority number is assigned to each substituent of the chiral center. The group with higher atomic number has higher priority. Take the chiral center of a bromochlorofluoromethane molecule as an example (Fig. 1.2(b)). The sequence of atomic numbers is Br>Cl>F>H and so is the sequence of priority numbers. Secondly, one orients the right hand so that the thumb points from the chiral center towards the lowest priority substituent (H atom in this example). Then one curls the fingers to connect the other three substitutents. If the connection is in a sequence of decreasing priority numbers then chiral center is assigned as (R), which is for "rectus" in Latin, meaning right. Otherwise, if the connection is in a sequence of increasing priority numbers, the chiral center is denoted as (S), which is for "sinister" in Latin, meaning left. Thus, in Fig. 1.2(b), the enantiomer on the left is called (S)-bromochlorofluoromethane and the enantiomer on the right is called (R)-bromochlorofluoromethane. Likewise, in Fig 1.1, the tartaric molecule on the left has two chiral centers. Both of them are of (R) configurations. Thus it is labeled as (R,R)-tartaric acid. This labeling method is extremely useful for complicated molecules which contain more than one chiral centers, because one can immediately deduce the configuration of an enantiomer from its name.

1.2.3 Importance of chirality

Chirality of nature

Since two enantiomers are chemically and physically very similar, it is quite surprising that they can usually be distinguished by living organisms, not only humans but also animals, plants and even bacteria. For instance, the smells of oranges and lemons differ in being two versions of the same molecule, limonene(see Fig. 1.4(a)). Pseudomonas putida, which is a bacterium, enantioselectively produces 1-bromo-2,3-benzenediol from the achiral bromobenzene (see Fig. 1.4(b)). In our bodies, all the amino acids are "left-handed" (Laevoratory), while sugars in DNA and RNA are all "right-handed" (Dextrorotatory). A truth that underlies these interesting phenomenons is that the nature is asymmetrical. All living systems on the earth are chiral environments. The uniform chirality of natural sugars (Dextrorotatory) leads to the larger scale chirality of all living structures. Thus, racemic polypeptides cannot form specific structures required by enzymes and a single wrong-handed monomer would be able to disrupt the stability of DNA helixes.

As all living systems are chiral, they will react differently to opposite enantiomers. Today, all drugs containing chiral centers are required to be analyzed in

4

(R)-limonene   (S)-limonene   Bromobenzene   1-Bromo-2,3-benzenediol

(a)   (b)
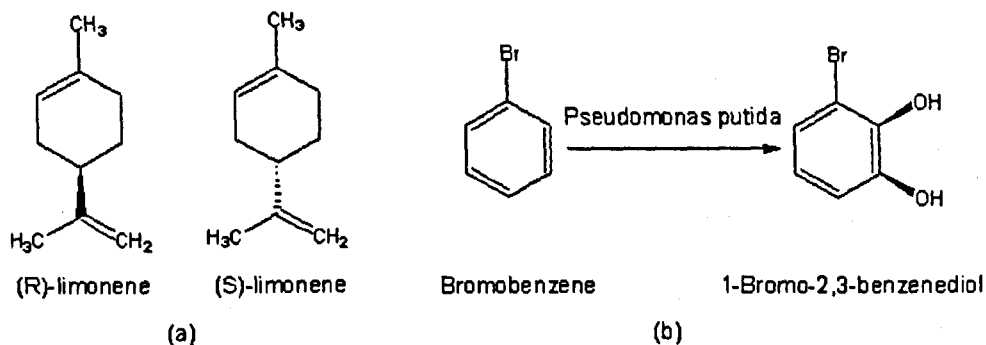
Figure 1.4: *(a)The (R)-limonene smells round and orangey while the (S)-limonene smells sharp and lemony. (b)Pseudomonas putida enantioselectively produces 1-bromo-2,3-benzenediol from the achiral bromobenzene*

terms of the effect of each enantiomer, and most of these drugs need to be sold in a homochiral form.

1.2.4 Chiral catalysis

In symmetric environments

In achiral environments, two enantimoers are chemically identical because the symmetric difference of two configurations does not lead to any kinetic or thermodynamic differences in reaction mechanisms. Take a cyanohydrin production reaction as an example (see Fig. 1.5). Here, the carbonyl group has two faces; a cyanide ion could approach the carbonyl group either from the front face or the back face and therefore yield two products: (R)-cyanohydrin and (S)-cyanohydrin. As the two reaction mechanisms are of equivalent probability, the product obtained is always a 1:1 mixture of two enantiomers. If, for instance, a homochiral form of the product is required, the separation step would be very difficult and expensive. Even if the separation can be achieved via resolution, i.e. crystallization of diastereomeric adducts, it would be highly wasteful (at least 50 percent of the product is wasted). In industry, producing one enantiomer through a resolution is always the last option unless recycling is possible.

(a)   (b)
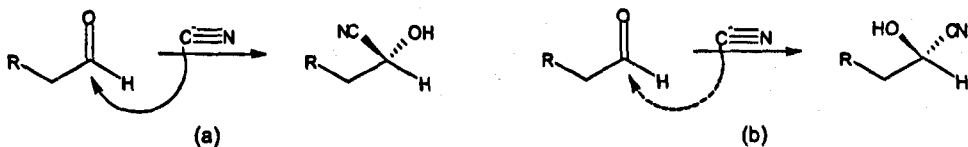
Figure 1.5: *(a)A cyanide ion attacks from the front face of carbonyl group. (b)A cyanide ion attacks from the back face of carbonyl group*

5

In asymmetric environments

Fortunately, enantiomerically enriched products can be directly made in asymmetric environments. If, for example, the above carbonyl-containing molecule possesses a chiral center in an α position, the two faces of carbonyl group would become unequal for an incoming cyanide ion. As shown in Fig. 1.6, the α carbon atom carries a carbonyl function group, a large-sized substituent L, a medium-sized substituent M and a small-sized substituent group S. When the cyanide ion approaches the carbonyl group, it will feel different stereo repulsions from the two faces and prefer to attack from the small substituent's side. The size of substituent L is related to the selectivity of the reaction. If the substituent L is large enough so that no cyanide ion can attack from its side, the produced cyanohydrin will be enantiomerically pure. Similar selectivity can also be obtained if the large-sized substituent is replaced or modified by a negative charged function group. Both kinds of substituents play a role in hindering the cyanide ion to attack from their side.
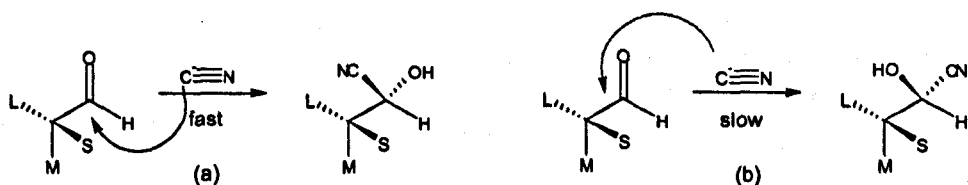
Figure 1.6: *(a)Cyanide ion preferentially attacks from front face of carbonyl group. (b)Cyanide ion attacking from back face of carbonyl group is hindered by the large-sized substituent.*

Chiral catalysis

Today, the most economical option to produce an enantiomercially pure product is using chiral catalysts: a small amount of chiral material which can generate a large amount of chiral product from achiral starting materials. Take the asymmetric epoxidation of allylic alcohols as an illustration, Fig. 1.7 (a) shows the configuration of the oxidized Ti catalyst. One approach is favored because the transition state needs to be stabilized by fitting the specific shapes of catalyst and reactant (see Fig. 1.7 (b) and (c)). In the last step, the Ti catalyst is regenerated when the epoxide is formed. In 2001, the Nobel Prize in Chemistry was awarded to research in this area (asymmetric homogeneous catalytic oxidation and asymmetric catalytic hydrogenation).

Heterogeneous catalysis in enantioselective chemistry

Most chiral catalysts cannot be derived from natural products but need to be carefully designed and synthesized. Producing a complex molecule like the one in Fig. 1.7(a) is often tortuous, time consuming and very expensive. It is a natural step to make chiral catalysts heterogeneous like for other catalytic applications.
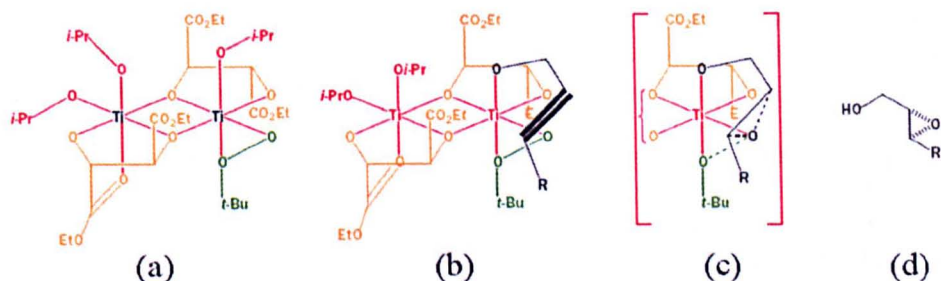
Figure 1.7: *(a) The oxidized chiral Ti catalyst (b) Only one approach of allylic alcohol can be stabilized (c) Transition states of the reaction (d) The formation of an epoxide* [4]

The past several decades have witnessed substantial advances in this area. But very few approaches are really successful. A crucial problem is that enantioselectivity is often lower over the heterogeneous catalysts than over the homogeneous counterparts[2]. Especially in the early years, the enantioselectivity was very low and often not even determined quantitatively.

In 1940, Nakamura opened the door of using chiral auxiliaries to modify heterogeneous catalysts [4]. Since then, absorbing functional organic molecules on metal surfaces to produce chiral architectures on two dimensional surfaces has been widely applied on different surfaces and chiral modifiers [5, 6]. In some reactions, this method has resulted in very high selectivity. As a benchmark[7, 8], the hydrogenation of methylacetoacetate (MAA) has been reported to be very stereodirected on a tartaric acid modified Ni surface. The enantiomeric excess (e.e.) is over 90%, see Fig. 1.8.

Despite the dramatic progress of this approach in scientific research, few re-
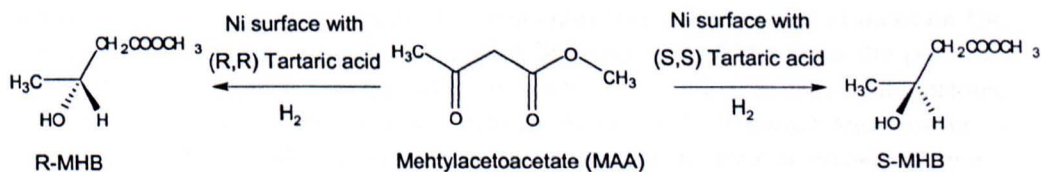
Figure 1.8: *On a (R,R)-tartaric acid modified Ni surface, the product are dominated by R-methyl-3-hydroxybutrate (MHB). On an (S,S)-tartaric acid modified Ni surface, the product are doninated by S-MHB*

actions are used on an industrial scale today [8]. The reason are that (i) only a particular combination of a metal, a modifier and a substrate lead to good enantioselectivity. The types of reactions are quite limited. (ii) Chiral modifiers and many metal precursors are expensive and not easily available. Over the years,

much effort has been devoted to improve the catalytic performance and lower the cost of catalysts. But progress is still limited by our lack of deeper understanding of the reaction mechanisms.

However, our empirical knowledge of most heterogeneous reactions is quite substantial [9]. For instance, we know that Pd and Pt-Rh are the best exhaust catalysts for NO removal [10], Pt, Pd and Ag are widely used in catalytic oxidations [9], Co, Fe and Rh are wonderful Fischer-Tropsch catalysts [11], and Ru and Fe are the best ammonia synthesis catalysts [9]. Nevertheless, only a detailed investigation of the underlying elementary steps, e.g. bonding of molecules onto a surface, diffusion of adsorbates, transitional states of a reaction and formation of the products can yield a deeper understanding of the mechanism of these catalytic processes. Due to the complex interactions among surface atoms and adsorbates and the coupling of their electronic states, it is unknown, at present, "what exactly happens on the surface at atomic scale". This is not only a problem for large organic molecules on complicated surfaces but also for atoms and small molecules on single crystal surfaces. Thus, the study of some "simple" and well-defined model systems will help to obtain a detailed insight into the nature of enantiodifferentiations taking place at surfaces.

1.3   The Scanning Tunneling Microscopy

To date, the ongoing scientific research on surface model systems has been driven greatly by the concurrent emergence of new experimental techniques developed to study surfaces under ultra-high vacuum (UHV) conditions. These techniques are capable of giving very detailed information on the geometric and electronic structure of the surface as well as its chemical compositions; they are based on the interaction of electrons, ions and photons with the surface. To give a few examples, we mention low-energy electron diffraction (LEED), high-energy ion scattering (HEIS) and angle-resolved ultraviolet photoelectron spectroscopy (ARUPS). However, the structural information obtained by these techniques is averaged over large areas compared to the characteristic atomic distances on the surface. By contrast, Scanning Tunneling Microscopy (STM) offers the possibility of direct, real-space determination of surface structures in three dimensions, including non-periodic structures. Since its invention by Binning and Rohrer in 1982 [12, 13], the STM has become the most ubiquitous tool in surface science.

As shown in Fig.1.9, the setup of an STM basically consists of a sharp tip which ideally has only a single apex atom, i.e. it is atomically sharp, mounted on a piezo-tube; this tip can be moved in all three dimensions on the picometer $(10^{-12}m)$ scale.

When the sharp tip is brought close to the surface, without actual contact (the distance then is in the range of a few Å), a narrow vacuum potential barrier is created. Provided an external bias voltage, electrons can tunnel through the vac-
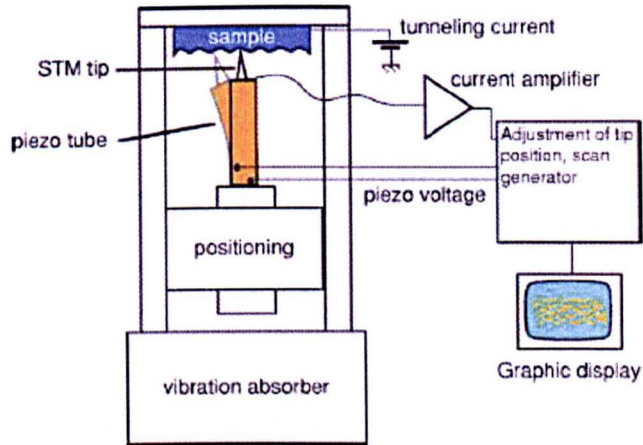
Figure 1.9: Schematic representation of STM operation. As the tip scans the surface, its height is adjusted to maintain a constant tunneling resistance. [14]

uum barrier from the tip into the surface and vice versa, leading to a tunneling current flowing between the two leads.

In STM operation, the current is usually set to a fixed value. When the tip scans the surface in two dimensions, its vertical position is varied by an electronic feedback on the piezo-crystal, so that the current remains constant. As the tip scans the surface in two dimensions, recording its height changes gives an electronic contour map of the surface. Provided elaborate vibration damping, the best instruments today are capable of a vertical resolution better than 1 pm, or one 200th of an atomic diameter[15].

Although this seems to be rather simple, the whole process is highly influenced by (i)the chemical composition of surface and tip, (ii)the tip-surface distance, (iii)the electronic structure of both systems, (iv)the chemical interactions between surface and tip atoms and (v)the electrostatic interactions of sample and tip[15, 16, 17, 18]. Hence the information that an STM map reveals is much more than simply a surface topography; the interpretation of those images is consequently not straightforward.

Qualitatively, the generation of tunneling current can be described as follows: The potential barrier in the vacuum range of the two subsystems determines the decay of surface wave-functions in the same range. For a square potential barrier (see Fig. 1.10), the current decays exponentially with a constant exponent:

$$I(z) \propto e^{-(\frac{\sqrt{2m\phi}}{\hbar})z}.$$

Here, ϕ is the work function (the energy that required to move an electron from the fermi level into vacuum) of the two surfaces and z is the distance from the

9
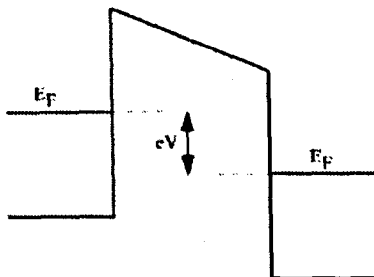
Figure 1.10: Schematic representation of the potential barrier between the two subsystems when a bias voltage is applied.

nucleus of the surface atoms to the nucleus of the tip apex. Although it is often a good approximation, in practice the work-function of sample and tip are generally not equal, the decay of electron states is not constant, and the density of states above the surface also varies with the exact location[15]. Since the high STM precision is based on its sensitivity of surface-tip distance, a better model is needed. In Chapter 2, we shall present a more refined model to calculate the tunneling current.

1.4 Succinic acid on a Cu(110) surface

Fortunately, building chiral surfaces is not simply limited to the approach of adsorbing chiral molecules on achiral surfaces[19, 20]. Recent STM experiments have reported a surprising chiral behavior of intrinsically achiral organic molecules when adsorbed on a metal surface : at around 400K, succinic acid molecules adsorbed on Cu(110) formed two thermally stable and asymmetric superstructures, in which succinate trimers extend in rows along the $[\bar{1}12]$ and $[1\bar{1}2]$ directions[21], see Fig. 1.11. An atomistic understanding of the nature of this intriguing ordering process would obviously enhance the possibility of developing future enantioselective heterogeneous catalysts from cheap and achiral starting materials. The two simple and well known subsystems (Cu(110) surface and succinic acid) suggest a suitable model system for experimental and theoretical investigations.

From the experimental side, in addition to STM, Low Energy Electron Diffraction (LEED) has been used to determine the (1 1, -9 0) and (9 0, -1 1) symmetry of chiral domains. Data obtained by Temperature Programmed Desorption (TPD) indicated that the adsorbate-metal interactions are so strong that intra-molecular bonds breaks prior to metal-molecule bonds. More detailed adsorption features have been identified by Reflection Absorption Infrared Spectroscopy (RAIRS) analysis[21]. As the coverage of adsorbates is increased, the disappearance of a characteristic $\nu_{C=O}$ stretch vibration (at $1700 cm^{-1}$) reveals there is no carboxylic groups present in the adsorbates. By contrast, the appearance of doublets

10
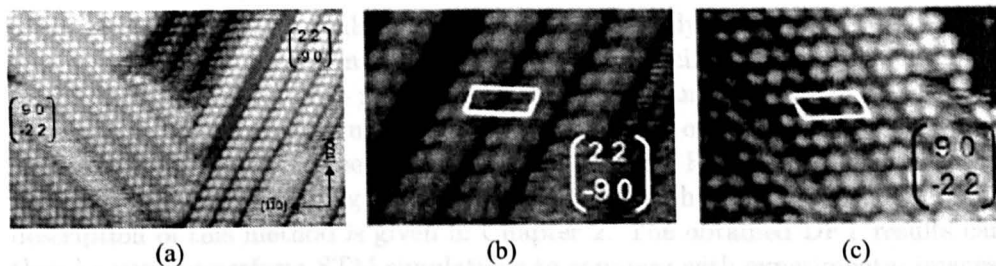
Figure 1.11: (a) An STM image (260Å ×170Å) of succinic acid molecules adsorbing on a Cu(110) surface. (b) and (c) are enlarged images for the two chiral domains. Parameters : $V = -0.21V, I = 0.15nA$ [21]
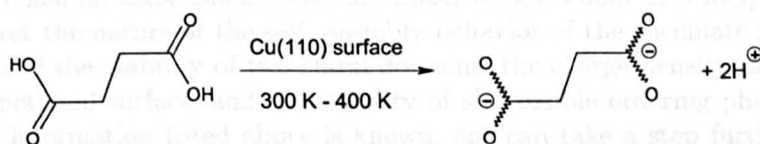
Figure 1.12: Contacting a Cu(110) surface, succinic acid molecules doubly deprotonate and form chemical bond with Cu surface via its two bidentate carboxylate functionality

at $1425cm^{-1}$ and $1408cm^{-1}$ strongly suggest coupled vibrations of the identical COO^- oscillators on each molecule[21, 22, 23]. The adsorbing unit has therefore been assigned to a doubly deprotonated bisuccinate, which bonds to the Cu surface via its two bidentate carboxylate functions (see Fig. 1.12). The produced hydrogen atoms are known to diffuse on Cu surfaces and undergo recombinative desorption into the gas phase as H_2[24, 25, 26].

1.4.1 Motivation of theoretical studies

In spite of extensive experimental investigations of this model system, the atomic mechanism of this self organization process is still unclear. For instance, the local adsorption geometry of a bisuccinate molecule is unknown, the nature of intermolecular interactions have not be characterized, and even more importantly, the electronic properties and/or the electron charge distribution of the modified chiral surface have not been fully investigated.

Generally, experimental analysis alone cannot disentangle these questions unambiguously. Because the information confined within the atomic or electronic scale can neither be obtained directly from experimental measurements nor interpreted unambiguously. Nevertheless, the missing information can usually be filled in by

11

precise first-principles calculations (treating many-body systems on the basis of first principles of quantum mechanics, without introducing any empirical parameters). With the help of first- principles methods, one can often obtain an accurate description of electronic structures of solids, surfaces, or clusters with moderate computational efforts. In the last few decades, Density Functional Theory (DFT) has been the primary driving force of developments in this area [27, 28], a detailed description of this method is given in Chapter 2. The obtained DFT results can then be used to perform STM simulations to compare with experimental images.

In addition to the lack of knowledge in the microscopic regime, the self assembly processes also depends on a statistical interplay of all microscopic process at finite temperatures. Thus the obtained first-principles description should be further treated statistically to investigate the phase diagram of this substrate-adsorbates system. When all these calculations are finished, one would in principle be able to interpret the nature of the self assembly behavior of the succinate molecules, the origin of the stability of two chiral domains, the charge density distribution of the functional surface, and the stability of all possible ordering phases. Only if all the information listed above is known, one can take a step further by designing a heterogeneous catalysis circle or refining the electronic properties of the functional surfaces.

1.4.2 Previous theoretical investigations

Even though "simple" and potentially very important to a number of application areas, the succinate-Cu(110) system has not been systemically investigated by theorists yet. To our best knowledge, only a few basic DFT calculations have been performed by Antonio [29] and Darling [30] in 2001 and 2007 respectively. According to Antonio's calculations [29], at ground state, a single bisuccinate acid
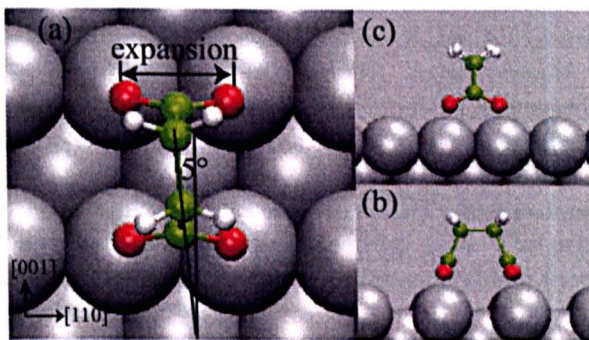
Figure 1.13: (a) Top view of the adsorption configuration of bisuccinate on a Cu(110) surface. (b) and (c) are the side views.

straddled straight across the long-bridge between Cu rows and bonded to the on-

top sites of top-layer Cu atoms with its four oxygen atoms (see Fig.1.12). Upon adsorption, the molecular skeleton undergos a slight deformation by twisting its backbone about 5° away from the [100] direction. Similarly to tartaric acid on Ni surfaces[31], this spontaneous distortion of the molecular skeleton destroys both the molecular and local surface mirror plane, and yields a chiral site on a Cu(110) surface. Since a bisuccinate molecule is achiral, this twist can either be clockwise or counterclockwise with an equal probability. As a consequence, the ratio of the two enantiomeric chiral succinates should be 1:1, which is consistent with the experimental result. Based on this theoretical adsorption model, the chiral domains have been characterized as the self-assembly of the homochiral adsorbates. In addition to the chiral twist of molecules, a lateral structural deformation has been found at the Cu substrate. Underneath each bisuccinate adsorbate, metal atoms display a Cu-Cu expansion in [110] direction, which consequently results in a stress to the surrounding surface atoms. This surface stress was suggested to be responsible for the formation of succinate trimers because it would dynamically hinder four bisuccinates siting next to each other in [110] direction. Based on the

(a)   (b)
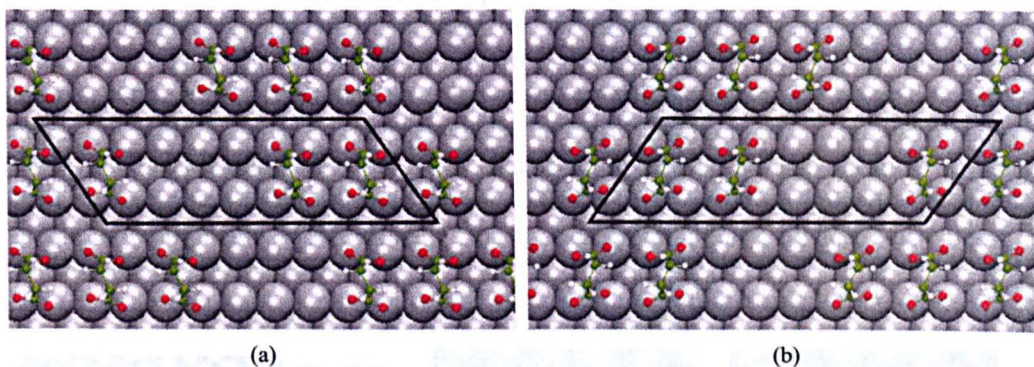
Figure 1.14: Antonio's theoretical model of the self assembled structure of succinates on a Cu(110) surface: (a) and (b). In this model, the two chiral domains are composed by stripes of succinate trimers. The growth direction of stripes are determined by the chirality of the footprint of each succinate molecule.

experimental symmetric data and the DFT calculations, a geometric model of chiral domains has been proposed, see Fig. 1.14. The organizational chirality has therefore been attributed to the spontaneous chiral footprint and surface stress induced by adsorbates.

Darling's calculations were carried out for a different ordering pattern of succinate molecules on a Cu(110) surface. Instead of two mirror enantiomorphous chiral domains, the experimental STM images indicate that the succinate molecules are self-organized to form chiral lines extending along the [001] direction. In the [110] direction, however, the adjacent chiral lines are mirror images of each other. Thus the overlayer is globally racemic and displays a p(4x2) symmetry. See, Fig. 1.15 (a). Based on the experimental symmetry, a set of Density Functional The-

ory calculations have been performed. Liu and Darling suggest that the chiral lines in Fig. 1.15 (a) are formed by a mixture of local chiral twists of succinate adsorbates in a 1:1 ratio. In [30], Darling has calculated the energetics of two possible adsorption configurations of succinate molecules on a Cu(110) surface. The first one is exactly the Antonio adsorption model (Fig. 1.15(b)), while the second one is shown in Fig. 1.15(c), where the Cu atoms are still bonded to the COO functional groups of the succinate molecule. However one of the bonded Cu-atom pair undergoes an offset by one Cu lattice constant in the [110] direction and thus gives rise to a diagonal adsorption site rather than a rectangular one. Interestingly, the adsorption energy per molecule in the second model has been reported to be lower than the first one by 15.9 kJ/mol. Comparing the $O \cdots H$ distances in the relaxed configurations, this energy difference has been attributed to the distortion of the molecular configuration and the formation of weak hydrogen bonds in and between the adsorbates [30].

It is quite interesting for many a few aspects. First of all, which adsorption model is the "real" ground state adsorption configuration? Secondly, since these two calculations were carried out based on two different coverages of adsorbates, is it possible that the adsorption configuration of succinate depends on the coverage? What kind of interactions play a role in the self-assembly process? Are there any other ordered structures which have not be observed? If yes, what are they, and how stable they could be? Surprisingly, all these aspects of obvious importance to an atomic understanding of this self organization behavior as well as for any potential application of this model system, have hitherto not been addressed.

(a)   (b)   (c)
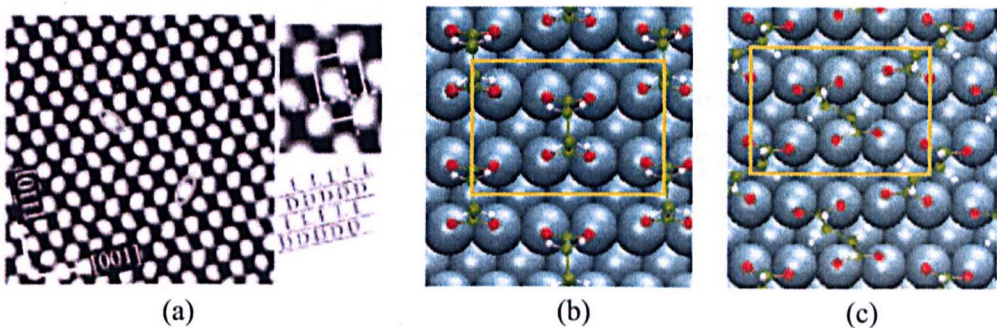
Figure 1.15: (a) The p(4x2) STM topography of the succinate molecules on a Cu(110) surface ($80\mathring{A} \times 80\mathring{A}$) and an enlarge image of the local chiral features ($17\mathring{A} \times 20\mathring{A}$). The angles of the long axes of the oval shaped features and the [001] direction are $\pm 65°$. Parameters: $I = 0.36nA$, $V = -0.88V$. (b) Theoretical p(4x2) unit cell based on Antonio's adsorption model. (c) Theoretical p(4x2) unit cell based on Darling's new adsorption model [30].

1.5   Objects of this work

In the previous DFT calculations all the unit cells were designed in order to perfectly fit the experimental images. However, as is well known, in most cases of molecular adsorption it is very difficult to relate the observed electron densities with the absolute atomic positions [15]. Besides, as will be discussed later in Chapter 2, the energetic results of molecular interactions obtained with DFT are only relative differences of energy levels in the configuration space at zero temperature. Under the STM operating conditions, the succinate adsorbates and the copper surface are heated up to 400K. The mechanism of the self-assembly process therefore depends not only on the description of a few energy levels in the microscopic scale, but also the interplay of these energy levels in the mesoscopic regime at realistic experimental conditions. Thus, in principle, a reliable theoretical description of such self-assembly process should include both accurate evaluations of all possible energy levels in the microscopic regime and a statistical mechanics treatment of their interplay at the mesoscopic scale. To this end, it is far from enough to just analyze the placement of a molecule on a particular surface unit cell. Instead, multi-scale modeling is required.

In this work, the microscopic energy levels of simple ordered lattice configurations are characterized by first principle DFT calculations. For this method often represents a fair compromise between sufficiently high accuracy and feasible computational cost. The energy parameters of the disordered structure, due to the high level of complexity, are evaluated by an effective cluster expansion method based on a set of DFT calculations. The resulting energetics of all configurations in the microscopic scale are then employed to parameterize a Lattice Gas Hamiltonian, which can subsequently be used to address the mesoscopic ordering behavior at finite temperatures by means of Monte Carlo simulations. Based on the results of systematic Monte Carlo simulations, we aim at an atomic understanding of the mechanism of self-assembly and an ab-inito prediction of a full phase diagram of this adsorbate-substrate system. The accuracy of prediction can then be checked by further DFT calculations and STM simulations.

Chapter 2

Theoretical Background

2.1 Electronic Structure Calculations

In order to obtain a microscopic description of chemical reactions taking place at molecule-metal interfaces, one must in principle solve the Schrödinger equation for an enormous number (N) of electrons, which interact according to the laws of quantum mechanics. Since the problem is highly nontrivial and the computational effort grows very rapidly with increasing N, one has to find a method which can simplify the mathematical description of the system. Nowadays, the most successful and most often used technique to solve the many-electron problem is based on the density functional (DFT) method of Hohenberg and Kohn. In this section, the basics of Density Functional Theory together with the algorithms implemented for calculations are described.

2.1.1 The Schrödinger Equation

The ultimate goal of most theoretical approaches to quantum systems is finding the wavefuction of a many-body system consisting of M ions and N electrons, where N and M are of the order of Avogadro's constant. For a stationary state, if the many-body wavefuction Ψ is known, any physical property A can be calculated by taking the expectation value of the associated (Hermitian) operator $\hat{A}$.

$$A[\Psi] = \langle \Psi | \hat{A} | \Psi \rangle \tag{2.1}$$

In quantum mechanics, stationary states Ψ and the corresponding energies E of a many-body system are determined by the time independent Schrödinger equation.

$$\hat{H}\Psi = E\Psi \tag{2.2}$$

where the many-body Hamilton operator $\hat{H}$ includes the following contributions:
(i) The kinetic energy of the N electrons:

$$\hat{T}_e = -\sum_{k=1}^{N} \frac{\hbar^2}{2m} \nabla_{r_k}^2 \tag{2.3}$$

where m is the mass of an electron and $\nabla^2_{\mathbf{r}_k}$ is the Laplacian operator acting on the kth electron.

(ii) The kinetic energy of the M ions:

$$\hat{T}_{Ion} = -\sum_{I=1}^{M} \frac{\hbar^2}{2M_I} \nabla^2_{\mathbf{R}_I} \tag{2.4}$$

here, M_I is the mass of the Ith ion and $\nabla^2_{\mathbf{R}_I}$ is the Laplacian operator acting on the Ith ion.

(iii) The electron-electron interaction (considering only the electrostatic interaction):

$$\hat{V}_{e-e}(\{\mathbf{r}_k \sigma_k\}) = \frac{1}{2}\frac{1}{4\pi\epsilon_o} \sum_{\substack{k,k' \\ k \neq k'}}^{N,N} \frac{e^2}{|\mathbf{r}_k - \mathbf{r}_{k'}|} \tag{2.5}$$

Where "$\{\mathbf{r}_k \sigma_k\}$" denote the position and spin of the kth electron.

(iv) The interactions of the ions:

$$\hat{V}_{Ion-Ion}(\{\mathbf{R}_I\}) = \frac{1}{2}\frac{1}{4\pi\epsilon_o} \sum_{\substack{I,I' \\ I \neq I'}}^{M,M} \frac{e^2}{|\mathbf{R}_I - \mathbf{R}_{I'}|} Z_I Z_{I'} \tag{2.6}$$

Here, the "$\{\mathbf{R}_I\}$" and "Z_I" are used to represent the coordinate and charge of the ion I.

(v) The electron-ion interaction:

$$\begin{aligned}
\hat{V}_{e-Ion}(\{\mathbf{r}_k \sigma_k\}, \{\mathbf{R}_I\}) &= -\frac{1}{4\pi\epsilon_o} \sum_{k=1}^{N} \sum_{I=1}^{M} \frac{e^2}{|\mathbf{R}_I - \mathbf{r}_k|} Z_I \\
&= -\frac{e^2}{4\pi\epsilon_o} \sum_{k=1}^{N} \int \frac{\eta^+(\mathbf{R})}{|\mathbf{R} - \mathbf{r}_k|} d^3\mathbf{R} \tag{2.7}
\end{aligned}$$

where $e\eta^+$ denotes all charge densities of the ions, assuming the positive charge is distributed continuously.

Consequently, we may rewrite the time independent many-body Schrödinger equation as:

$$(\hat{T}_e + \hat{T}_{Ion} + \hat{V}_{e-e} + \hat{V}_{Ion-Ion} + \hat{V}_{e-Ion})\Psi(\{\mathbf{r}_k \sigma_k\}, \{\mathbf{R}_I\}) = E\Psi(\{\mathbf{r}_k \sigma_k\}, \{\mathbf{R}_I\}) \tag{2.8}$$

where $\Psi(\{\mathbf{r}_k \sigma_k\}, \{\mathbf{R}_I\})$ is a function of the coordinates of the interacting $M + N$ particles. It has long been known that there is no analytical solution for equation (2.8). We therefore have to use some, justified, approximations.

2.1.2 The Born-Oppenheimer Approximation

The first important assumption used in many-body electronic structure calculations was formulated in 1927 by Born and Oppenheimer[32]. The approximation is based on the fact that the electrons are much lighter than the ions (even the lightest ^1H atom is a factor of 1800 heavier than an electron), therefore the characteristic frequencies of electrons and ions are very different. The electronic and ionic motions can be separated by assuming that electrons move so quickly compared to the movement of ions that they are always in adiabatic equilibrium (in their ground state) with respect to ionic geometries. The consequence is that the wavefunction for the whole system can be decoupled into separate wavefunctions for the electrons and the ions, the total energy can then be considered as a function of the ionic configurations in phase space only. For a given ionic geometry $\{\mathbf{R}_I\}$, the eigenstates Φ_ν of the N electrons are defined by

$$\hat{H}^e(\{\mathbf{R}_I\})\Phi_\nu(\{\mathbf{R}_I\},\{\mathbf{r}_k\sigma_k\}) = E^e_\nu\{\mathbf{R}_I\}\Phi_\nu(\{\mathbf{R}_I\},\{\mathbf{r}_k\sigma_k\}) \qquad (2.9)$$

where

$$\hat{H}^e = \hat{T}_e + \hat{V}_{e-Ion} + \hat{V}_{e-e}. \qquad (2.10)$$

The overall wavefuction of the $M+N$ system can therefore be represented by the product of ionic wavefuntion $\Lambda_\nu(\{\mathbf{R}_I\})$ and the electronic eigenfunctions of each ionic configuration $\Phi_\nu(\{\mathbf{R}_I\},\{\mathbf{r}_k\sigma_k\})$.

$$\Psi = \sum_\nu \Lambda_\nu(\{\mathbf{R}_I\})\Phi_\nu(\{\mathbf{R}_I\},\{\mathbf{r}_k\sigma_k\}) \qquad (2.11)$$

The corresponding many body Hermitian operator can be written as

$$\hat{H} = \hat{H}^e + \hat{V}_{Ion-Ion} + \hat{T}_{Ion} \qquad (2.12)$$

here, only $\hat{T}_{Ion}$ cannot be interchanged with Λ_ν.

$$\hat{T}_{Ion}(\Lambda_\nu\Phi_\nu) = \sum_\nu \sum_{I=1}^M -\frac{\hbar^2}{2M_I}\nabla^2_{\mathbf{R}_I}(\Lambda_\nu\Phi_\nu) \qquad (2.13)$$

with

$$\begin{aligned}
\nabla^2_{\mathbf{R}_I}(\Lambda_\nu\Phi_\nu) &= \nabla_{\mathbf{R}_I}[\Lambda_\nu(\nabla_{\mathbf{R}_I}\Phi_\nu) + (\nabla_{\mathbf{R}_I}\Lambda_\nu)\Phi_\nu] \\
&= \Lambda_\nu(\nabla^2_{\mathbf{R}_I}\Phi_\nu) + 2(\nabla_{\mathbf{R}_I}\Lambda_\nu)(\nabla_{\mathbf{R}_I}\Phi_\nu) + (\nabla^2_{\mathbf{R}_I}\Lambda_\nu)\Phi_\nu \quad (2.14)
\end{aligned}$$

Multiplying both sides of the many-body Schrödinger equation by Φ^*_μ and integrating over the electronic coordinates, we obtain

$$\begin{aligned}
<\Phi_\mu|\hat{H}|\Psi> &= E\Lambda_\mu \\
&= (E^e_\mu + T_{Ion} + V_{Ion-Ion})\Lambda_\mu \\
&\quad -\sum_\nu \sum_{I=1}^M \frac{\hbar^2}{2M_I}[\langle\Phi_\mu|\nabla^2_{\mathbf{R}_I}|\Phi_\nu\rangle\Lambda_\nu + 2\langle\Phi_\mu|\nabla_{\mathbf{R}_I}|\Phi_\nu\rangle\nabla_{\mathbf{R}_I}\Lambda_\nu]
\end{aligned}$$

18

where the second line arises from the contribution of electron-phonon coupling. However, according to the Born-Oppenheimer approximation, electrons are assumed to respond instantaneously to the ionic motion and thus the movement of ions shall not induce transitions from Φ_ν to Φ_μ. Therefore, the matrix elements $\langle\Phi_\mu|\nabla_{\mathbf{R}_I}^2|\Phi_\nu\rangle$ and $\langle\Phi_\mu|\nabla_{\mathbf{R}_I}|\Phi_\nu\rangle$ vanish for $\mu \neq \nu$. For $\mu = \nu$, on the other hand, the term $\langle\Phi_\mu|\nabla_{\mathbf{R}_I}|\Phi_\mu\rangle$ also vanishes because $\langle\Phi_\mu|\nabla_{\mathbf{R}_I}|\Phi_\mu\rangle = \frac{1}{2}\nabla_{\mathbf{R}_I}\langle\Phi_\mu|\Phi_\mu\rangle = 0$, and the diagonal element $-\frac{\hbar^2}{2M_I}\langle\Phi_\mu|\nabla_{\mathbf{R}_I}^2|\Phi_\mu\rangle$ can usually be neglected.

The ionic wavefunction of the M+N system now satisfies the following equation

$$(\hat{T}_{Ion} + \hat{V}_{Ion-Ion} + \hat{H}_\mu^e)\Lambda_\mu = E\Lambda_\mu \qquad (2.15)$$

The trajectory of ions can, due to their large mass, generally be described as classical particles using classical (Newton) equations of motion. The Coulomb summation of $V_{Ion-Ion}$ converges very slowly in both real and reciprocal space, and is usually calculated by the Ewald Method [33]. In the the succeeding sections, the emphasis are put on the calculations of the electronic energy E^e.

2.1.3 The Hartree Approximation

Up to this point, the only unknown term is the electronic energy E^e, which consists of the kinetic energy of the electrons, the potential energy of electrons in the field of the nuclei, and electron-electron interaction. Here, we rewrite the electronic Hamiltonian

$$\hat{H}^e = \hat{T}_e + \hat{V}_{e-Ion} + \hat{V}_{e-e} \qquad (2.16)$$

The attractive potential exerted on the electron at position $\mathbf{r}$ due to the nucleis is often called the "external potential", $V_{ext}(\mathbf{r})$. Despite the decoupling of electronic and ionic motions, the operator $\hat{V}_{e-e}$ still couples with coordinates of all electrons. The solution of the Schrödinger equation with $\hat{H}^e$ is therefore still the many-body wavefunction, which depends on 3N spatial coordinates and N spin coordinates.

The second historically very important approximation is the Hartree approximation [34]. The formulation of this approximation is that each electron is assumed to move in a smooth potential field generated by all ions and other electrons. The movement of an electron therefore does not explicitly depend on the positions of other electrons but on the averaged charge density of the N-electron system. As a consequence, the Hartree approximation maps the original problem of N interacting electrons to a problem of N independent electrons moving in the mean field due to all other electrons. If we write the Hamiltonian for the kth electron as $\hat{H}_k$, then the total Hamiltonian reads

$$\hat{H}^e = \hat{H}_1^e + \hat{H}_2^e + ... + \hat{H}_N^e \qquad (2.17)$$

19

Since the spins of the electrons are not taken into consideration, the electrons are free to commute with each other, the wavefunction $\Psi^{Hartree}$ can therefore be written as the product of N one-electron wavefunctions. In principle, $\Psi^{Hartree}$ can be any arbitrary function in the N-particle Hilbert space, which can be differentiated twice and normalized.

$$\Psi^{Hartree} = \varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2)...\varphi_N(\mathbf{r}_N) \tag{2.18}$$

The eigenstates of the N-electron system can then be constructed from one-electron eigenstates.

$$\hat{H}^e \Psi^{Hartree} = [\epsilon_1(\mathbf{r}_1) + \epsilon_2(\mathbf{r}_2) + ... + \epsilon_N(\mathbf{r}_N)]\Psi^{Hartree} \tag{2.19}$$

The expectation value of $\hat{H}^e$ is obtained from

$$\langle \Psi^{Hartree}|\hat{H}^e|\Psi^{Hartree}\rangle = \sum_{k=1}^{N} \int \varphi_k^*(\mathbf{r}_k)\left[-\frac{\hbar^2}{2m}\nabla^2 + \hat{V}_{ext}(\mathbf{r}_k)\right]\varphi_k(\mathbf{r}_k)d^3\mathbf{r}_k$$
$$+\frac{1}{2}\frac{e^2}{4\pi\epsilon_o}\sum_{\substack{k,k' \\ k\neq k'}}^{N,N}\iint \frac{|\varphi_k(\mathbf{r}_k)|^2|\varphi_{k'}(\mathbf{r}_{k'})|^2}{|\mathbf{r}_k - \mathbf{r}_{k'}|}d^3\mathbf{r}_k d^3\mathbf{r} \tag{2.20}$$

In order to find the best set of $\varphi_k(\mathbf{r}_k)$, we introduce a Lagrange multiplier ϵ_k to guarantee that the one-electron wavefunctions are normalized, and minimize the above equation with respect to the wavefunctions φ_k^*

$$\frac{\delta}{\delta\varphi_k^*}\left[\langle\hat{H}^e\rangle - \sum_{k=1}^{N}\epsilon_k\int|\varphi_k(\mathbf{r}_k)|^2 d^3\mathbf{r}_k\right] = 0 \tag{2.21}$$

Then, we obtain the N single particle (Hartree) equations which are used to determine the functions $\varphi_k(\mathbf{r}_k)$

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + \hat{V}_{ext}(\mathbf{r}_k)\right]\varphi_k(\mathbf{r}_k) + \sum_{\substack{k'=1 \\ k'\neq k}}^{N}\frac{e^2}{4\pi\epsilon_o}\left\langle\varphi_{k'}(\mathbf{r}_{k'})\left|\frac{1}{|\mathbf{r}_{k'}-\mathbf{r}_k|}\right|\varphi_{k'}(\mathbf{r}_{k'})\right\rangle\varphi_k(\mathbf{r}_k) = \epsilon_k\varphi_k(\mathbf{r}_k)$$
$$\tag{2.22}$$

The above function can be rewritten in a single-particle equation form :

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v^{eff}(\mathbf{r}_k)\right]\varphi_k(\mathbf{r}_k) = \epsilon_k\varphi_k(\mathbf{r}_k) \tag{2.23}$$

with

$$v^{eff}(\mathbf{r}_k) = \hat{V}_{ext}(\mathbf{r}_k) + \frac{e^2}{4\pi\epsilon_o}\int\frac{\rho(\mathbf{r}_{k'})}{|\mathbf{r}_{k'}-\mathbf{r}_k|}d^3\mathbf{r}_{k'} - \frac{e^2}{4\pi\epsilon_o}\int\frac{|\varphi_k(\mathbf{r}_k)|^2}{|\mathbf{r}_{k'}-\mathbf{r}_k|}d^3\mathbf{r}_{k'} \tag{2.24}$$

The first term is the external potential due to the ions, the second term is called "Hartree potential", $V_{Hartree}$, which can also be obtained from the Poisson Equation

$$\nabla^2 V_{Hartree}(\mathbf{r}) = -\frac{e^2}{\epsilon_o}\rho(\mathbf{r}) \tag{2.25}$$

and the third term is the "self interaction correction" for the Hartree potential. As can be seen, the "effective potential" itself depends on the solutions $\varphi_k(\mathbf{r}_k)$. The Hartree equations are therefore just "effective" (not real) single-particle equations.

Since the "effective potential" is initially unknown, because the charge density $\rho(\mathbf{r})$ is not known, the Hartree equations are solved by the self-consistent field (SCF) method [33].

2.1.4 The Hartree-Fock approximation

The Hartree-Fock approximation is an extension of the Hartree approximation to include the permutation symmetry of the wavefuncion. According to the Pauli exclusion principle, the total wavefunction must be antisymmetric with respect to an interchange of electrons.

$$\Psi\left(\varphi_1(\mathbf{r}_1,\sigma_1),\varphi_2(\mathbf{r}_2,\sigma_2),...,\varphi_N(\mathbf{r}_N,\sigma_N)\right) = -\Psi\left(\varphi_2(\mathbf{r}_2,\sigma_2),\varphi_1(\mathbf{r}_1,\sigma_1),...,\varphi_N(\mathbf{r}_N,\sigma_N)\right)$$
(2.26)

where $\mathbf{r}_k, \sigma_k$ denote the position and spin of the kth electron respectively. Therefore no two electrons can have the same set of quantum numbers and electrons with the same spin cannot occupy the same state simultaneously.

In order to fulfill the Pauli principle, the many-body wavefunction is represented by the Slater determinant of an $N \times N$ matrix of single particle wavefunctions, which has the form of a linear combination of all possible permutation of the single-particle wavefunctions.

$$\Psi^{HF} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{r}_1,\sigma_1) & \varphi_1(\mathbf{r}_2,\sigma_2) & \cdots & \varphi_1(\mathbf{r}_N,\sigma_N) \\ \varphi_2(\mathbf{r}_1,\sigma_1) & \varphi_2(\mathbf{r}_2,\sigma_2) & \cdots & \varphi_2(\mathbf{r}_N,\sigma_N) \\ \vdots & \vdots & & \\ \varphi_N(\mathbf{r}_1,\sigma_1) & \varphi_N(\mathbf{r}_2,\sigma_2) & \cdots & \varphi_N(\mathbf{r}_N,\sigma_N) \end{vmatrix}$$

The factor $1/\sqrt{N!}$ makes sure that the many-body wavefunction is normalized. Following the same method of minimizing the expectation value of $\langle \Psi^{HF}|\hat{H}^e|\Psi^{HF}\rangle$ as we did in Hartree approximation to determine the best set of single-particle functions, we obtain the Hartree-Fock equations :

$$\begin{aligned} \epsilon_k \varphi_k(\mathbf{r}_k,\sigma_k) = & \left[-\frac{\hbar^2}{2m}\nabla^2 + \hat{V}_{ext}(\mathbf{r}_k) \right] \varphi_k(\mathbf{r}_k,\sigma_k) \\ & + \sum_{k'=1}^{N} \frac{e^2}{4\pi\epsilon_o} \left\langle \varphi_{k'}(\mathbf{r}_{k'},\sigma_{k'}) \left| \frac{1}{|\mathbf{r}_{k'}-\mathbf{r}_k|} \right| \varphi_{k'}(\mathbf{r}_{k'},\sigma_{k'}) \right\rangle \varphi_k(\mathbf{r}_k,\sigma_k) \\ & - \frac{e^2}{4\pi\epsilon_o} \sum_{k'=1}^{N} \delta_{\sigma_k,\sigma_{k'}} \int \frac{\varphi_{k'}^*(\mathbf{r}_{k'})\varphi_k(\mathbf{r}_{k'})\varphi_{k'}(\mathbf{r}_k)}{|\mathbf{r}_{k'}-\mathbf{r}_k|} d^3\mathbf{r}_{k'} \end{aligned}$$
(2.27)

The third line, which is not included in Hartree equations, is known as the "exchange interaction". When $k' = k$, this term equals the "self interaction correc-

tion" in the Hartree's approximation exactly. For $k' \neq k$, the exchange interaction is negative when the considered electrons are of different spins. Therefore, in Hartree-Fock equations, each electron is surrounded by an "exchange hole", a small volume where like-spin electrons are not present (Pauli correlation) [28].

The Hartree-Fock approximation is obviously better than the Hartree approximation. However, the induced exchange interaction is a non-local integral which is state-dependent. This makes the Hartree-Fock equations very difficult to solve. In addition, since the Pauli interaction is not explicit included in the Hamiltonian but is taken into account via the constraint of an "antisymmetric wavefunction", the dynamic Pauli correlation is not included in Hartree-Fock approximation. In fact, due to Coulomb repulsion, each electron will displace other electrons (both like spin and unlike spin) in its neighbourhood. The electrons do not move independently. Since Hartree-Fock theory uses a mean field approximation to describe the effective potential due to other electrons, the Coulomb correlation interactions are also missing. A direct consequence is that the Hartree-Fock approximation cannot describe the electronic screening in crystals properly. Consequently, if one uses this approximation in metals, he(or she) obtains a zero density of states at the Fermi Level. For semiconductors, the energy gap is always overestimated [35]. Thus, better approaches to determine the many-body ground state are required.

2.1.5 Density Functional Theory

In contrast to traditional electronic methods, which attempt to track all individual electrons to find solutions $\Psi\{\varphi_k(\mathbf{r}_k, \sigma_k)\}$ to the Schrödinger equation of N interacting electrons moving in an external electrostatic potential, density functional theory provides a much simpler approach, i.e. a scalar function, the density of electron charge $\rho(\mathbf{r})$ is used to describe this many-body system. In 1998, Walter Kohn was honored with the Nobel Prize in Chemistry for the development of Density Functional Theory.

The Hohenberg-Kohn theorems

The central theorem of DFT is that the problem of finding the ground state of a many-body system defined by many-body wavefunctions can be replaced by the physically equivalent problem of finding the ground state charge density of the system. The charge density is defined as

$$\rho(\mathbf{r}) = \left\langle \Psi \left| \sum_{k=1}^{N} \delta(\mathbf{r} - \mathbf{r}_k) \right| \Psi \right\rangle \tag{2.28}$$

The formal justification of this approach, which is known as the Hohenberg-Kohn theorem, was proposed by Hobenberg and Kohn in 1964 [36]. The first part of the theorem states that:

"The external potential $V_{ext}(r)$ is (to within a constant) a unique functional of

22

the ground state charge density $\rho_o(r)$; since, in turn $V_{ext}(r)$ fixes $\hat{H}^e$, we see that the full many particle ground state is a unique functional of $\rho_o(r)$ ".

For a system with non degenerate ground state, the first theorem means that there is a bijective mapping between the the ground state charge density $\rho_o(\mathbf{r})$ and the external potential $V_{ext}(\mathbf{r})$. An immediate consequence is that the ground state electron density of a many electron system can uniquely determine the external potential, which is sufficient to construct the full Hamilton operator and hence in principle to calculate the physical properties in all states of the system without the knowledge of the many electron wavefunction. Since, however, only the ground state charge density has the one-to-one correspondence with the external potential, the density of an excited state cannot be used.

With this theorem the energy of an N-electron system with a local external potential $V_{ext}(\mathbf{r})$ can be described by the Hamiltonian

$$\hat{H}^e = \hat{T}_e + \hat{V}_{e-e} + \sum_{k=1}^{N} \hat{V}_{ext}(\mathbf{r}_k). \qquad (2.29)$$

The total energy is a functional of the ground state electron density $\rho_o(\mathbf{r})$,

$$
\begin{aligned}
E[\rho_o] &= T_e[\rho_o] + V_{e-e}[\rho_o] + \int V_{ext}(\mathbf{r})\rho_o(\mathbf{r})d^3r \\
&= T_e[\rho_o] + V_{e-e}[\rho_o] + V_{ext}[\rho_o] \\
&= F[\rho_o] + V_{ext}[\rho_o] \ .
\end{aligned}
\qquad (2.30)
$$

Here, the Hoenberg-Kohn functional $F[\rho_o]$ is a new quantity given as a sum of two, system independent (universally valid for any system), parts.

$$F[\rho_o] = T_e[\rho_o] + V_{e-e}[\rho_o] = \langle \Psi | \hat{T}_e + \hat{V}_{e-e} | \Psi \rangle \qquad (2.31)$$

As can be seen, the first theorem determines the basic strategy of DFT to deal with the many-body problem : Not the many-body wavefunction is the quantity which will be determined, but the ground state charge density $\rho_o(\mathbf{r})$. Since this function depends on only three independent spatial coordinates rather than on 3N coordinates of the wave-function, it is theoretically a great simplification of the numerical task and makes DFT computationally feasible even for large systems. At this point, one has to solve the following problem: How can we be sure that a certain density is really the ground state density we are looking for? A formal prescription, how this problem should be tackled, is given by the second part of Hohenberg-Kohn theorem: the Variational Principle.

For the N-electron system with external potential $V_{ext}(r)$, the energy of an N-electron trial density, $\rho(r)$, is always greater than or equal to the true ground state energy of the system. Moreover, equality is achieved if and only if $\rho(r)$ is a ground state density for this system.

$$E_o \equiv E_{V_{ext}}[\rho_o] < E_{V_{ext}}[\rho]$$

with the constraint

$$\int \rho(\mathbf{r})d^3\mathbf{r} = N$$

The second theorem gives a minimization principle which can be used to determine the ground-state density if we calculate the total energy for every trial density. Because if we happen to use the correct charge density for the total energy calculation, then the potential calculated from this charge density will be the correct groundstate potential, and then the total energy will be a minimum. Provided the form of the total energy functional is known, the ground state density can be determined.

The Kohn-Sham Scheme

Since both the exact ground state charge density ρ_o and the Hohenberg-Kohn functional $F[\rho]$ are unknown, the problem remains now, how the Hohenberg-Kohn theorems can be used in practice to determine the actual ground state. The most common and successful applications are based on the Kohn-Sham scheme [37].

First, we can write the electron-electron interaction V_{e-e} as

$$V_{e-e}[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_k)\rho(\mathbf{r}_{k'})}{|\mathbf{r}_k - \mathbf{r}_{k'}|} d^3\mathbf{r}_k d^3\mathbf{r}_{k'} + E_{ncl}[\rho] = J[\rho] + E_{ncl}[\rho] \qquad (2.32)$$

Here $J[\rho]$ is just the classical Hartree interaction as we discussed before and $E_{ncl}[\rho]$ contains all non-classical parts. Now the complete Hohenberg-Kohn energy functional can be given as follows

$$\begin{aligned} E[\rho] &= T_e[\rho] + V_{e-e}[\rho] + V_{ext}[\rho] \\ &= \underbrace{T_e[\rho]}_{unknown} + \underbrace{J[\rho]}_{known} + \underbrace{E_{ncl}[\rho]}_{unknown} + \underbrace{E_{ext}[\rho]}_{known} \end{aligned} \qquad (2.33)$$

At this point, the basic problem is the unknown functional of the kinetic energy of interacting electrons with respect to the charge density. In 1965 Kohn and Sham gave a solution to this problem, where they formally split this functional into two parts

$$T_e[\rho] = T_s[\rho] + T_c[\rho] \qquad (2.34)$$

Here the first part T_s is the kinetic energy of a model system of N non-interacting electrons. This can be expressed in a one particle approach similar to Hartree-Fock approximation, thus is well known.

$$T_s[\rho] = -\frac{\hbar^2}{2m} \sum_{k=1}^{N} \langle \varphi_k | \nabla^2 | \varphi_k \rangle \qquad (2.35)$$

While the second part, which is still unknown, contains the difference of the real functional $T_e[\rho]$ and the one particle term $T_s[\rho]$. Now, the electronic energy can be expressed as

$$\begin{aligned} E[\rho] &= T_s[\rho] + J[\rho] + V_{ext}[\rho] + E_{ncl}[\rho] + T_c[\rho] \\ &= \underbrace{T_s[\rho]}_{known} + \underbrace{J[\rho]}_{known} + \underbrace{V_{ext}[\rho]}_{known} + \underbrace{E_{xc}[\rho]}_{unknown} \end{aligned} \qquad (2.36)$$

24

Here the $E_{xc}[\rho]$ is the so-called exchange-correlation functional, it is given by

$$E_{xc}[\rho] = E_{ncl}[\rho] + T_c[\rho] \qquad (2.37)$$

To sum the above, we finally get

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + \hat{V}_{ext}(\mathbf{r}) + \hat{V}_{Hartree}(\mathbf{r}) + \hat{V}_{xc}(\mathbf{r}) \right] \varphi_k(\mathbf{r}) = \epsilon_k \varphi_k(\mathbf{r}) \qquad (2.38)$$

This equation which is called the Kohn-Sham orbital equation, has the form of the Schrödinger equation. Here, φ_k and ϵ_k are the Kohn-Sham orbitals (wavefunction of non-interacting systems) and the Kohn-Sham eigenvalues of electronic states k, respectively. $\hat{V}_{Hartree}$ is the Hartree potential. The exchange-correlation potential, $\hat{V}_{xc}$, is given by the functional derivative

$$V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} \qquad (2.39)$$

The central assertion of Kohn-Sham Scheme is that for any interacting system, one can find a non-interacting model system, with some local effective potential $V_{eff}(\mathbf{r})$, that has the ground-state density which exactly equals the ground-state density of the interacting system. The difference in the kinetic energy between the interacting electron system and the model non-interacting system, as well as the many-particle effects, the exchange and correlation energy contribution are taken into account in the term $E_{xc}(\mathbf{r})$.

The minimization process corresponds to the determination of the solution of Kohn-Sham equation, which will again lead to the Lagrange multipliers ϵ_k. Since the potentials depend on the electron charge density, which in turn depends on the one-particle wavefunction φ_k, the Kohn-Sham equations are solved self-consistently (see Fig. 2.1).

Solving the Kohn-Sham equations

The numerical solution of Kohn-Sham orbital equations needs a reasonable description of the one-particle wavefunction φ_k. A mathematically important simplification often used in modelling of macroscopic systems are periodic boundary condition. In this context, a plane-wave basis set is a natural choice because any periodic function can be expressed as a sum of plane waves. The single-particle wavefuntions can then be expressed by

$$\varphi_k = \sum_{p=1}^{P} c_p^k \phi_p^{basis} \qquad (2.40)$$

where P is the total number of plane waves which are used to represent the kth Kohn-Sham orbital φ_k, ϕ_p^{basis} denotes the pth plane wave in the basis, and c_p^k is the coefficient of ϕ_p^{basis} in the Fourier expansion. Substituting equation (2.40)

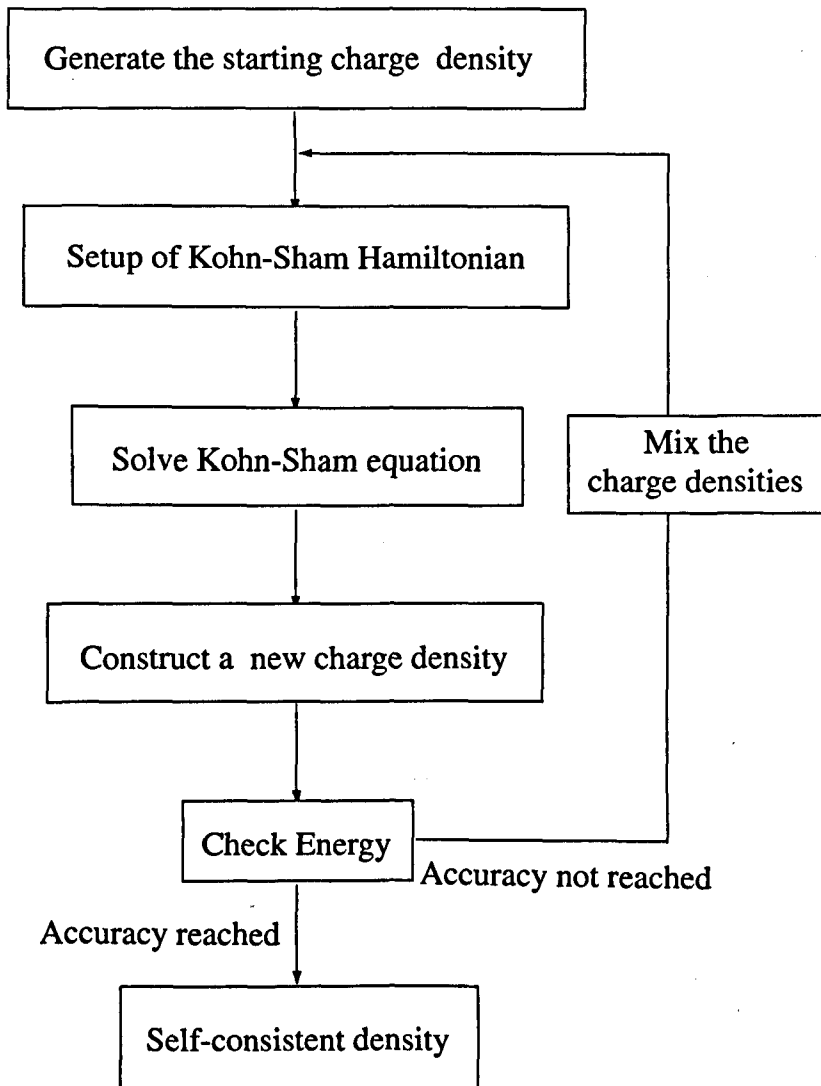Figure 2.1: Schematic representation of self-consistent field method

into equation (2.38), and from the left multiply with $\langle\phi_i^{basis}|$ $(i = 1, 2, ..., P)$, we obtain

$$
\begin{bmatrix} \cdots & \cdots & \cdots \\ \vdots & \langle\phi_i^{basis}|\hat{H}|\phi_j^{basis}\rangle - \epsilon_k\langle\phi_i^{basis}|\phi_j^{basis}\rangle & \vdots \\ \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} c_1^k \\ \vdots \\ c_P^k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}
$$

The eigenvalues and eigenfunctions can then be solved by common matrix operations. Each eigenvalue $\epsilon^k(\mathbf{k})$ corresponds to an eigenfunction $\varphi^k(\mathbf{k}, \mathbf{r})$, which is constructed by P plane waves. An exact representation of Kohn-Sham orbitals φ_k will need an infinite number of basis functions ($P \to \infty$). In practice, however, only a limited number of plane waves are used to generate a function that is "close" to φ_k. For each eigenfunction $\varphi^k(\mathbf{k}, \mathbf{r})$, we have

$$
\varphi^k(\mathbf{k}, \mathbf{r}) = \sum_{G=0}^{G_{max}} c_{G,k}^k e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} \tag{2.41}
$$

Here, $\mathbf{k}$ is the wave vector in the first Brillouin zone and $\mathbf{G}$ is the reciprocal lattice vector, which is a linear combination of the primitive reciprocal vectors. The plane-wave basis is $\mathbf{k}$ dependent, i.e. the same set of P plane wave basis can be used to represent eigenfunctions with the same $\mathbf{k}$ but a different band index k. However, for the eigenfunctions at a different $\mathbf{k}$ point, a new basis needs to be used. This process should be repeated for an increasing number of $\mathbf{k}$ points until the first Brillouin zone is sampled densely enough. The parameter G_{max} is used to truncate the number of plane waves in the basis. Increasing the number of G vectors in (2.41) plays the same role as increasing P in (2.40). In current DFT calculations, this truncation is often expressed by the "cut-off energy", E_{cut}:

$$
E_{cut} = \frac{\hbar^2 G_{max}^2}{2m} \tag{2.42}
$$

Pseudopotentials

In spite of the simple mathematical picture of a plane-wave basis, this method has turned out to be entirely inapplicable to real atomic potentials. The reason for this failure lies in the strong attractive potential of the core region of an ion. In this region, core electrons are tightly bound to the nucleus and therefore possess high kinetic energy (a high cut-off energy is required). The valence electron wavefunctions, on the other hand, in order to be orthogonal to the core electron wavefunctions (as required by the exclusion principle) must oscillate rapidly in the core region. A correct description of all electronic states would therefore need an enormously large number of plane waves. However, in most calculations, we are only interested in valence electronic states, because the core electronic states do not contribute much to the interatomic interactions, but only play the role of screening the original ionic potential so that a valence electron does not feel a full but only a screened ionic potential. An effective method to simplify the electronic calculation is therefore to replace the electronic states in the core region by a

"(relativistic) effective core potential", which is also known as "pseudo potential". In this method, the core electrons are calculated only for a reference configuration (usually a sphere with a given radius r_c) and are kept fixed thereafter. The pseudo wavefunctions are constructed to be identical to the real all electron wavefunctions outside the core region and be node-less inside the core region. Such kind of pseudopotentials dramatically reduce the number of plane waves in the basis and directly yield the valence electronics states (but with a modified description of the attractive core region). In this work, all DFT calculations were performed using ultrasoft pseudopotentials.

Physical meaning of the Kohn-Sham orbitals

Although it is customary to regard the bonding between atoms and molecules in terms of the interactions among electronic orbitals, there are no such orbitals inherent to the many-particle system where the electron eigenstates are spread out and commonly overlap with more than just a few particles. Since the connection between the interacting and non-interacting systems is only through the constraint that they have the same charge density, the Kohn-Sham orbitals $\varphi_k(\mathbf{r}_k, \sigma_k)$ cannot be interpreted as the orbitals of the many-electron system. Likewise, one should not expect any straight interpretation of the Kohn-Sham orbital energies ϵ_k either. However, as an exception, the eigenvalue of the highest occupied Kohn-Sham orbital is equal to the negative of the ionization energy of the many-body system [38].

The Exchange-Corrlation functionals

Up to this point, the framework of DFT was developed under the assumption that the exchange-correlation energy functional is known. However, its explicit expression is not known. Thus, further approximations have to be made. The first attempt of finding an expression of $E_{xc}[\rho(\mathbf{r})]$ was based on a model system of a homogeneous interacting electron gas, where $V_{ext}(\mathbf{r})$ and $\rho(\mathbf{r})$ are constants. The exchange-correlation functional can be expressed as

$$E_{xc}[\rho] = \int \epsilon_{xc}[\rho]\rho(\mathbf{r})d^3\mathbf{r}, \qquad (2.43)$$

where $\epsilon_{xc}[\rho]$ represent the exchange-correlation energy per particle of the uniform electron gas. Since $\rho(\mathbf{r})$ is just a constant, $\epsilon_{xc}[\rho]$ becomes a function of the local charge density $\rho(\mathbf{r})$. This approximation is called Local-Density Approximation (LDA). It means that the many-body system can be divided into infinitesimally small volumes, each of which contributes the same exchange-correlation energy as the same volume of homogeneous electron gas.

$$E_{xc}^{LDA}[\rho] = \int \epsilon_{xc}^{LDA}(\rho(\mathbf{r}))\rho(\mathbf{r})d^3\mathbf{r} \qquad (2.44)$$

The expression of the function $\epsilon_{xc}^{LDA}(\rho(\mathbf{r}))$ can be further split into exchange and correlation contributions.

$$\epsilon_{xc}^{LDA}(\rho(\mathbf{r})) = \epsilon_x^{LDA}(\rho(\mathbf{r})) + \epsilon_c^{LDA}(\rho(\mathbf{r})) \qquad (2.45)$$

The exchange part can be calculated by using the Hartree-Fock approximation (exchange interactions of an electron in a homogeneous electron gas) and has the form

$$\epsilon_x^{LDA}(\rho(\mathbf{r})) = -\frac{3}{4}\sqrt[3]{\frac{3\rho(\mathbf{r})}{\pi}} \qquad (2.46)$$

The correlation part $\epsilon_c^{LDA}(\rho(\mathbf{r}))$, however, is only numerically known from quantum Monte Carlo (QMC) simulations done by Ceperly and Alder in 1980 [39]. The obtained results can then be parameterized for DFT calculations, the two most widely used reresentations today were developed by Vosko, Wilk and Nusair in 1980 [40] and Perdew and Wang in 1992 [41].

Although LDA is a very drastic approximation for actual systems, it often gives surprisingly good results which are comparable or even better than the Hartree-Fock aprroximation. The reason is that the spherical average of wrong qualitative description of the exchange-correlation hole in LDA is very similar to the spherical average of the real exchange-correlation hole [28]. A typical problem of using LDA, however, is that the binding energy is usually overestimated, as a result, the bond lengths and equilibrium volume are underestimated.

An improvement in the description of the correlation energy is the Generalized Gradient Approximation (GGA), which introduces a function of the electron spin-densities $\rho_\uparrow(\mathbf{r}), \rho_\downarrow(\mathbf{r})$ and their gradients $\nabla\rho_\uparrow(\mathbf{r}), \nabla\rho_\downarrow(\mathbf{r})$. The exchange-correlation energy functional in GGA is expressed as

$$E_{xc}^{GGA} = \int f\left(\rho_\uparrow(\mathbf{r}), \rho_\downarrow(\mathbf{r}), \nabla\rho_\uparrow(\mathbf{r}), \nabla\rho_\downarrow(\mathbf{r})\right) d^3\mathbf{r} \qquad (2.47)$$

In addition to the local charge density $\rho(\mathbf{r})$, the semi-local function also takes into account the electronic charge density in a direct neighborhood of the position $\mathbf{r}$ and therefore results in better binding energy comparing with LDA. For this reason, most current DFT calculations use GGA rather than LDA.

In contrast with the unique definition for ϵ_{xc} in LDA, there exists more than one options to incorporate the charge density gradient. As a consequence, there are more than one GGA functionals. The one used in this work is the PW91 functional, which developed by Perdew and Wang [42].

2.2 STM simulations

As we discussed in the first section, the first principles calculations of the electronic properties of a single metal or semiconductor surface based on DFT methods often require vast amounts of computer time. Today, the straightforward calculation of the electronic properties and the conductance of a tunneling junction consisting of the two leads and an intermediate region is still only a remote possibility. This means that the effects of all calculations of the tunneling current are limited by underlying assumptions about the interaction between the scanning tip

and the sample surface. Over the last decades, quite a few models have been put forward to calculate the tunneling current[43]. In increasing order of theoretical difficulty, the four main approaches are: (i)the Tersoff-Hamann approach[44, 45], (ii)the transfer Hamiltonian or the Bardeen approach[46], (iii)the scattering or Landauer-Bütticker approach[47], and (iv) the Keldysh or non-equilibrium Greens function approach[48]. In this section, we will give an detailed description of the Tersoff-Hamann approach (perturbation approach) and Julian Chen's modified Bardeen's approach. Further information can be found in the review papers [49, 50].

2.2.1   Tersoff-Hamann Approach
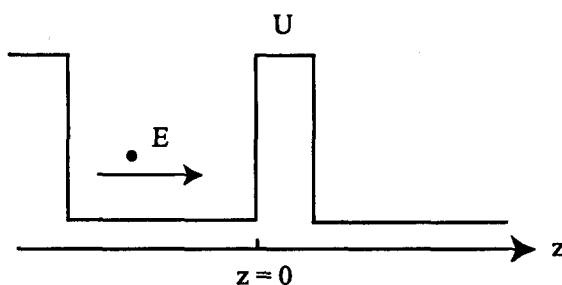
Figure 2.2: Schematic representation of the one-dimension tunneling in a square potential well

Considering the one-dimensional case, an electron with energy E moving in a finite square well potential $U(z)$ (see Fig. 2.2), the Schrödinger equation reads:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dz^2}\psi(z) + U(z)\psi(z) = E\psi(z).$$
(2.48)

In the classically allowed region, $E > U(z)$, the solutions are plane waves

$$\psi(z) = \psi_0(z)e^{ikz}$$
(2.49)

where

$$k = \frac{\sqrt{2m(E-U)}}{\hbar}$$
(2.50)

In the classical forbidden region, $E < U(z)$, the solutions are

$$\psi(z) = \psi_0(z)e^{-\kappa z}$$
(2.51)

with

$$\kappa = \frac{\sqrt{2m(U-E)}}{\hbar}$$
(2.52)

As can be seen, the electron states decay exponentially in z-direction. Consequently, there is always a non-vanishing probability (proportional to $|\psi_0(z)|^2 e^{-2\kappa z}$)

of finding the electron in the barrier region z. Assuming that the surface and tip have the same work function ϕ (the minimum energy required to take an electron from bulk to the vacuum) and only elastic tunneling is taken into account, the probability w of a surface electron with the energy of E_{S_n} to present in the tip state which have the same energy as E_{S_n} is

$$w \propto |\psi_{S_n}(0)|^2 e^{-2\kappa z} \qquad (2.53)$$

where z is the tip-surface distance, $\psi_{S_n}(0)$ is the nth surface state and

$$\kappa = \frac{\sqrt{2m\phi_{S_n}}}{\hbar} \qquad (2.54)$$

However, there will not be net current between the two leads until a bias voltage V_{bia} is provided. The higher V_{bia} is, the more surface states come into play. Suppose the applied bias voltage is very small so that it will not significantly change the surface states, the tunneling current should be proportional to the number of surface states within the energy interval eV_{bia}.

$$I(z) \propto \sum_{E_{S_n} > E_F - eV_{bia}}^{E_{S_n} < E_F} |\psi(z, E_{S_n})|^2 \qquad (2.55)$$

where, E_F denotes to the Fermi level of the surface states and E_n is the eigenstates of the surface electrons. The obtained tunneling current map is nothing but the Local Density of States (LDOS) of the surface around the Fermi Level.

Due to its simple mathematic expression, the Tersoff-Hamann method has been incorporated into nearly all DFT codes. Nevertheless, in realistic cases, the work functions of the tips and surfaces are usually different, and the decay of the electrons states is not a constant. This method has been analyzed to be reliable only when the surface-tip distance is large ($> 5 - 6$Å) and when the feature size is well above the typical length scale of electron states (about $1 - 2$Å) of an STM tip [43]. In addition, since the Tersoff-Hamann method is based the electronic structure of the sample surface alone (contributions form an STM tip are not included), it fails in cases where the properties of STM tips play an important role in imaging mechanisms.

2.2.2 Bardeen Approach

In Bardeen's tunneling theory, the tip and the sample surface are assumed to be sufficient far apart. When the bias voltage is low, the tunneling current can be regarded as the net effect of independent scattering events that transfer electrons across the tunneling barrier. Thus, instead of treating the combined surface-tip system as one quantum mechanical system, the approach takes the electronic structures of the two subsystems separately and then calculates the electron transfer between the tip and surface using time dependent perturbation theory. The method is based on three further assumptions: (i) the surface-tip interactions are

very weak, (ii) the electron-electron interactions can be neglected, (iii) occupation probabilities of the surface and the STM tip are independent from each other and do not change during tunneling. The total current is expressed by the following equation.

$$I = \frac{4\pi e}{\hbar} \int_{-\infty}^{\infty} [f(E_F - eV + \epsilon) - f(E_F + \epsilon)] \rho_S(E_F - eV + \epsilon) \rho_T(E_F + \epsilon) |M|^2 d\epsilon \quad (2.56)$$

where ϵ is denoted to the value of bias voltage applied in an STM scanning and the function $f(E, T)$ is the Fermi distribution function

$$f(E, T) = \frac{1}{e^{[(E - E_F)/k_B T]} + 1} \quad (2.57)$$

and M denotes the tunneling matrix element, which is determined by the overlap of the surface wavefunction (ψ) and the tip wavefunction(χ) at a separation surface ($z = z_0$).

$$M = \frac{\hbar}{2m} \int_{z=z_0} \left(\chi^* \frac{\partial \psi}{\partial z} - \psi \frac{\partial \chi}{\partial z} \right) dS \quad (2.58)$$

The original derivation of the equations can be found in [46]. As can be seen, in Bardeen's approach not only the tip wavefunction but also the working temperature and the bias voltage have to be taken into consideration. Thus, this model enables a quantitative prediction of tunneling current according to the real experimental conditions. In practice, however, when an STM tip comes close to a sample surface, the wavefunctions of the two subsystems are no longer those of the isolated tip and surface systems, the interactions between the two leads then change the tunneling mechanism substantially[15, 51, 52].

2.2.3   Modified Bardeen Approach

In STM operation, the tunneling of an electron from one electrode to the other is highly influenced by the interactions (such as van der Waals interactions, electrostatic interactions, chemical interactions and magnetic interactions) between an STM tip and a sample surface. The main challenge to a theoretical description of the true tunneling process is that the order of the importance of these effects depends on the tip-surface distance and therefore on the tunneling conditions. The essence of the Modified Bardeen Approach (MBA) is to account for the long distance (10 ~ 100Å) interactions, e.g. van der Waals force, with the stationary-state perturbation theory and treat the short distance (3 ~ 10Å) interactions, e.g. resonance interaction, by using the time-dependent perturbation theory [53]. A brief introduction of time-dependent perturbation theory and the explicit derivation of MBA is given below.

The Schrödinger picture

Due to the interactions between the tip and a surface, the physical properties of a subsystem will typically change over time, the appropriate vectors and/or operators may also change over time. There are three distinct, equivalent, standard

approaches to compute the time dependency of these quantities, each of them is referred to as a different "picture".

In the Schrödinger picture, the time dependency of the actual physical systems is carried by the state vectors themselves, the energy operator $\hat{H}$ does not depend on time.

$$\hat{H} \equiv -\frac{\hbar^2}{2m}\nabla^2 + \hat{V}(\mathbf{r}) \tag{2.59}$$

Then the time-dependent wavefunction $\Psi(\mathbf{r}, t)$ develops according to

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = \hat{H}\Psi(\mathbf{r}, t) \tag{2.60}$$

The time-evolution of a state $\Psi(\mathbf{r}, t)$ can usually be expressed by a time-evolution operator $\hat{U}$, which satisfies

$$\Psi(\mathbf{r}, t) = \hat{U}(t - 0)\Psi(\mathbf{r}, 0) \tag{2.61}$$

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = i\hbar\frac{\partial}{\partial t}\hat{U}(t - 0)\Psi(\mathbf{r}, 0) = \hat{H}\hat{U}(t - 0)\Psi(\mathbf{r}, 0)$$

$$\Rightarrow i\hbar\frac{\partial}{\partial t}\hat{U}(t - 0) = \hat{H}\hat{U}(t - 0) \Rightarrow \hat{U}(t - 0) = e^{-\frac{i}{\hbar}\hat{H}t}$$

Thus we get

$$\Psi(\mathbf{r}, t) = e^{-\frac{i}{\hbar}\hat{H}t}\Psi(\mathbf{r}, 0) \tag{2.62}$$

The change with time in this case is part of the wavefunctions.

Heisenberg picture

The Heisenberg picture can be derived from the Schrödinger picture, if we put all the time-dependency into the linear operators. The wavefunction, on the other hand, maintains exactly the same at whatever time. From (2.62) we have

$$\Psi(\mathbf{r}, 0) = \Psi(\mathbf{r}, t)e^{\frac{i}{\hbar}\hat{H}t} \tag{2.63}$$

$\Psi(\mathbf{r}, 0)$ does not depend on time, because

$$\begin{aligned}
\frac{\partial}{\partial t}\Psi(\mathbf{r}, 0) &= \frac{i}{\hbar}\hat{H}e^{i\hat{H}t/\hbar}\Psi(\mathbf{r}, t) + e^{i\hat{H}t/\hbar}\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \\
&= \frac{i}{\hbar}\hat{H}e^{i\hat{H}t/\hbar}\Psi(\mathbf{r}, t) + e^{i\hat{H}t/\hbar}[-\frac{i}{\hbar}\hat{H}\Psi(\mathbf{r}, t)] \\
&= \frac{i}{\hbar}\hat{H}e^{i\hat{H}t/\hbar}\Psi(\mathbf{r}, t) - \frac{i}{\hbar}\hat{H}e^{i\hat{H}t/\hbar}\Psi(\mathbf{r}, t) \\
&= 0
\end{aligned}$$

The expectation value of an operator $\hat{A}$ in the state $\Psi(\mathbf{r}, t)$ reads

$$\langle\Psi(\mathbf{r}, t)|\hat{A}|\Psi(\mathbf{r}, t)\rangle = \langle\Psi(\mathbf{r}, 0)|e^{i\hat{H}t/\hbar}\hat{A}e^{-i\hat{H}t/\hbar}|\Psi(\mathbf{r}, 0)\rangle \tag{2.64}$$

The time dependent operator is defined as

$$\hat{A}_H = \hat{A}_H(t) = e^{i\hat{H}t/\hbar}\hat{A}e^{-i\hat{H}t/\hbar} \tag{2.65}$$

Interaction Picture

The interaction picture is a combination of the last two pictures for the case when the Hamiltonian operator can be suitably divided into two (or possibly more) parts, one of which is well-understood. For instance, sometimes we have a situation, that the energy operator is composed of a time independent part H_o and a time dependent potential $V(t)$

$$\hat{H} \equiv \hat{H}_o + \hat{V}(t)$$

In this case, we can use the interaction picture, because it separates the ground-state, described by $\hat{H}_o$, from the interaction, described by $\hat{V}(t)$.

The wavefunction in the interaction picture is defined by

$$\Psi_\tau = e^{i\hat{H}_o t/\hbar} \Psi(\mathbf{r}, t)$$

The operator A(t) in the the interaction picture is given by

$$A_\tau(t) = e^{i\hat{H}_o t/\hbar} A(t) e^{-i\hat{H}_o t/\hbar}$$

The changes of the wavefunction in the interaction picture depend only on V_τ. We have

$$
\begin{aligned}
\frac{\partial}{\partial t}\Psi_\tau(\mathbf{r}, t) &= \hat{H}_o \frac{i}{\hbar} e^{i\hat{H}_o t/\hbar} \Psi(\mathbf{r}, t) + e^{i\hat{H}_o t/\hbar} \frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \\
&= \hat{H}_o \frac{i}{\hbar} e^{i\hat{H}_o t/\hbar} \Psi(\mathbf{r}, t) + e^{i\hat{H}_o t/\hbar} \frac{-i}{\hbar}(\hat{H}_o + \hat{V}(t))\Psi(\mathbf{r}, t) \\
&= -\frac{i}{\hbar}[e^{i\hat{H}_o t/\hbar} \hat{V}(t) e^{-i\hat{H}_o t/\hbar}][e^{i\hat{H}_o t/\hbar} \Psi(\mathbf{r}, t)] \\
&= -\frac{i}{\hbar}\hat{V}_\tau(t)\Psi_\tau(\mathbf{r}, t)
\end{aligned}
$$

$$\Rightarrow i\hbar \frac{\partial}{\partial t}\Psi_\tau(\mathbf{r}, t) = \hat{V}_\tau(t)\Psi_\tau(\mathbf{r}, t)$$

There is also an equation of motion for the operator $\hat{V}_\tau(t)$

$$\frac{d}{dt}\hat{V}_\tau(t) = \frac{i}{\hbar}[\hat{H}_o, \hat{V}_\tau(t)] + \frac{\partial}{\partial t}\hat{V}_\tau(t)$$

These three pictures allow us to determine the evolution of a system once the ground state is known.

Time dependent perturbation theory

Since we know that

$$i\hbar \frac{\partial}{\partial t}\Psi_\tau(\mathbf{r}, t) = \hat{V}(t)\Psi_\tau(\mathbf{r}, t)$$

We can easily integrate from t_o to t and get the integral equation

$$i\hbar[\Psi_\tau(\mathbf{r}, t) - \Psi_\tau(\mathbf{r}, t_o)] = \int_{t_o}^{t} dt' \hat{V}_\tau(t')\Psi_\tau(\mathbf{r}, t')$$

$$\Psi_\tau(\mathbf{r}, t) = \Psi_\tau(\mathbf{r}, t_o) + \frac{1}{i\hbar} \int_{t_o}^{t} dt' \hat{V}_\tau(t') \Psi_\tau(\mathbf{r}, t')$$

Note that the function $\Psi_\tau(\mathbf{r}, t)$ occurs twice : once on the left and then in the integral. We can use this equation successively, such a succession is called a **perturbation series**.

Zero-order: $\Psi_\tau(\mathbf{r}, t)$ in the integral is zero

$$\Psi_\tau^0(\mathbf{r}, t) = \Psi_\tau(\mathbf{r}, t_o)$$

First-order: $\Psi_\tau(\mathbf{r}, t)$ in the integral is $\Psi_\tau(\mathbf{r}, t_o)$

$$\Psi_\tau^1(\mathbf{r}, t) = \Psi_\tau(\mathbf{r}, t_o) - \frac{i}{\hbar} \int_{t_o}^{t} dt' \hat{V}_\tau(t') \Psi_\tau(\mathbf{r}, t_o)$$

Second-order: $\Psi_\tau(\mathbf{r}, t)$ in the integral is $\Psi_\tau^1(\mathbf{r}, t)$

$$\Psi_\tau^2(\mathbf{r}, t) = \Psi_\tau(\mathbf{r}, t_o) - \frac{i}{\hbar} \int_{t_o}^{t} dt' \hat{V}_\tau(t') \Psi_\tau(\mathbf{r}, t_o) - \frac{1}{\hbar^2} \int_{t_o}^{t} dt' \hat{V}_\tau(t') \int_{t_o}^{t'} \Psi_\tau dt'' \hat{V}_\tau(t'') \Psi_\tau(\mathbf{r}, t_o)$$

We could use this solution to create the solution of $\Psi_\tau^3(\mathbf{r}, t)$, which would be the fourth-order, and so on, in principle till we arrive at a constant solution. This implies that our wavefunction $\Psi_\tau^n(\mathbf{r}, t)$ does not change from the approximation $\Psi_\tau^{n-1}(\mathbf{r}, t)$

In our model, we generally limit the series to the first order perturbation expansion. Then, our solution is given by

$$\Psi_\tau^1(\mathbf{r}, t) = \Psi_\tau(\mathbf{r}, t_o) - \frac{i}{\hbar} \int_{t_o}^{t} dt' \hat{V}_\tau(t') \Psi_\tau(\mathbf{r}, t_o)$$

Transition to the first order

Now let us assume that we have one electron in a state ψ_m at t=0, the interaction picture of this state is then

$$\psi_{\tau m}(\mathbf{r}, t) = e^{i\hat{H}_o t/\hbar} \psi_m(\mathbf{r}, t)$$

$$\psi_m(\mathbf{r}, t) = e^{-i\hat{H}_o t/\hbar} \psi_m(\mathbf{r})$$

$$\psi_{\tau m}(\mathbf{r}, t) = \psi_m(\mathbf{r})$$

Next we want to calculate the wavefunction ψ_m at a later time, using our first order relation:

$$\psi_{\tau m}^1(\mathbf{r}, t) = \psi_{\tau m}(\mathbf{r}) - \frac{i}{\hbar} \int_{t_o}^{t} dt' \hat{V}_\tau(t) \psi_{\tau m}(\mathbf{r})$$

To get the transition probability from state ψ_m to another state ψ_n, we need to calculate the overlap of these two states:

$$\left| \int \psi_{\tau n}^*(\mathbf{r})\psi_{\tau m}(\mathbf{r},t)d^3r \right|^2 = P_{m\to n}(t)$$

This involves:

$$\int d^3r\psi_{\tau n}^*(\mathbf{r})\psi_{\tau m}(\mathbf{r},t) = \int d^3r\psi_{\tau n}^*(\mathbf{r})\psi_{\tau m}(\mathbf{r}) - \frac{i}{\hbar}\int_{t_o}^t dt' \int \psi_{\tau n}^*(\mathbf{r})e^{iH_ot/\hbar}\hat{V}(t')e^{-iH_ot/\hbar}\psi_{\tau m}(\mathbf{r})$$

$$= \delta_{nm} - \frac{i}{\hbar}\int_{t_o}^t dt' \int d^3r \left(\psi_n^*(\mathbf{r})\hat{V}\psi_m(\mathbf{r}) \right)e^{i(\omega_n-\omega_m)t'}$$

Here,

$$\omega_n - \omega_m = \frac{<\hat{H}_o>_n - <\hat{H}_o>_m}{\hbar}$$

And for the probability if n$\neq$m, we therefore get

$$P_{m\to n}(t) = \left| \frac{i}{\hbar}\int_0^t dt' e^{i(\omega_n-\omega_m)t'}(\psi_n^*\hat{V}(t')\psi_m) \right|^2$$

$$= \frac{1}{\hbar^2}f(\frac{\omega_{nm}}{2})\left| (\psi_n^*\hat{V}(t')\psi_m) \right|^2 \qquad (2.66)$$

where

$$f(\frac{\omega_{nm}}{2}) = \left[\frac{sin(\frac{\omega_{nm}t}{2})}{\frac{\omega_{nm}t}{2}} \right]^2$$

As known, the maximum value of $f(\frac{\omega_{nm}}{2})$ is t^2 and this function vanishes for

$$\omega_{nm}t = 2n\pi \quad n = 1,2,3...$$

If t is large enough, we can express $f(\frac{\omega_{nm}}{2})$ by a delta-functional

$$\delta(\frac{\omega_{nm}}{2})\pi = \frac{f(\frac{\omega_{nm}}{2})}{t} = \frac{sin^2(\frac{\omega_{nm}t}{2})}{(\frac{\omega_{nm}}{2})^2t} \qquad (2.67)$$

Substituting (2.67) into (2.66), the time transition velocity or the number of transitions per second is:

$$\frac{P_{n\to m}(t)}{t} = \frac{2\pi}{\hbar}\delta(E_n - E_m)\left| (\psi_n^*\hat{V}(t)\psi_m) \right|^2 \qquad (2.68)$$

Julian Chen's Theory

In the Modified Bardeen Approach, developed by C. Julian Chen [53], the surface potential is given by U_S and the tip potential is given by U_T. The potentials of the two components are required to fulfill the following two conditions.

$$(1) \quad U_S + U_T = U \qquad (2) \quad U_S U_T = 0$$

For $t < 0$, the tip potential U_T is switched off, and the Schrödinger Equation for the sample system is

$$(-\frac{\hbar^2}{2m}\nabla^2 + U_S)\psi_u = E_u\psi_u \qquad (2.69)$$

For $t \geq 0$, the tip potential U_T is switched on and the sample system starts to evolve according to the time dependent Schrödinger Equation

$$(-\frac{\hbar^2}{2m}\nabla^2 + U_T + U_S)\psi = i\hbar\frac{\partial\psi}{\partial t} \qquad (2.70)$$

The tip system is also described by a Schrödinger Equation

$$(-\frac{\hbar^2}{2m}\nabla^2 + U_T)\psi_\nu = E_\nu\psi_\nu \qquad (2.71)$$

The key point in Chen's derivation is to develop the wavefunction ψ in terms of the unperturbed tip states ψ_ν

$$\psi = \sum_\nu a_\nu(t)\psi_\nu e^{-iE_\nu t/\hbar} \qquad (2.72)$$

$$a_\nu(t) = (\psi_\nu, \psi_\mu)e^{-i(E_\mu - E_\nu)t/\hbar} + C_\nu(t) \qquad (2.73)$$

where for $t = 0$ we assume $C_\nu(0) = 0$. Now, we substitute equation (2.73) into equation (2.72)

$$\begin{aligned} \psi &= \sum_\nu[(\psi_\nu, \psi_\mu)e^{-i(E_\mu - E_\nu)t/\hbar} + C_\nu(t)]\psi_\nu e^{-iE_\nu t/\hbar} \\ &= \psi_\mu e^{-iE_\mu t/\hbar} + \sum_\nu C_\nu(t)\psi_\nu e^{-E_\nu t/\hbar} \end{aligned} \qquad (2.74)$$

We can see from equation (2.74) that the wavefunction ψ is now a linear combination of the original state ψ_μ and all the tip states ψ_ν. All the changes of our system with time are included in the term $C_\nu(t)$. The time-dependent coefficient $C_\nu(t)$ can be evaluated with:

$$C_\nu(t) = \frac{1}{i\hbar}\int_0^t e^{i(E_\nu - E_\mu)t'/\hbar}(\psi_\nu U_T\psi_\mu)dt' \qquad (2.75)$$

And the transition probability per unit time is

$$P(t) = \frac{2\pi}{\hbar}\delta(E_\mu - E_\nu)|(\psi_\nu U_T\psi_\mu)|^2 \qquad (2.76)$$

Calculation of the current

According to the Schrödinger Equation, the decisive matrix element $M_{\mu\nu}$ for the transition from state ψ_μ of the surface to ψ_ν of the tip can be rewritten to a Bardeen like form. The integral encompasses only the region of the tip, because the potential U_T is zero outside.

$$M_{\mu\nu} = (\psi_\nu U_T\psi_\mu) = \int_{\Omega_T} dt\psi_\mu U_T\psi_\nu^* = \int_{\Omega_T} dt\psi_\mu\left(E_\nu + \frac{\hbar^2}{2m}\nabla^2\right)\psi_\nu^* \qquad (2.77)$$

Due to the delta functional, E_μ and E_ν must be equal, and therefore, the matrix element can also be written as

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int_{\Omega_T} dt(\psi_\mu \nabla^2 \psi_\nu^* - \psi_\nu^* \nabla^2 \psi_\mu) \qquad (2.78)$$

Note that the surface potential is zero in the region of integration.

With the help of Gauss's theorem, the integral can be transformed into a surface integral over the separation surface S, while the operator of the kinetic energy becomes a gradient:

$$M_{\mu\nu} = -\frac{\hbar^2}{2m} \int_S d\vec{S}(\psi_\nu^* \nabla \psi_\mu - \psi_\mu \nabla \psi_\nu^*) \qquad (2.79)$$

According to Chen's theory, the matrix element has the dimension of energy. From Chen's point of view, its physical meaning is the lowering of energy due to the interplay of the two states. Integrating over all the states of the tip and the sample, taking the occupation probabilities into account, the tunneling current is

$$I = \frac{4\pi e}{\hbar} \int_{-\infty}^{+\infty} d\varepsilon[f(E_F - eV + \varepsilon) - f(E_F + \varepsilon)] \times \sum_{\mu,\nu} |M_{\mu\nu}|^2$$
$$\times \delta(\varepsilon - \varepsilon_\mu)\delta(\varepsilon - \varepsilon_\nu) \qquad (2.80)$$

Where ε is the eigen values of the tip and surface in the vacuum, $f(E) = [1 + exp(E - E_F)/k_B T]^{-1}$ is the Fermi distribution, $\rho_S(E_F)$ is the density of state (DOS) of the sample at the fermi level, and $\rho_T(E_F)$ is DOS of the tip at the fermi level.

Numerical Implementation

The wavefunction for the surface is

$$\psi_\mu = \sum_{\vec{G_S}} A_S(\vec{G_S})e^{-K(\vec{G_S})z}e^{i(\vec{K}+\vec{G_S})\vec{r}_\parallel} \qquad (2.81)$$

The wavefunction for the tip is

$$\psi_\nu = \sum_{\vec{G_T}} A_T(\vec{G_T})e^{K(\vec{G_T})z}e^{i(\vec{K}+\vec{G_T})\vec{r}_\parallel} \qquad (2.82)$$

Here A is the amplitude, $e^{-K(\vec{G_S})z}$ and $e^{K(\vec{G_T})z}$ describe the decay of the surface and tip states into the vacuum and $e^{i(\vec{K}+\vec{G})\vec{r}_\parallel}$ is a two dimensional planewave. K is defined by taking into account the kinetic energy associated with the given reciprocal lattice vector, $U - \frac{\hbar^2}{2m}K(\vec{G})^2 + \frac{\hbar^2}{2m}(\vec{K} + \vec{G})^2 = E$ and U is the barrier height.

$$\psi_\mu \nabla \psi_\nu^* = \psi_\mu \frac{\partial}{\partial z}\psi_\nu^*$$
$$= \sum_{\vec{G_S}\vec{G_T}} A_S(\vec{G_S})A_T^*(\vec{G_T})e^{i(\vec{K}+\vec{G_S})\vec{r}_\parallel}e^{-i(\vec{K}+\vec{G_T})\vec{r}_\parallel}e^{-K(\vec{G_S})z}K(\vec{G_T})e^{K(\vec{G_T})z} \qquad (2.83)$$

38

$$\int_S \psi_\mu \nabla \psi_\nu^* = \sum_{\vec{G}_S \vec{G}_T} A_S(\vec{G}_S) A_T^*(\vec{G}_T) K(\vec{G}_T) e^{K(\vec{G}_T)z} e^{-K(\vec{G}_S)z} \int d\vec{S}\, e^{i(\vec{G}_S - \vec{G}_T)\vec{r}_\parallel}$$

$$(2.84)$$

$$\int d\vec{S}\, e^{i(\vec{G}_S - \vec{G}_T)\vec{r}_\parallel} = \int dx e^{i(G_S - G_T)x} dy e^{i(G_S - G_T)y} \qquad (2.85)$$

This integration can be performed analytically. For our calculations we used the BSKAN code, which is described in [43].

2.2.4 Applications of the first-principles STM simulations

Interpret the STM images

It is well known that an STM image is a map of electronic states of a surface. Its formation can be affected by many factors: e.g. the tunneling current, the bias voltage, the configurations of an STM tip and also the interactions between the tip and the surface. As a result, the interpretation of the microscopic surface structures of a given STM image has been regarded as one of the most delicate problems in surface science. The main applications of the first-principle STM simulations are therefore to interpret the images mapped in STM measurements.

In principle, with the theory introduced above, one can perform STM simulations for all possible configurations under real experimental conditions: the same tip material, bias voltage and current value. The obtained images can then be compared with the constant current map produced from the STM experiment. The configuration, which is consistent with the experimental result, is very likely the surface structure of we are looking for. Although this strategy is easy to understand, the number of possible structures for a surface system are usually too large to be calculated. In addition, since the first-principles STM simulations are as time consuming as DFT calculations, the strategy used in practice was the following.

- Step 1: Based on other experimental results, e.g. LEED, propose some possible structures of the adsorbate-substrate system.

- Step 2: Perform the first-principle calculations to evaluate the low temperature adsorption energy per adsorbate of those configurations.

- Step 3: Select only a few of configurations, which are associated with the lowest adsorption energies for STM simulations. Since most STM experiments are carried out at very low temperatures, the surface structure measured by STM should be in (or very close to) the ground state configuration determined with DFT.

- Step 4: Compare the simulated images with the STM experimental results. Ideally, only the global ground state configuration in the simulations can reproduce the experimental results.

As can be seen, the whole structure search process is based on first-principles calculations, except the first step. The accuracy of theoretical investigations using this strategy therefore significantly depends on the how close the proposed structures are to the global ground state. Since the initial "guess" plays such an important role in the search of the ground state configuration, this strategy is efficient and reliable only when the surface substrate is simple and the adsorbates possess only very few degrees of freedoms.

Predict the adsorption configuration for large organic molecules

When the adsorbate we are interested in is large, the determination of the ground state adsorption site is always very difficult in first-principle calculations:

- Large adsorbates usually require large super cells for DFT calculations. As the size of super cell increases, the time spent for each configuration also increases.

- For large adsorbates, it is difficult to predict the actual adsorption site of the molecule. In order to determine the ground state configuration, in principle, one needs to scan the positions of the molecule in two dimensions on the surface. After each move, the total energy of the configuration has to be calculated by DFT.

- After completing the two dimensional scan, the ground state adsorption configuration is the structure associated with the lowest total energy.

This strategy is very time consuming and inefficient. By contrast, first-principles STM simulation can predict possible ground state configurations very efficiently. First of all, the STM tip used for simulation is prepared by attaching an adsorbate to the tip apex. Secondly, instead of the constant current mode, the modified STM tip should be kept at a constant hight during scanning (constant height mode). As the tip scans the clean substrate surface, a two dimensional map of tunneling current is generated. In this map, the points associated with high current values indicate the area, where the overlaps of the wavefunctions between the adsorbate and the substrate are large. Since the most stable adsorption configuration corresponds to the largest overlap of the wavefunctions, the ground state configuration is marked by the highest current value in the simulated STM image. Based on the result of this qualitative searching method, one can then perform DFT calculations for a quantitative evaluation [54].

Study the imaging selectivity of modified tips

Due to the on-going refinement of STM techniques, today even the atomic-scale electronic properties of adsorbate-substrate systems can be explored. Recently, the selective measurement of individual molecular states or surface states has become a focus of research [55, 56]. These investigations are of great interest for molecular electronics and nano-scale device fabrication [57, 58, 59, 60].

However, selectively imaging molecular or metal surfaces states in STM measurements is usually prevented by the complex coupling of the electronic states of the molecules, surfaces, and STM tips. Collaborating with the group of Hongjun Gao in Beijing, we studied the state selectivity of a modified tungsten STM tip functionalized by the attachment of organic molecules [18].

At low temperature and under UHV condition, the STM measurements were carried out on the perylene-Ag(110) system. The chemical formula of a perylene molecule is $C_{20}H_{12}$. The density of states of an isolated molecule in the vacuum is shown in Fig. 2.3. The highest occupied molecular orbital (HOMO), i.e. the π states, can be visualized by a partial charge density calculation, see Fig. 2.4(c).
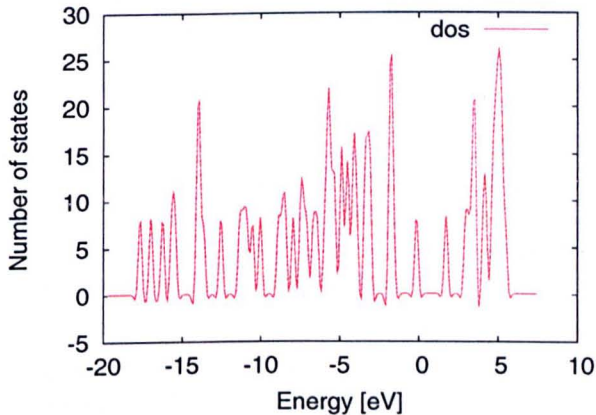

Figure 2.3: Density of states of a perylene molecule in the vacuum

As seen in Fig. 2.4(a), the molecule is flat and consists of five benzene rings. To isolate the interaction between one adsorbate and the substrate from the intermolecular interactions, we chose the coverage of about 0.6ML to determine the optimized adsorption configuration, see Fig. 2.4(a). In total, twenty possible structures varying in the orientations and adsorption sites have been analyzed. The molecule is only weakly bound to the surface, the adsorbate-substrate interactions are mainly due to the overlap between the π states of the molecule and the electronic states of the surface. The ground state configuration is shown in Fig. 2.4(a), in which the center of the molecule is above the on-top site of the surface; the longitudinal axis is parallel to the [1$\bar{1}$0] direction. With this adsorption geometry, the overlap of electronic states is maximized and leads to the adsorption energy of -0.53 eV. Comparing the partial charge density of perylene in the vacuum and adsorbed on a Ag(110) surface (see Fig. 2.4(b) and (c)), it can be seen that the molecular states of perylene remain virtually unchanged, indicating only a physisorption of the molecule [18].

Previous low energy electron diffraction (LEED) experiments showed that, on Ag(110) surfaces, the local adsorption configuration of perylene molecules does
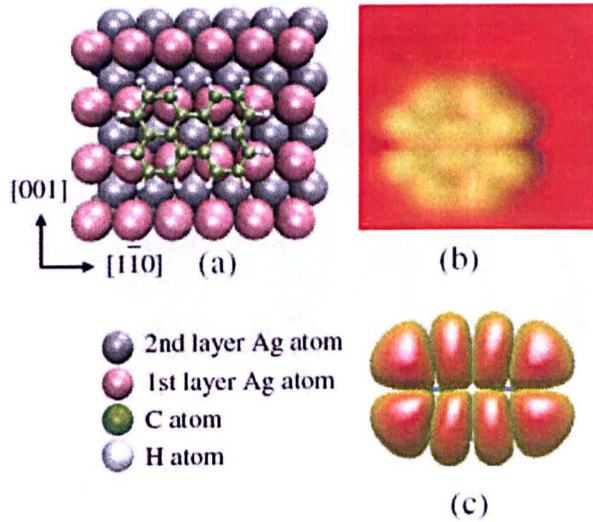
Figure 2.4: The ground state adsorption configuration of perylene on the Ag(110) surface. (b) Partial charge density of perylene molecule deposited on the Ag(110) surface. (c) Partial charge density of an isolated perylene molecule in the vacuum (from -0.7 to +0.7 eV relative to the Fermi level)

not depend much on the coverage [61, 62]. Consequently, we can use the ground state configuration shown in Fig. 2.4(a) to analyze the perylene-Ag(110) system for different coverages. Throughout this work, all STM simulations are based on the electronic structure of this configuration.

Fig. 2.5(a) shows a high-resolution STM image of a perylene monolayer on the Ag(110) surface. Each perylene molecule is imaged as four bright lobes. First-principles STM simulations indicate that this image is obtained with a clean tungsten tip (see Fig. 2.5(b) and (c)). The twist of the molecule in the experimental images is probably due to lateral intermolecular interactions of the adsorbates. Comparing Fig. 2.5 with Fig. 2.4(b), we see that the STM image obtained with a clean tungsten tip is roughly equivalent to the local charge density, i.e., the π states, of the perylene molecules on Ag(110). The surface states do not show up in either experimental images or STM simulations.

Functionalizing an STM tip is rather simple and straightforward in the experiments: move bring the tip close to the surface and apply voltage pulses. When this tip is used for STM scanning, the obtained image changes to a different mode. We thus conclude that the electronic properties of the tip have changed. The most likely option in this case is that the tungsten tip has picked up a perylene molecule. Tuning the bias voltage and current, this functionalized tip can be used to block the tunneling from the perylene states the energy levels of perylene adsorbates and image only the surface state electrons. As seen in Fig. 2.6 (a),
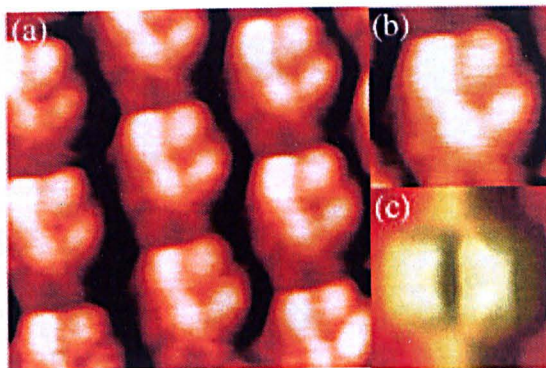
Figure 2.5: A monolayer of perylene molecules adsorbed on Ag(110) (a) A high-resolution STM image, 3nm × 3nm. (b) An enlarged image of one perylene molecule in (a), 1.2nm × 1.2nm. (c) Simulated STM image with a clean tungsten tip. Parameters: V = +1.8V, I = 90pA.

with a bias voltage of -0.8V and tunneling current of 27.5pA, part of the Ag surface appears as a protrusion. Varying the bias voltage to -0.67V and changing the current to 161.5pA (see Fig. 2.6(b)), the whole Ag surface appears and Ag rows in [1$\bar{1}$0] direction become visible. If the bias voltage is further increased to -1.5V, a current of 149pA yields a high resolution of the Ag surface while perylene adsorbates are seen as depressions, see Fig. 2.4(c). In order to ascertain that the state selectivity is due to a perylene-functionalized tip, a set of STM simulations have been carried out using the experimental tunneling parameters. As shown in Fig. 2.6 (d)-(i), the simulated images with a perylene modified tip are consistent with the features obtained experimentally. The STM simulations thus clearly demonstrate that the functionality of the tip is due to the adsorbed molecule. In addition, we also note that varying the adsorption geometry of the perylene on the tungsten tip, e.g., by rotating the molecule with respect to the vertical axis or changing the adsorption site, does not significantly affect the results of the STM simulations. This indicates that the perylene molecule on the STM tip rather than the underlying tungsten surface, is dominating the obtained STM image.

To understand the nature of this effect, an additional partial charge density calculation for a perylene molecule on a W(100) surface has been performed. As shown in Fig. 2.7(b), the π orbitals of the perylene molecule merge completely with the charge density of the metal surface, thereby changing the electronic structure of the clean tungsten tip. The functionality of this tip, i.e., the discrimination of perylene states, can be attributed to a mismatch of energy levels of perylene molecules on the Ag surface and on the modified STM tip.

To summarize, first-principles STM simulations successfully identified the functionality of modified STM tip. In addition, it also indicates that functionalizing STM tips with organic molecules can induce energy selectivity in the electronic
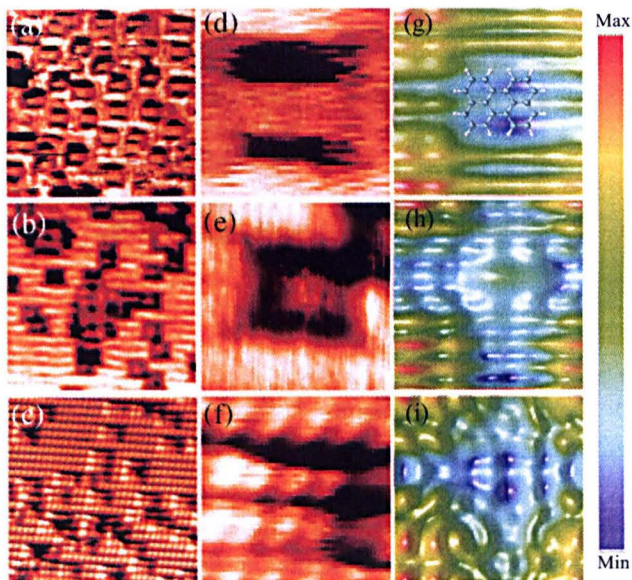
Figure 2.6: (a)-(c) STM images of perylene molecules adsorbed on Ag(110) surfaces. (d)-(f) High resolution details of a single feature in (a)-(c) respectively. (g)-(i) the corresponding STM simulation of (d)-(f) with a tungsten tip functionlized by a perylene molecule. The position of the molecule on the surface is indicated in (i). Parameters: (a) 10nm × 10nm, 0.5ML, V = -0.8V, I = 27.5pA; (b) 7.9nm × 7.9nm, 0.6ML, V = -0.67V, I = 161.5pA; (c) 10nm × 10 nm, 0.3 ML, V = -1.5V, I = 149pA

transport through organic interfaces.

2.3 Multi-scale modeling

With First Principle Density Functional Theory calculations and Scanning Tunneling Microscopy simulations, one can study the electronic structures and physical properties of any low temperature, low pressure "elementary process". By the term "elementary process", we mean a physical process taking place within the electronic regime, which is limited within a length scale of a few nanometers and a time scale of several picoseconds, see Fig.2.8.

This applies, for instances, to the interaction between a particle and a substrate, the rearrangement of electron density during the formation of a chemical bond, or the diffusion of an adsorbate on a surface, etc.. However, the physical processes within the mescoscopic and macroscopic regimes are either too complicated or too inefficient to be treated by DFT, even under the condition of "0K" and "ultra high vacuum (UHV)".

But it is well known that the organic catalysis functional surface in typical het-
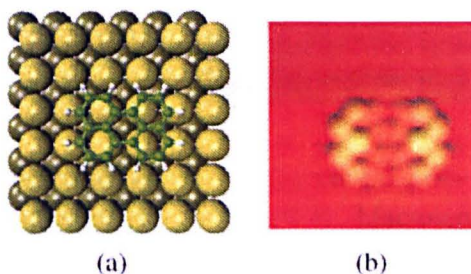
<div align="center">(a)   (b)</div>

Figure 2.7: (a) Perylene molecule attached to a W(100) surface. (b) Partial charge density of the perylene modified tungsten surface (from -0.7eV to +0.7eV relative to the Fermi level).

erogeneous catalytic cycles are usually exposed to pressures of the order of bar and temperatures well above room temperature. While the individual, elementary processes occur on microscopic timescales, the actual catalytic behavior is determined by their concerted interplay. As a consequence, the self-assembly process of e.g. the succinic acid on a Cu(110), cannot be interpreted based on the information from electronic regime alone. The proper theories to apply in mesocopic and macroscopic regimes are statistical mechanics and thermodynamics. A predictive modelling of such a multi-scale behavior must therefore combine an accurate description of all individual processes on the microscopic scale with a statistical mechanics treatment of their interplay at the mesoscopic level. Thus, an atomistic understanding of the nature of this self-ordering process can only be obtained through a multi-scale approach, which consists of electronic structure calculations, thermodynamics and statistical mechanics (see Fig.2.8). The rest of section 2.3 first reviews some basics of the theories in mesocopic regime, and then explains how statistically to deal with the information obtained from the microscopic regime.

2.3.1 Thermodynamics

In contrast to the theoretical methods described in section 2.1 and 2.2, which mainly concern physical properties of a system in the microscopic regime, thermodynamics deals with the macroscopic or bulk equilibrium properties of matter.

Thermal equilibrium

A system is in thermal equilibrium if the state of the system does not change under time-independent external conditions. As an illustration, consider a gas of N identical particles confined in a volume V. For simplicity, we assume that all the particles behave like hard spheres, so that the particles always move in straight lines with constant velocity except during collisions. If one particle moves much faster than all the others, this particle will collide with its neighbors as it moves. Because its neighbors move very slowly, this particle will tend to slow down after
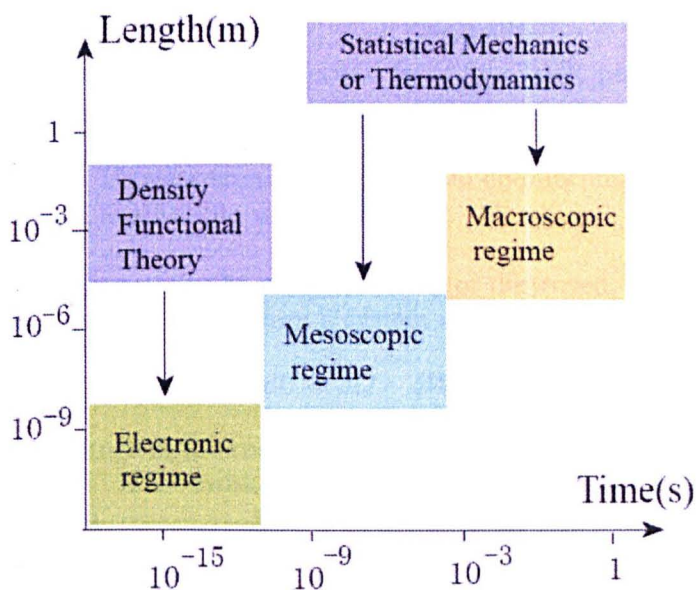
Figure 2.8: Schematic representation of the time and length scales used in chemistry and physics. The elementary processes occuring within the electronic scale can usually be investigated by DFT. Statistical mechanics and thermodynamics are required to study the observable phenomena taking place in the mesoscopic and macroscopic regimes [63].

each collision, while its neighbors will pick up energy from collisions with this particle and move faster than before. After sufficient time, the momentum of all particles will average and remain unchanged from then on. In fact, not only the momentum, but also other observable properties, such as the density of particles, will average as long as enough time is allowed to elapse. We therefore expect a uniform thermodynamic state throughout the volume V. However, if the external conditions are changed, the corresponding thermal equilibrium configurations will also evolve. These equilibrium configurations are called "thermodynamic states".

A profound consequence of the thermal equilibrium picture is that systems in thermal equilibrium exhibit uniform dynamics throughout. Since the dynamics is uniform, each thermodynamic states can be uniquely defined by only a small number of "state variables" such as number of particles "N", the volume in which particles are confined "V", the pressure that a gas is exposed to "P", the temperature "T" etc.. A change of a thermodynamic state can therefore always be referred to a change of one or more state variables. More importantly, since the state variables are independent of the path in which a particular thermodynamic state is achieved, one can always choose a convenient path to calculate the transformation of thermodynamic states. This makes thermodynamics very powerful in the study of macroscopic physical properties of a system in equilibrium.

Laws of thermodynamics

The relation of state variables are derived from four basic postulates, which are called "laws of thermodynamics".

"The Zeroth law": If two systems are in thermal equilibrium with a third, they are also in thermal equilibrium with each other.

"The First law": Energy can be neither created or destroyed, but can only transferred betweens systems, or between a system and its surroundings.

$$dU = \delta Q - \delta W \qquad (2.86)$$

Here, dU is the change of internal energy of the system, δQ is the heat added to the system and δW is the work done by the system. The energy of an isolated system is conserved.

"The second law": When two systems are brought into thermal contact, there exists no thermodynamic transformation whose sole effect is to extract a quantity of heat from the cold system to the hot system. This law plays an important role of indicating the spontaneous direction of physical and chemical processes.

"The third law": all perfect materials have the same entropy S at $T = 0K$, and this value may be taken to be $S = 0 JK^{-1}mol^{-1}$; at higher temperatures, S is always positive. However, it is impossible to cool any system to $T = 0K$.

The laws of thermodynamics are not obtained from mathematic derivations but were developed from countless observations of systems at the macroscopic level. Thus, the validity of thermodynamics is completely independent from the approximations we used in the microscopic scale. However, since thermodynamics uses only a few state variables to fully describe the average behavior of a large number of particles in thermal equilibrium, it alone cannot give any insight into the physical states of individual particles at the microscopic scale. For an overview, see ref. [64].

2.3.2 Statistical mechanics

The aim of Statistical Mechanics is to construct a bridge between the over-elaborate functions of the microscopic variables and the obscure generalities of the state variables in a macroscopic system at thermal equilibrium. Take again a gas of non-interacting identical particles of mass m as an illustration. At a set of fixed external conditions e.g. (N,V,P), each particle occupies different position coordinates and momentum vectors as time elapses. As a consequence, the whole well defined thermal equilibrium system will experience different microscopic configurations (or microstates). If all the possible microscopic configurations are collected in a set called the "state space" Ω, the number of states R contained in Ω should depend on the size of the macroscopic system and the number of degrees

of freedom of each particle.

For each microscopic configuration, the value of a macroscopic state variable can be obtained by calculating the corresponding microscopic variables for all particles, because the macroscopic property is statistical in nature. For instance, if at a particular moment $t = t_o$, our model system is in the microstate i, the energy of the system can be written as

$$E(t = t_o) = E_i = \sum_{k=1}^{N} \frac{p_k{}^2}{2m}(t = t_o) \qquad (2.87)$$

Although the E_i are well defined, the microscopic state of our system changes its microstates very rapidly. As an illustration, one mole of a typical gas will change its microstates roughly 10^{32} times per second, into another microstate i' [67]. The equation (2.87) thus fails to describe the averaged energy of the system for $(t \to \infty)$. Under these circumstances, it becomes necessary for us to know the probability P_i (statistical weight) of each microscopic configuration i in the state space Ω. Once P_i is known, the expectation value of any observable A in the equilibrium system is given by:

$$< A >= \sum_{i=1}^{R} A_i \times P_i \qquad (2.88)$$

For instance, the expectation value of energy is $< E >= \sum E_i \times P_i$, which is also the internal energy U of the system. Thus, instead of studying the time evolution of microscopic variables for each particle, the key idea of statistical mechanics is to evaluate the probability of finding the equilibrium system in one or another microstate.

The most straightforward way to find the probability P_i of state i is via "time averages": to observe a thermal equilibrium system over a appropriate time interval. During the time of observation, the system will pass through a very large number of microstates. Based on the assumption that "For a system in equilibrium, all accessible microstates are equally probable.", a reproducible value of P_i can be obtained by simple counting. In general, however, a complete set of simultaneous measurements is not accessible. As a result, only a subset of the state space Ω is averaged.

Alternatively, the probability P_i can also be obtained from an "ensemble average" approach, in which a whole set of identical systems are measured and each microstate is equally averaged over all possibilities. An ensemble is a set of sampled configurations of a system, where each individual sample may have a different microstate while sharing three common macroscopic properties [68]. To set up an ensemble, we take a sample system of specified volume V, number of particles N and think of it as replicated M times, all these M systems are thought in thermal equilibrium with each other. Since an ensemble is a collection

of imaginary replications of the system, M can be any large number, when appropriate, M can be infinite. Thus we expect that all the accessible microstates are contained in an assemble with correct proportions. The probability of a particular microstate P_i is then just the number of sampled configurations in that particular state n_i, divided by the total number of samples M. The statement that "the time average and the ensemble average are equivalent" is known as the "ergodic hypthesis [69].

Depending on the nature of the macroscopic constraints, we have different ensembles. If all samples are "isolated systems", i.e. neither energy nor matter can be exchanged between samples, the ensemble is called a "microcanonical ensemble" (N, V, E common). Since "all microstates of an isolated system are equally probable" [69], we have $P_i = n_i/M = 1/R$. If the condition of "isolated systems" is replaced by "closed systems" and all samples are in thermal contact with a heat bath, the corresponding ensemble is called a "canonical ensemble" (N, V, T common). In a closed system, P_i is proportional to $exp(-E_i/k_BT)$, where k_B is the Boltzmann constant (1.380658×10^{-23} J/K). If we proceed further and consider samples as "open systems", which means both energy and matter can be exchanged between the samples, the ensemble is called a "grand canonical ensemble" (N, V, μ common, where μ is the chemical potential in each sample system). The value of P_i in a open system is proportional to $exp[-(E_i + \mu N_i)/k_BT]$. See Fig. 2.9.
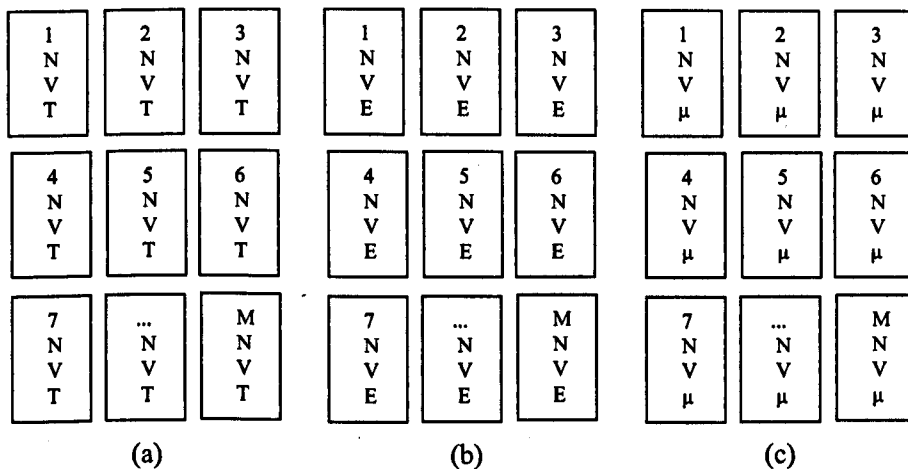
Figure 2.9: A schematic representation of (a) a canonical ensemble, (b) a microcanonical ensemble and (c) a grand canonical ensemble

Once a proper type of ensembles has been chosen, all statistical properties of that system are contained in the partition function of the ensemble [69]. Considering the most widely employed conditions of typical STM operations, where a certain coverage of adsorbates are scanned at a constant temperature in UHV, it is a nat-

ural choice to use the canonic ensemble for numerical modeling. The partition function of a canonical ensemble is

$$Z = \sum_i^R e^{-E_i/k_B T} \qquad (2.89)$$

where E_i is the energy level if the system is at the microstate i, T is the temperature in Kelvin, and k_B is the Boltzmann constant. The sum in (2.89) is over all possible microscopic configurations contained in state space.

Direct connections between the partition function and thermodynamic state variables are established by the Helmholtz free energy [70]:

$$F = -k_B T ln Z \qquad (2.90)$$

All other thermodynamic state variable can be calculated from this relation. For instance, the internal energy of a system is $U = -T^2 \partial(F/T) \partial T$. The equation (2.90) therefore plays the role of a bridge between thermodynamics and statistical mechanics.

The probability that a system is in the microstate i is given by the Boltzman distribution [69].

$$P_i = \frac{1}{Z} e^{-E_i/k_B T} \qquad (2.91)$$

Substituting equation (2.91) into (2.88), we obtain a much better expression for the expectation value of any observable A.

$$<A> = \sum_{i=1}^R \frac{1}{Z} e^{-E_i/k_B T} A_i \qquad (2.92)$$

However, there are still two unknown variables in Equation (2.92). The energy level of each microstate E_i and the value of the partition function Z. The former can generally be calculated in the microscopic regime, the method of this calculation, which is called "cluster expansion", will be discussed later in this chapter. The latter, however, can usually not be calculated, because the number of possible microstates R can be astronomically large. For a typical system (e.g. a mole of gas), which is an assembly of $N = 10^{23}$ particles, the number of terms contained in the partition function would be of order N^N! This is far beyond our computational capacity. Nevertheless, as we will show in succeeding sections, this obstacle of applying equation (2.92) can be removed by the Monte Carlo method.

2.3.3 The lattice gas model

The archetypal lattice gas model is a two dimensional interacting system, where each lattice site can either be occupied by a particle or vacant. By the term "sites", we mean the well-defined positions, where atoms or molecules adsorb on a surface. The occupancy variable $\sigma = 1,0$ is used to represent occupied

or unoccupied sites, respectively. If the particle interacts with its four nearest neighbors, the Lattice Gas Hamiltonian (LGH) is:

$$E = \sum_i \sigma_i E_o + \frac{1}{2} \sum_{i,j} V \sigma_i \sigma_j \qquad (2.93)$$

where V is the nearest neighbor interaction, and E_o is the adsorption energy that controls the coverage of the adsorbates on the surface. A large positive E_o often corresponds to a low coverage of particles, while a large negative E_o often indicates that most of the lattice sites are occupied. As an illustration, we use oxygen atoms and a Pd(100) surface, see Fig 2.10(a). Oxygen atom adsorbs on the fcc hollow site of the Pd surface. In this model, the nearest neighbor distance will be the lattice constant a_o. The energy E of a particular lattice configuration is calculated by investigating the four nearest neighbors of each site. A pair interaction V can be calculated for two adsorbates separated by a distance a_o. Once the values of E_o and V are known, the energies of all lattice configurations can be calculated exactly. The probability for each lattice configuration is then given by equation (2.91).
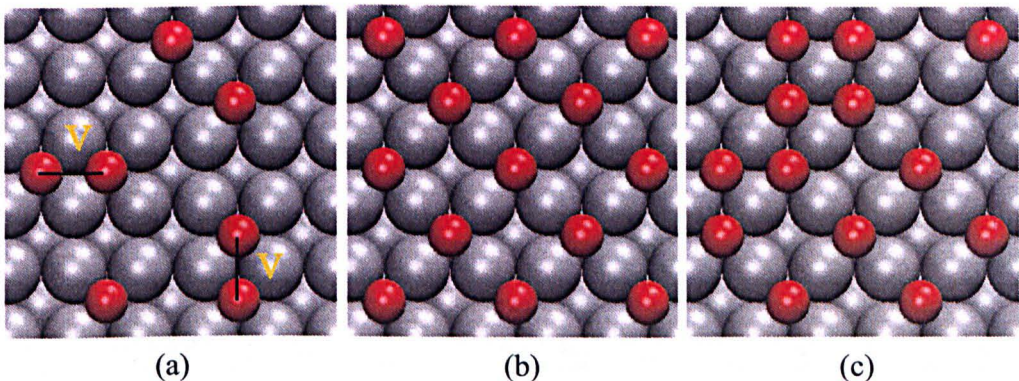

(a)   (b)   (c)

Figure 2.10: (a) A schematic representation of the lattice gas model, where only the binding energy E_o, and the interaction V between nearest neighbor are considered (b) an example of an ordered lattice configuration (c) an example of a disordered lattice configuration

Even though the lattice gas model appears to be a dramatic simplification of real systems, it is one of the most useful methods to study the thermodynamic properties at surfaces. Once thermal equilibrium has been reached, the coverage of the surface remains constant, and the adsorbates spend most time in their adsorption sites, while the diffusion and desorption processes take place only in very short intervals. Since for the purpose of this work, we are mainly interested in the equilibrium properties of succinate molecules on Cu(110), we concentrate in the following on equilibrium states of adsorbates.

In most real systems, the energy value of a certain lattice configuration cannot be simply expressed by equation (2.93), but has to be calculated explicitly. In practice, this can be done by zero temperature quantum mechanical methods, e.g. DFT. The obtained results, however, are not limited to low temperature applications, because the energy levels of a lattice gas model remain unchanged as the temperature increases. The temperature dependent quantity is just the occupancy of these energy levels, i.e. the Boltzmann distribution.

Unfortunately, only the lattice configurations which are perfectly ordered (for instance, Fig. 2.10(b)), can be simulated by today's DFT codes. For large disordered configurations (for example, Fig. 2.10 (c)), the quantum mechanical calculations become extremely inefficient. Since direct first-principles calculations are not an option, we have to resort to approximations. One obvious approach is to express the energy values by the lattice gas hamiltonian in equation (2.93). However, the number of energy parameters should be much larger than just two. Apart from pair interactions, we also need take into account trio interactions, quatro interactions and higher orders. Each energy parameter can be explicitly evaluated by first principle calculations. This approach is called the "effective cluster interaction method". Some typical interaction energy parameters are shown in Fig. 2.11.
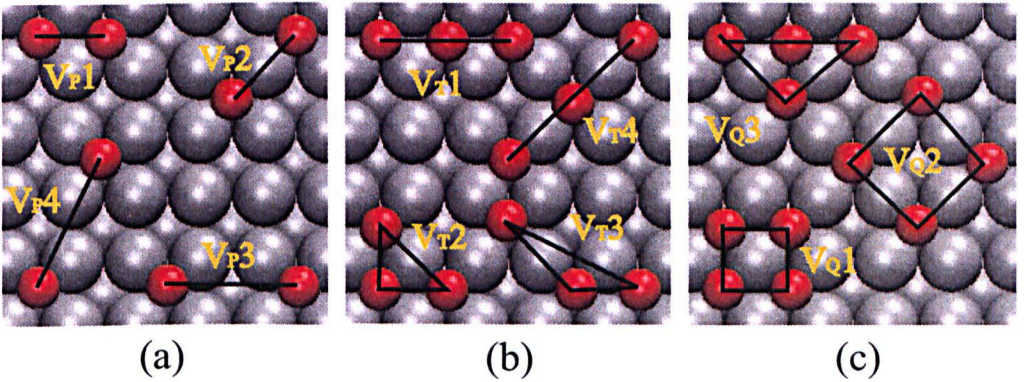

(a)   (b)   (c)

Figure 2.11: Some typical energy parameters of (a) pair interactions, (b) trio interactions and (c) quatro interactions

The corresponding lattice gas hamiltonian is given by

$$E = \sum_{i}^{all\ sites} \sigma_i E_o + \frac{1}{2} \sum_{i \neq j}^{all\ site\ pairs} V_{P_{(i,j)}} \sigma_i \sigma_j + \frac{1}{3} \sum_{i \neq j \neq k}^{all\ site\ trios} V_{T_{(i,j,k)}} \sigma_i \sigma_j \sigma_k + \cdots \quad (2.94)$$

where E_o is the isolated binding energy of the adsorbate on the surface, V_P and V_T denote to the pair interactions and trio interactions respectively. We again take oxygen on Pd(100) as an illustration. Suppose that the size of our unit cell is large enough, so that the lattice configurations do not feel the existence of oxygen atoms in their translational symmetric images. (See Fig. 2.12.) Then the energy
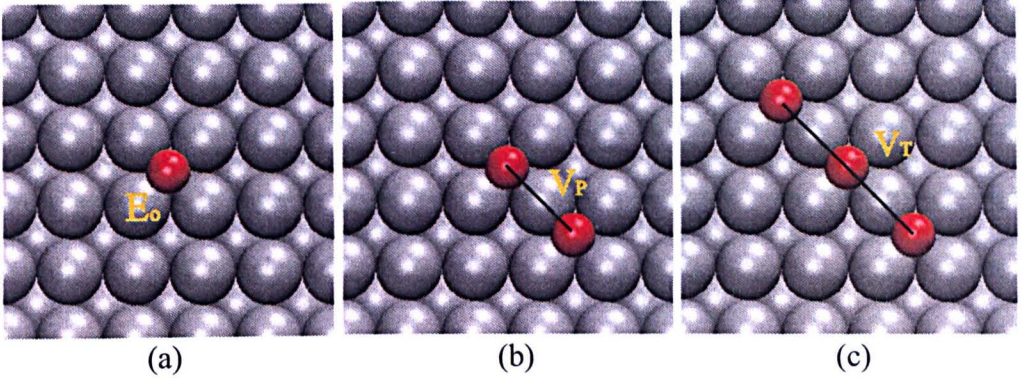
Figure 2.12: (a) An isolated oxygen on a Pd(100) surface, there is only one energy parameter : E_o (b) For paired configurations, there are two energy parameters, E_o and V_P (c) Three energy parameters in this configuration: E_o, V_P and V_T.

parameters can be evaluated by the equations below:

$$
\begin{aligned}
E_o &= E(\text{unit cell (a)}) - E(\text{clean surface}) - 1/2E(O_2) \\
V_P &= E(\text{unit cell (b)}) - E(\text{clean surface}) - E(O_2) - 2E_o \\
V_T &= E(\text{unit cell (c)}) - E(\text{clean surface}) - 3/2E(O_2) - 3E_o - 2V_P
\end{aligned}
$$

If all interactions are considered in the sums of Eq. (2.94), then, one could describe the energy value E_i for any lattice configuration i. However, for complex disordered lattice configurations, the number of energy parameters can be very large rendering the calculations very inefficient. Fortunately, the expansion of a lattice gas hamiltonian converges very rapidly with respect to the size of the considered clusters. We therefore expect that the energy of a lattice configuration can be expressed only by the short-distance pair interactions and small-size trio interactions.

2.3.4 Monte Carlo Methods

So far, the only unknown parameter in Eq. (2.92) is the exact value of the partition function for an ensemble. As discussed in section 2.3.2, the main difficulty which prevents an explicit evaluation of a partition function is the large number of microstates associated with the system. Fortunately, Eq. (2.92) can still be applied to thermal equilibrium systems by averaging the microstates generated by Monte Carlo technique.

Importance sampling

Let us first suppose that the value of the partition function Z is known. Then a straightforward next step is to calculate the energy E_i for each point i in the state space Ω. The value of E_i can then be substituted to equation (2.92) to obtain the expectation value of a macroscopic quantity. Unfortunately, it is not a

feasible approach. As one can expect, the total number of points R for a typical mesoscopic system is too large to be calculated. Of more importance, even if such a procedure could be performed, the obtained results would be subject to a large statistical error. Since a complete enumeration through phase space is not an option, we therefore consider a random method as a possible alternative : average only a (randomly generated) subspace of Ω to represent the expectation value of A with a reasonable accuracy. For the sake of efficiency, the number of sampling points M should be much smaller than the total number of microscopic configurations R. The average value of A is give by

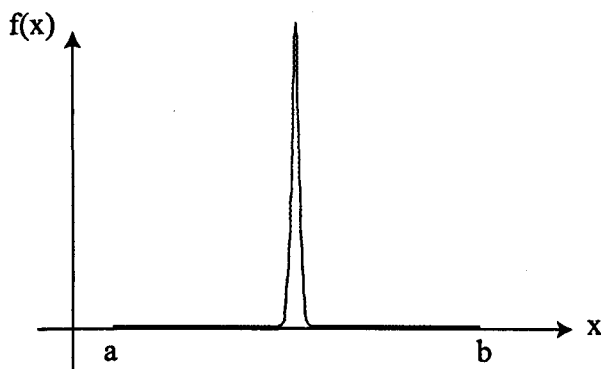$$A_M = \frac{\sum_{i=1}^{M} A_i e^{-E_i/k_B T}}{\sum_{i=1}^{M} e^{-E_i/k_B T}} \tag{2.95}$$

Figure 2.13: evaluation of the expectation value of a one dimensional function f(x) with a random sampling method

The accuracy of such a random sampling method can be illustrated by a one dimensional integral between two limits (see Fig. 2.13). The expectation value of f(x) should be calculated by a numerical integral over all points between a and b :

$$< f(x) >= \int_a^b f(x)dx/(b-a).$$

In a random sampling method, m points on the line are randomly selected and the corresponding values of $f(x)$ at these points $f(x_1), f(x_2), ..., f(x_m)$ can then be measured. The average value of $f(x)_m$ is given by

$$f(x)_m = \frac{1}{m} \left[f(x_1) + f(x_2) + ... + f(x_m) \right].$$

It is quite clear that the quality of the obtained average depends largely on the locations and quantity of the selected sampling points. Since m must be a very small fraction of the total number of points, the obtained averages from such a

random sampling are generally neither close to the expectation value nor reproducible. In this example, the direct Monte Carlo sampling method fails mainly because there is a large region where the values of $f(x)$ are negligibly small, while the non-zero value region is going to be scarcely sampled by a random selection. Thus, it would be much more efficient to sample a large number of points in the region where f(x) has non-zero values and few elsewhere. In other words, the number of sampled points in a region should be proportional to their contribution in the expectation value $< A >$. This is a the basic idea of "importance sampling" in a Monte Carlo simultaion.

Similar with the one dimensional model above, in macroscopic equilibrium, there are always a large number of microstates whose Boltzmann factors P_i are very small (effectively zero) due to the high value of configuration energy E_i (equation (2.91)). Only a very small proportion of the state space corresponds to low energy configurations where sufficiently large Boltzmann factors are weighted. This is also a reflection of the nature of the state space, and the low energy region coincide with the physically observed phases such as a solid, liquid, etc. [71]. Since only an extremely small part of the state space is important in determining the averages, we resort to an "importance sampling" Monte Carlo method: the microstates are not sampled with equal probability, but according to their population in the ensemble. The states associated with a significant value of the Boltzmann factor are to be more easily to selected and sampled than the others. These selected microstates, when sampled in adequately large enough numbers, should constitute a very good approximation to a canonical ensemble. The expectation value $< A >$ of a macroscopic property can then be calculated as a simple arithmetic mean over the sampled microstates, Eq. (2.95).

Markov chain

In order to perform such a importance sampling, one needs to know the Boltzmann factor for each microstate, so that the sampled points can be properly distributed in the state space. Theoretically, this decision can be made via the Boltzmann equation, Eq. (2.91). However, as shown in equation (2.89), the precise value of the partition function Z needs to be calculated in the first place. Thus, there is a catch that brings us back to the original problem of summing over an impossibly large number of states.

Luckily, Metropolis and his co-workers demonstrated that it is possible to select the "important states" randomly by generating a "Markov chain" of configurations according to certain probabilistic rules. If the number of configurations in the chain is sufficiently large, the selected states are qualified to constitute an equilibrium canonical ensemble [72].

A Markov chain is a discrete stochastic process in time. Consider a system with a finite set of possible states $(s_1, s_2, ..., s_R)$. At each time $(t_0, t_1, t_2, t_3, ...)$ of a discrete sequence, the system can be described as being in one of these possible

states. Let s_{t_n} be the state in which the system is at time t_n. The state s_{t_n} is random and can be any state s_i in the configuration space Ω with a conditional probability P_i. For a system in equilibrium the probability of $s_n = s_i$ is given by the Boltzmann distribution $P_i = exp(-E_i/k_BT)/Z$, which is independent of the time index t_n. In a discrete time stochastic evolution, the system starts from an arbitrary initial state s_{t_0}, experiences a number of states $s_{t_1}, s_{t_2}, ...$ and ends at s_{t_n}. For any state evolution process, the joint probability $P(s_{t_0}, s_{t_1}, s_{t_2}, ..., s_{t_n})$ can be calculated by

$$
\begin{aligned}
P(s_{t_0}, s_{t_1}, s_{t_2}, ..., s_{t_n}) = \quad & P(s_{t_0}) \times \\
& P(s_{t_1}|s_{t_0}) \times \\
& P(s_{t_2}|s_{t_1}, s_{t_0}) \times \\
& \cdots\cdots\cdots \times \\
& P(s_{t_{n-2}}|s_{t_{n-3}}, s_{t_{n-4}}, \cdots, s_{t_1}, s_{t_0}) \times \\
& P(s_{t_{n-1}}|s_{t_{n-2}}, s_{t_{n-3}}, \cdots, s_{t_1}, s_{t_0}) \times \\
& P(s_{t_n}|s_{t_{n-1}}, s_{t_{n-2}}, \cdots, s_{t_1}, s_{t_0}).
\end{aligned}
$$

The sequence of the state variables $s_{t_0}, s_{t_1}, s_{t_2}, ..., s_{t_n}$ constitute a Markov chain if the probability for each s_{t_i} at time t_i is independent of all states but only the immediate predecessor $s_{t_i - 1}$:

$$
P(s_{t_n}|s_{t_{n-1}}, s_{t_{n-2}}, \cdots, s_{t_1}, s_{t_0}) = P(s_{t_n}|s_{t_{n-1}}). \tag{2.96}
$$

The Markov assumption implies that the probability of a system to visit a particular state s_i at time t_{n+1} depends only on the current state s_{t_n} but not the previous history of the system. Thus, for all Markov chains, we have

$$
P(s_{t_0}, s_{t_1}, s_{t_2}, ..., s_{t_n}) = P(s_{t_0})P(s_{t_1}|s_{t_0})P(s_{t_2}|s_{t_1}) \cdots P(s_{t_n}|s_{t_{n-1}}). \tag{2.97}
$$

The Equation (2.96) can therefore be used to represent the transition probability of moving from the microstate i to microstate j in a single step:

$$
W_{ij} = W(s_i \rightarrow s_j) = P(s_i|s_j). \tag{2.98}
$$

Detailed balance

Up to this stage, the task has been to calculate the transition probability W_{ij} for all microstates. Once this is done, one would be able to construct a Markov chain of configurations according to the transition probabilities and subsequently obtain good approximation to an equilibrium canonical ensemble. The expectation value of any macroscopic property can then be evaluated by averaging the sampled configurations.

Let $P(s_i, t_n)$ denote the probability of finding the system in the state s_i at time

t_n. Then, the transition probability for all microstates should satisfy the master equation

$$\frac{\partial P(s_i, t_n)}{\partial t} = -\sum_j W_{ij} P(s_i, t_n) + \sum_j W_{ji} P(s_j, t_n) \qquad (2.99)$$

and the sum rule

$$\sum_i P(s_i, t_n) = 1 \qquad (2.100)$$

for any discrete time $t_n = (t_0, t_1, t_2, ...)$. Since $P(s_i, t_n)$ is a real number between zero and 1, the system would come to equilibrium as $(n \to \infty)$. Once the equilibrium has been reached, as discussed in 2.3.2 for any equilibrium systems, $P(s_i, t_n)$ would become a constant (Boltzmann distribution). As a consequence, the right hand side of (2.99) must become zero, i.e.

$$\sum_j W_{ij} P(s_i, t_n) = \sum_j W_{ji} P(s_j, t_n). \qquad (2.101)$$

Equation (2.101) is referred as the condition of the "global equilibrium". Numerically, in order to avoid the "dynamic equilibrium" (a Markov chain which is trapped in a cycle of limited states), we impose a stronger (sufficient but not necessary) condition called "detailed balance" to make sure a global equilibrium is achieved. The expression of the detailed balance condition is given by

$$W_{ij} P(s_i) = W_{ji} P(s_j) \qquad (2.102)$$

which implies that

$$\frac{W_{ij}}{W_{ji}} = \frac{P(s_j)}{P(s_i)} = exp[-(E_j - E_i)/k_B T] = exp(-\Delta E/k_B T) \qquad (2.103)$$

where

$$\Delta E = E_j - E_i. \qquad (2.104)$$

As seen, the ratios of the equilibrium transition probabilities between two microstates is just the ratios of their Boltzmann weights. Thus, we do not need the precise value of the partition function to construct an approximation of an ensemble. What we need now is an algorithm which takes a system from one microstate to another in such a way that Eq. (2.103) is satisfied. This algorithm is known as the "Metropolis algorithm" [72].

Metropolis Algorithm

Let us take the lattice gas model as an example. The Metropolis algorithm preforms stochastic dynamics (a Markov chain) which produces, starting from an arbitrary initial microstate s_{t_0}, a sequence of microstates

$$s_{t_0} \to s_{t_1} \to s_{t_2} \to \cdots s_{t_n} \to s_{t_{n+1}} \to \cdots .$$

57

The transition of $s_{t_i} \rightarrow s_{t_{i+1}}$ for each Metropolis algorithm should be non-deterministic, i.e. the sequence of configurations in a Markov chain is not predetermined and fulfill the ergodicity condition, i.e. all states are accessible from each other with the finite length of a Markov chain. In addition, considering that the "important states" in an ensemble are usually confined in a rather narrow energy region, the selected state at time t_n should not differ much from the state at time t_{n-1}. The whole process of the Mertropolis algorithm can then be described as follows:

1. Select an initial microstate s_{t_0} randomly form the state space Ω with equal probability $1/R$ for all possible states.

2. Let s_i to be the state at time t_{t_0}, then we randomly and with equal probability choose an occupied site and an empty site in configuration s_i and swap their occupation variable to obtain another configuration s_j.

3. Calculate the lattice energies for these two configurations with the cluster expansion method.

4. The energy difference of state s_i and state s_j is defined by $\Delta E = E_j - E_i$. According to the detailed balance condition, we have the transition probability $W_{ij} = exp(-\Delta E/k_B T)$

5. Accept s_j as the state s_{t_1} in the Markov chain with a probability that consists with the detail balance condition. We rewrite the transition probability from state s_i to and other state s_j as

$$W_{ij} = S_{ij} \times A_{ij}, \tag{2.105}$$

where S_{ij} denotes the selection probability of s_j from s_i, and A_{ij} represents the acceptance probability of s_j to be the state s_{t_1}. The Equation (2.103) can therefore be written is following

$$\frac{W_{ij}}{W_{ji}} = \frac{S_{ij} \times A_{ij}}{S_{ji} \times A_{ij}} = e^{-\Delta E/k_B T}. \tag{2.106}$$

According to the selection method described in step 2, it is very clear that

$$S_{ij} = S_{ji}$$

Thus, we have

$$\frac{A_{ij}}{A_{ji}} = e^{-\Delta E/k_B T} \tag{2.107}$$

However, in order to make the algorithm more efficient, the acceptance ratio should be as large as possible so that the system could visit a wide region in the state space. Considering that the value of an transition probability cannot be larger than one, the value of acceptance probabilities is defined as $P(s_i \rightarrow s_j) = min\left(1, e^{-\Delta E/k_B T}\right)$. In other words

$$P(s_i \rightarrow s_j) = \begin{cases} 1 & \text{if } \Delta E < 0 \\ e^{-\Delta E/k_B T} & \text{if } \Delta E > 0 \end{cases}$$

In practice, we accept the state s_j by

$$s_{t_1} = \begin{cases} s_j & \text{if } \Delta E < 0 & s_j \text{ is accepted} \\ s_j & \text{if } P(s_i \rightarrow s_j) > ran & s_j \text{ is accepted} \\ s_i & \text{if } P(s_i \rightarrow s_j) < ran & s_j \text{ is rejected} \end{cases}$$

where ran is a random number, whose value is uniformly distributed between zero and one.

6. A Markov chain of microstates can then be produced by iterating the above steps i.e. repeatedly accepting or rejecting a proposed trial "move".

To sum up, in the Meropolis algorithm a random walk through the whole configuration space is required (step 1 and repeating step 2), so that the sequence of microstates in a Markov chain is not predefined. This also ensures that the ergodicity condition is fulfilled, because in the lattice gas model, any state s_j can be obtained by switching occupation variables finite times from another state s_i. Step 3 to Step 5 guarantee that the random walk is in a "importance-weighted" fashion, which satisfies the condition of detailed balance. The Markov chain of configurations are therefore generated with a time ordered path. Given sufficient long "time" the sampled points should be able to construct an "effective ensemble". However, one needs to note that the "time" in this instance is referred to as "Monte Carlo time". There is no direct relation between the Monte Carlo time and real physical time. As a consequence, the Monte Carlo method described above is only used to study the physical properties in equilibrium rather than kinetic properties.

Chapter 3

Calculations on Succinate/Cu(110)

3.1 Density Functional Theory Calculations

All the DFT calculation in this work were preformed with the Vienna Ab-initio Simulation Package (VASP), a plane wave based density functional code developed by G. Kresse and J. Furthmüller [73, 74, 75, 76]. Even though the detailed computational method has been already described in section 2.1, there are quite a few parameters that need to be determined before a set of systematic theoretical investigations starts. Take the density of sampled k-points in the first brillouin zone as an example: although a denser the k-mesh leads to a better description of the electronic structure, the calculations will be substantially slower. As a consequence, the values of these parameters need to be selected in the region of the convergence threshold, where both the reliable numerical accuracy and a reasonable amount of computing time are balanced.

3.1.1 Computational Details

The typical computational parameters in a VASP calculation are given below.

- The density of sampled k-points in the irreducible first brillioun zone.

- The cut-off energy for plane wave basis set. All plane-waves with a kinetic energy smaller than E_{cut} are included in the basis set, see equation (2.42).

- The thickness of the slabs.

- The vacuum region between slabs in the z direction.

Based on the results of a set of test calculations, the copper surfaces throughout this work are modeled by super-cells that contain a four-layer slab, in which the top two layers of Cu atoms are allowed to relax in three dimensions, and a vacuum range of about 12Å, which is large enough to ensure that the metal surface interacts only negligibly with its periodic images. The Γ centered mesh with 3x2x1 and 1x1x1 special k-points is used for 6x6x4 and 8x6x4 unit cells,

respectively. This density of sampled k points gives a relative energy convergence of about 5 meV per unit cell. In all systems, the energy cutoff for plane waves is 400 eV; the ionic relaxation is stopped when the forces on all relaxed atoms are smaller than 0.01eV per Angstrom. In addition, as has been discussed in section 2.1.5, the ion-electron interaction in all systems are described by ultrasoft pseudopotentials and the exchange correlation potentials are calculated by the generalized gradient approximation (GGA) of Perdew and Wang (PW91) [42]. Due the limited computational capacity, the tests of the GGA functions and pseudopotentials are omitted.

3.1.2 Bulk Cu

Cu is a 4d transition metal element whose atomic structure of $[Ar]3d^{10}4s^{1}$. The bulk copper crystallized in a face centered cubic (fcc) structure. The equilibrium
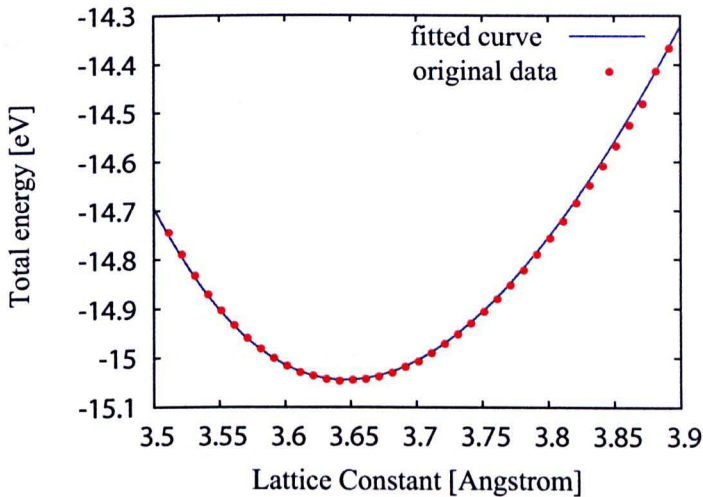

Figure 3.1: Determining the equilibrium lattice constant a_o a Cu bulk by the Murnaghan fit. The red dot represents the original calculation data, while the blue line is the fitted curve. The most optimized lattice constant obtain from this fitting is 3.64 Å.

lattice constant a_o (nearest neighbor distance between two Cu atoms in the bulk) in this work is determined by minimizing the total energies of the primitive unit cell. In practice, VASP calculations are carried out for an increasing lattice constants in the range of 3.50 to 3.9 Angstrom with an interval of 0.01 Angstrom. The obtained the energetics are then used for a Murnaghan equation state of fit [77, 78, 79]:

$$E(V) = E(V_o) + \frac{B_o V}{B'_o(B'_o - 1)}[B'_o(1 - \frac{V_o}{V}) + (\frac{V_o}{V})^{B'_o} - 1]. \qquad (3.1)$$

Here V_o is the equilibrium unit cell volume, which is related to a_o by $V_o = a_o^3$. The bulk modulus B(T,V) is related to the curvature of E(V) by:

$$B(T, V) = V \left. \frac{\partial^2 E(V)}{\partial V^2} \right|_T .$$

(3.2)

Based on the results of the Murnaghan fit (see Fig. 3.1), a_o has been determined to be 3.640 Angstrom. This value is 0.6% larger than the experimental data (3.608 Angstrom) [80]. The difference is attributed to the fact that the binding energies are always slightly under estimated by GGA functionals. The obtained equilibrium bulk modulus is 139 GPa, which is also consist with the results obtained from experiments and all-electron GGA-PBE calculations [80, 81], see Table 3.1.

Table 3.1: Comparison of the calculated bulk variables with the experimental data and other theoretical results [80, 81]

Variables	Murnaghan fit	Experimental data	GGA-PBE
Equilibrium lattice constant (Å)	3.640	3.608	3.628
Equilibrium bulk modulus (GPa)	139	138	139

3.1.3 Succinic acid molecule in gas phase

The 3D configuration of the succinic acid molecule is constructed according to the bonding parameters (bond lengthes and bond angles) of carbon, oxygen and hydrogen atoms in organic compounds [80]. As can be seen in Fig. 3.2, the four
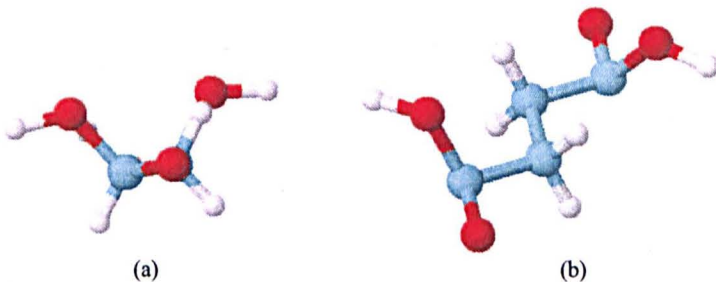

(a)   (b)

Figure 3.2: The molecular configuration of a succinic acid molecule in the gas phase. The blue balls represent the carbon atoms, while the red ones and white ones represent the oxygen atoms and hydrogen atoms respectively. (a) side view (b) top view.

carbon atoms and two of the oxygen atoms are in the same plane, the initial C-C

and C-O bond lengthes are set to be 1.50Å and 1.30 Å respectively. The typical H-C-H and O-C-O bond angle is 109.5 degree and the C-H bond length is known to be 1.01 Å [80].

The total energy of the succinic acid molecule in gas phase is calculated with a supercell geometry: the molecular configuration shown in Fig. 3.2 is placed in a 20Å × 20Å × 20Å unit cell. This size is large enough to prevent any interactions from periodic images in three dimensions. Since the above initial configuration is only a reasonable guess, all atoms in this structure are allowed to relax according to their instantaneous ground states. Considering the absence of dispersion in such a small Brillouin zone, we limit the calculation to only one k-point. In addition, this calculation is performed spin polarized so that the Hund's role principle is fulfilled [4]. With this procedure, the obtained total energy for a fully relaxed molecular configuration is $-86.075eV$.

3.1.4 An isolated succinate molecule on a Cu(110) surface

The details of the molecular adsorption configuration of succinic molecule on a Cu(110) surface are experimentally unknown. The only information we have got is that the molecule deprotonates and bonds to the Cu surface as a succinate, in which the four C-O bonds are chemically equivalent. With so little information, a complete ab-initio search of the ground state adsorption structure is difficult. However, as the ground state adsorption configuration is used for all DFT calculations, the accuracy of obtained energetics may significantly affect the quality of our Monte Carlo simulations.

Since the molecules are known to form chemical bonds with the metal surface, the nature of this chemical adsorption must be associated with the Cu-O interactions. As shown in the STM images of ref. [21], the long axis of the imaged adsorbates is roughly parallel with the [100] direction of the metal substrate, the adsorbates' structure seems consistent with the Antonio configuration [29]. We thus use an Antonio succinate configuration to determine the binding site of the an isolated succinate molecule on a Cu(110) surface.

In practice, (see Fig.3.3) the adsorption position of a succinate molecule is varied by putting the oxygen atoms above the on-top sites, short-bridge sites, fcc-hollow sites and long-bridge sites of the Cu surface and perform DFT calculations. During relaxations, only the two bottom layers of Cu atoms are fixed. The adsorption energy of each configuration is determined by

$$E_{ads} = E_{unit\ cell} + E_{H_2} - E_{clean\ surface} - E_{succinic\ acid}. \qquad (3.3)$$

Here, E_{H_2} is the energy of one hydrogen molecule in the gas phase, and $E_{succinic\ acid}$ is the energy of one succinic acid molecule in the gas phase. The energetic values obtained from DFT calculations are listed in Table 3.2. As can be seen, the configuration in which the oxygen atoms are above Cu atoms leads to the maximum binding energy. Taking into account the large energy penalties for other sites,
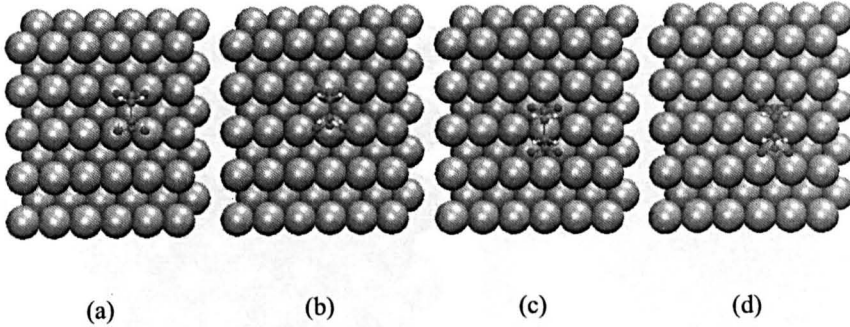
Figure 3.3: Four configurations are used to determining the binding sites. The oxygen atoms of a succinate molecule are located above (a) the on-top sites, (b) short-bridge sites, (c) fcc-hollow sites and (d) long-bridge sites.

the adsorption sites of succinate molecules should exclusively be the on-top sites of Cu(110).

Table 3.2: Determination of adsorption sites

Configuration	$E_{unit\ cell}$(eV)	E_{H_2}(eV)	$E_{clean\ surface}$(eV)	$E_{succinic\ acid}$(eV)	E_{ads}(eV)
On-top sites	-475.802	-6.771	-395.116	-86.075	-1.382
short-bridge sites	-474.824	-6.771	-395.116	-86.075	-0.404
fcc-hollow sites	-473.322	-6.771	-395.116	-86.075	1.098
long-bridge sites	-474.527	-6.771	-395.116	-86.075	-0.107

In order to obtain an understanding of the "spontaneous twist" of the adsorbates proposed in ref. [29], the initial configurations in Fig. 3.3 are designed so that the long axis of the succinate molecule is parallel with the [100] direction of the Cu surface. After the ionic relaxations, no obvious molecular twist is observed. However, if the backbone of the adsorbate is twisted, the relaxation lead to the Antonio's configuration, see Fig. 3.4(a). The resulting adsorption energy is 5meV higher than that of Fig. 3.3 (a). We thus agree that Antonio's adsorption model corresponds to a minimum in the energy space. Here, a rectangular footprint is formed on Cu(110). A simulated STM image of the molecule-surface system using the Tersoff-Hamann method is shown in Fig. 3.4(b)

An systematic comparison of the stabilities of the Darling and Antonio adsorption model has been performed on a Cu(110) 6x6 unit cell, see Fig. 3.5. The estimation of the relative molecular strain of a succinate molecule in these two models are carried out in a large vacuum unit cell, where the molecular structure in Fig. 3.5 (a) and (b) are kept fixed and all oxygen atoms are saturated by hydrogen. The surface strains are calculated by fixing all the surface atoms as

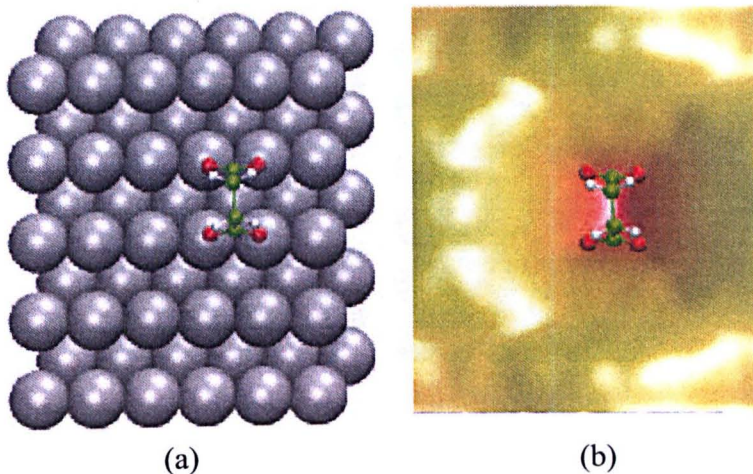(a)   (b)

Figure 3.4: (a) The relaxed configuration of a succinate acid molecule on a Cu(110) with a little twist. (b) Simulated STM image with Tersoff-Hamann method, $V_{bias} = -0.21V$.

they are in Fig. 3.5 (a) and (b), but without any of the molecular structures. The calculated results are listed in Table 3.3. As can be seen, the Darling model is 44 meV more stable than the Antonio model. This value coincides with the difference of the total energies due to surface strain and the molecular twist between these two configurations. We thus conclude that the Darling model is the ground state adsorption configuration, the physical reason is that this structure is closer to the ground state structure of the succinic acid in the gas phase and thus pays a lower energy penalty prior to adsorption. At low temperatures, all succinate molecules should exclusively take this structure on the Cu(110) surfaces.

Table 3.3: Comparison of the Antonio and Darling adsorption model

	Adsorption energy(eV)	relative molecular strain(eV)	surface strain(eV)
Antonio's model	-1.572	0.174	0.070
Darling's model	-1.616	0.000	0.140

An detailed structure analysis of the Darling model has been undertaken on a Cu(110) $8 \times 6 \times 4$ unit cell with a $2 \times 2 \times 1$ k-mesh, see Fig.3.6 (a). Interestingly, the four oxygen atoms in the succinate molecule are not all equivalent. The oxygen A and D are close to the exact on-top site of the copper surface, while oxygen B and C are about 0.72 Å away from the on-top sites. In the z direction, oxygen A and D are lower than oxygen B and C by about 0.12 Å. As a consequence, the oxygen-copper bond length of oxygen A and D (1.9 Å) is also 0.1Å shorter than
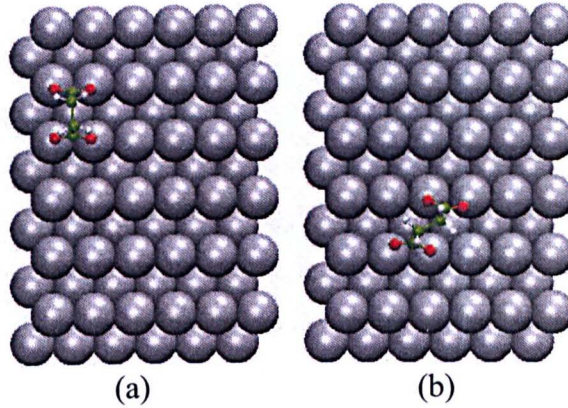
65

Figure 3.5: (The relaxed configuration of an isolated succinate molecule on a Cu(110) 6x6 surface unit cell (a)Antonio's adsorption model (b)Darling's model

that of oxygen B and C. The carbon back bone in this configuration is about $26°$ from the [001] direction. A simulated STM image with the Tersoff-Hamann method is shown in Fig.3.6 (b). Surprisingly, the map of local density of states (LDOS) shows two protrusions for the succinate molecule. The exact locations of
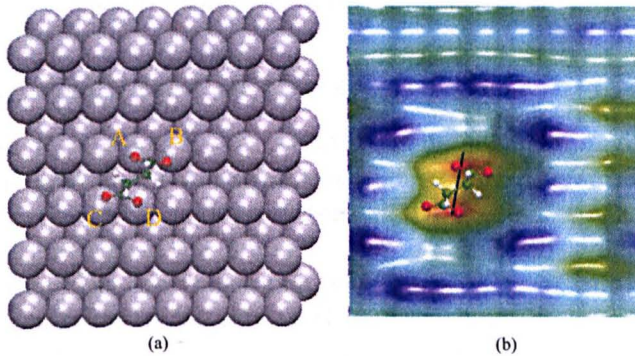

Figure 3.6: (a) The relaxed configuration of an isolated succinate molecule on a Cu(110)-8×6×4 unit cell (b) A simulated STM image with Tersoff-Hamann method, $V_{bias} = -0.21V$.

the two protrusions are found close to the oxygen A and D but shifted to the right (to oxygen A) or left (to oxygen B). The angle between the line that connects the two peaks in Fig.3.6 (b) and the [001] direction is about 9 degree. The large blue areas around the succinate molecule in LDOS map indicates that electron charge flows from the surface to the adsorbate; the adsorption of a molecule thus significantly changes the local electronic environment of the substrate.

Additionally, an analysis of the displacements of the top-most layer copper atoms
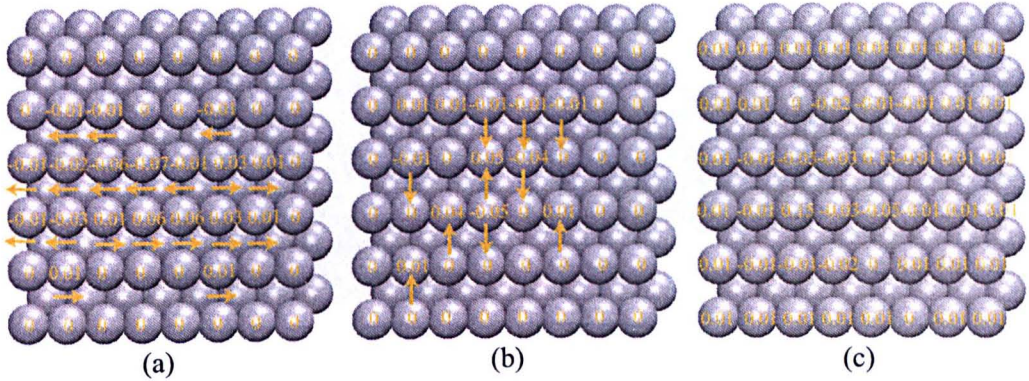
66

Figure 3.7: (a) The movement of the top-most layer copper atoms in the [110] direction, (b) the movement of the top-most layer copper atoms in the [001] direction, and (c) the movement of the top-most layer copper atoms in the z direction.

is shown in Fig. 3.7. As can be seen, upon adsorption of a succinate molecule the copper surface atoms undergo ionic movements in three dimensions. In the [001] direction and z direction, the movements of ions are significant only for the copper atoms directly bond to the oxygen. In the [110] direction, however, even the second nearest neighbor of the bonding sites give a displacement of 0.03Å.

3.1.5 Determination of the size of the Cu(110) unit cell

Once the ground state adsorption structure is determined, as discussed in 2.3.3, the next step is to evaluate the energies of higher order interacting structures, e.g. pairs, trios, etc. However, before a systematic DFT calculations, one needs to determine the suitable size of the unit cell. A proper unit cell should be large enough to not only contain the adsorbates in a particular coverage, but also minimize the forces on surface atoms at the boundaries. A too small unit cell can result in an error in the energy values. For instance, as shown in Fig. 3.8, unit cell (a) and (b) represent the same ordering of the adsorbates on the Cu(110) surfaces. All other calculating parameters, including the sampled k-mesh, are identical for these two structures. The only difference is the size of the unit cells. As shown by DFT calculations, the difference in the adsorption energy per molecule with respect to the same configuration is as much as 31meV. Another example is shown in Fig. 3.9. A succinate molecular chain along the [110] direction bas been separately constructed in two unit cells. The molecular structures before relaxations are completely identical. When the relaxation starts, however, the small unit cell does not have enough space to relax the surface atoms, the whole structure has therefore been broken, see Fig. 3.9(b). It seems that the molecular chain cannot be stabilized. By contrast, in the larger unit cell, the molecules are fully relaxed without significantly disturbing the lattice configuration of the copper surface. The DFT result for this unit cell shows that the molecular chain is one of the
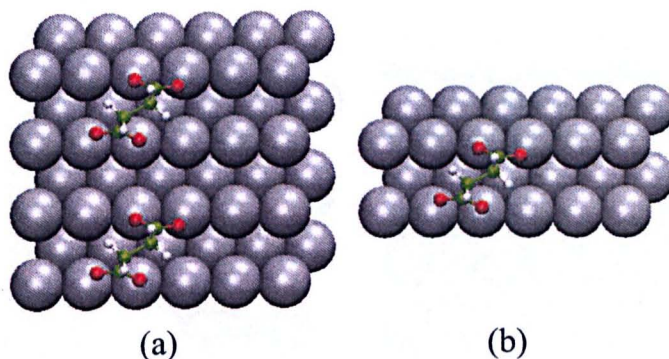
(a)   (b)

Figure 3.8: Unit cell (a) and (b) represent the same configuration of the succinate molecules on the Cu(110) surface. Due to the size difference of the unit cell, the binding energy per molecule in (a) and (b) differ by 31meV

stable structures on a Cu(110) surface.

Taking the computational efficiency and the size of the higher-order interacting structures investigated into account, the unit cell shown in Fig. 3.6(a) (20.59Å × 21.84Å × 16Å) has been chosen for all DFT calculations. Due to the limitation of computational resources, a systematic test of the maximum coverage for this unit cell was not possible. However, by estimating the surface area where the ionic movements occur due to the adsorption of one succinate molecule (see Fig. 3.7), such a unit cell should properly contain three adsorbates.

3.1.6 Lateral interactions

Pair Interactions

The energy parameters of all pair interactions in this work have been derived from the DFT results of two succinate molecules in one unit cell. All pair configurations taken into account are labeled by the relative positions of the two adsorbates. For instance, in Fig. 3.10(a), the two succinate molecules are separated by two lattice vectors in [110] direction, this structure is therefore referred to as "P2_0". In Fig. 3.10(b), the molecule on the left can be superimposed onto the other by moving two lattice vectors in [110] direction and then two lattice vectors in [001] direction, the configuration is therefore named as "P2_2". Following the same convention, Fig. 3.10(c) and (d) are called "P-1_2" and "P0_3", respectively. Since the pair interactions "Pi_j" and "P-i_-j" correspond to the same pair configuration, only one of them needs to be calculated with DFT.

For those configurations where the two adsorbates are very close to each other, the unit cell used for VASP calculations is regarded to be large enough to isolate the interactions from adjacent unit cells, see Fig. 3.11(a). Thus, the energy
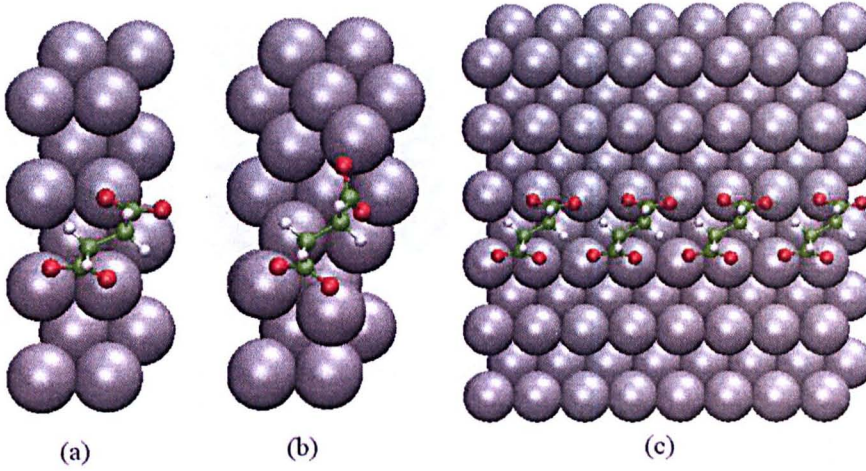
Figure 3.9: a) the initial configuration of one succinate in a Cu(110)-(2×4×4) unit cell. (b) the relaxed configuration of (a) with a DFT calculation. (c) the relaxed configuration of four succinate molecules in a row along the [110] direction of a Cu(110)-(8×6×4) unit cell

parameter of a pair interaction "Pi_j" is calculated with equation (3.4):

$$V_P(i, j) = \frac{1}{2}[E_{ads}(Pi\_j) - 2E_{ads}(\text{Isolate})] \tag{3.4}$$

Here, $E_{ads}(Pi\_j)$ denotes to the energy gained by the adsorption of two succinate molecules in pair structure Pi_j; while $E_{ads}(\text{Isolate})$ represents the binding energy of one succinate molecule in the unit cell 3.6(a). The adsorption energy of all pair configurations are evaluated by

$$E_{ads}(Pi\_j) = E(Pi\_j) + 2E(H_2) - E(\text{clean surface}) - 2E(\text{succinic acid}). \tag{3.5}$$

However, as the distance between two molecules in the unit cell increases, the intermolecular interactions between different unit cells become non-negligible. For instance, in Fig. 3.11(b), each succinate molecule equally interacts with two other adsorbates: one from its own unit cell, the other from an adjacent unit cell. Consequently, the pair-interaction energy of these configurations should also take the molecules from neighboring unit cells into account. As an illustration, in 3.11(b), the pair-interaction P0_3 is expressed by

$$V_P(0, 3) = \frac{1}{4}[E_{ads}(P0\_3) - 2E_{ads}(\text{Isolate})] \tag{3.6}$$

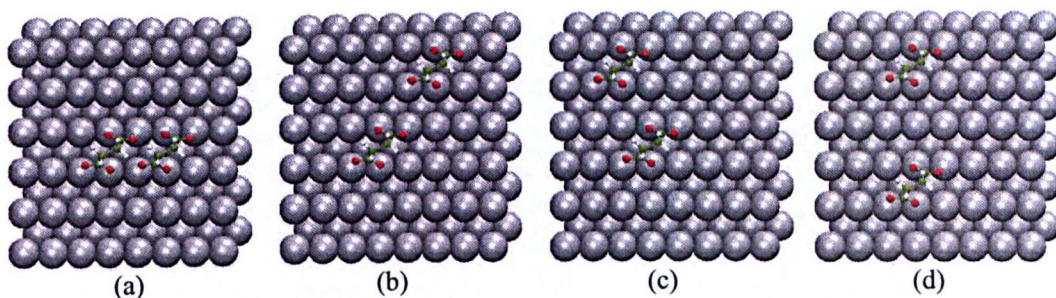Table 3.4 gives the pair-configurations calculated within the unit cell, and the values of their interaction energy.

Figure 3.10: Schematic representations of pair configuration (a) P2_0, (b) P2_2, (c) P-1_2 and (d) P0_3.

Generally, for chemisorbed organic molecules on metal surfaces, simple pair interactions can generally be divided into three classes.

- Direct interactions: when the separation distance of the two adsorbates is short, the wave functions of molecules may overlap to form chemical bonds.

- Indirect interactions: at short range, the adsorption of one molecule will change the electronic and ionic structure of the substrate. As a consequence, the first adsorbate will affect the adsorption of a second molecule in its vicinity.

- Nonlocal electrostatic effects: if the adsorbing species is charged, the electrostatic interactions between two adsorbates can be described by considering the molecules as charged points. If the adsorbates are neutral but the center of positive charges and negative charges are different, electrostatic effects are mainly dipole-dipole interactions.

As far as the succinate molecules on Cu(110) surfaces are concerned, Table 3.4 indicates a distance dependence of the pair interactions. As can be seen, all the strong interactions are confined with in an area shown in Fig. 3.12.

At low coverage, the succinate molecules do not overlap with each other on the Cu(110) surfaces. Thus, the only possible direct interactions between the adsorbates is hydrogen bonding. Having analyzed all the pair structures in Table 3.4, we found that only in configurations "P2_0" and "P1_-1", the shortest intermolecular oxygen-hydrogen distances are below five angstroms (2.53Å and 2.80 Å respectively). Correspondingly, the attractive hydrogen bonding in these two structures should be the strongest. However, as shown in Table 3.4, the DFT calculations indicate that the pair-interactions in "P2_0" and "P1_-1" are repulsive. As a matter of fact, these two configurations are the most unstable pair structures in the list. We thus conclude that the direct interactions do not play an important role in the ordering behavior of the succinate-Cu(110) system.
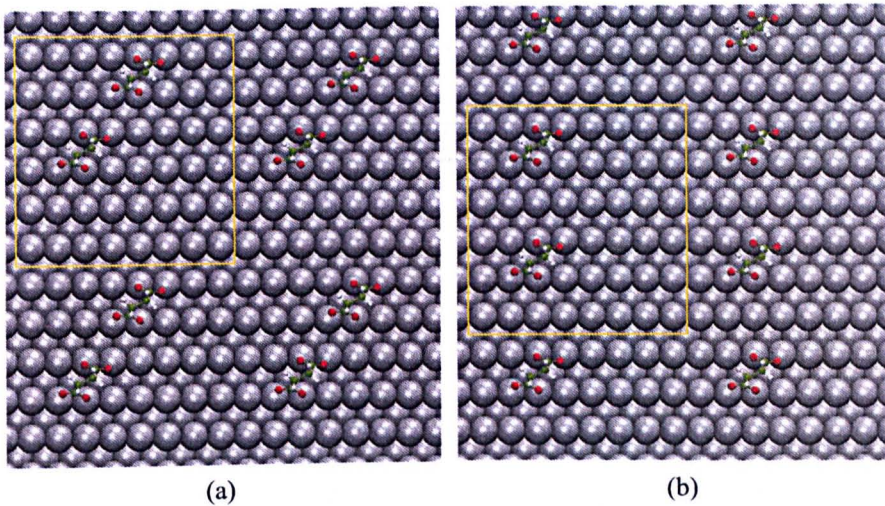
70

| (a) | (b) |

Figure 3.11: (a) the molecules in one unit cell are far away from the molecules in adjacent unit cells (b) the molecules in one unit cell interact with molecule in another unit cell

For the effects of indirect interactions, Fig. 3.6 and Fig. 3.7 showed very clearly that the adsorption of the first succinate molecule results in a significant change in the electronic structure and atomic positions around the binding sites. Of more importance, we found that the area of the local lattice distortions around an adsorption site (see Fig. 3.7) is very similar to the area of strong pair interactions highlighted in Fig. 3.12. We thus investigated the map of the local density of states for P1_-1 (a very repulsive pair interaction), P0_2 (the most attractive pair interaction) and P0_3 (one of the weakest pair interactions), see Fig. 3.13. As can be seen, the charge density contours for the P1_-1 and P0_2 are dramatically distorted due to molecular interactions. In the case of configuration P0_3, the local density of states around adsorbates are also different from that of Fig. 3.6 (b), but the symmetry of protrusions in Fig. 3.13 (f) is consistent with Fig. 3.13 (c). From these three examples, one can conclude that the strong interacting configurations are always associated with significant hybridizations of the local density of states between two adsorbates. The nature of the pair interactions can therefore be attributed to the indirect interactions.

The contributions of the nonlocal dipole-dipole interactions are estimated by an Electron Density Difference (EED) calculation [82]. When a molecule adsorbs on a surface, a charge transfer takes place at the interface. An EED analysis allows a visualization of this process.

1. Denote the charge density of the relaxed adsorbate-substrate system as CHG1.

2. Keep the molecular structure in the adsorbate-substrate system, remove all

Table 3.4: A list of the pair-configurations calculated with DFT and the values of their interaction parameters

Configurations	Adsorption energy(eV)	Pair-interaction parameters(eV)
P2_0	-2.641	0.097
P3_0	-2.817	0.009
P4_0	-2.796	0.010
P1_-1	-2.711	0.062
P2_-1	-2.870	-0.017
P4_1	-2.801	0.017
P3_1	-2.806	0.015
P0_2	-2.898	-0.032
P1_2	-2.877	-0.021
P-1_2	-2.886	-0.025
P2_2	-2.877	-0.021
P-2_2	-2.849	-0.007
P3_2	-2.872	-0.019
P0_3	-2.860	-0.006
P1_3	-2.848	-0.003
P2_3	-2.851	-0.004

surface atoms, and calculate the charge density CHG2.

3. Keep the surface atoms unchanged in the adsorbate-substrate system, remove the molecular structure, and calculate the charge density CHG3.

4. The EED can then be obtained by subtracting CHG2 and CHG3 from CHG1.

Fig. 3.14 shows the results of EED calculation for an isolated Succinate molecule on the Cu(110) surface. The surface unit cell is imaged in a transparent mode, so that the charge transfer inside the surface slab is visible. The blue color in Fig. 3.14 indicates the areas of charge depletion; while the yellow color shows areas of charge accumulation. As can be seen, in x-y dimensions (see Fig. 3.14 (a) and (c)), the centres of negative charge accumulations and positive charge accumulations completely overlap. However, since the centre of the blue area is lower than that of the yellow area in z direction (see Fig. 3.14 (b) and (d)), there is a net charge flow from the copper surface to the adsorbed molecule.

As illustrated in Fig. 3.15(b), a dipole moment is a vector that represents the arrangement of charges. Its magnitude is defined as
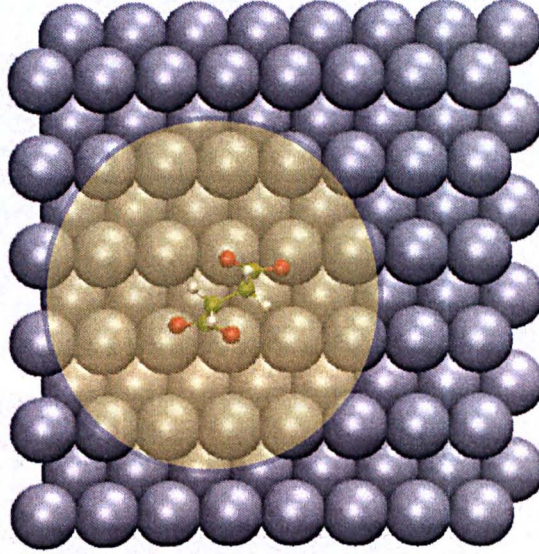
$$\mu = qR. \tag{3.7}$$

Figure 3.12: Only the pair configurations in which the center of the other succinate molecule is also in the highlighted area give strong pair interactions.

When two dipoles are parallel (as in Fig. 3.15(a)), the potential energy of two dipoles is expressed as below.

$$V_{dipole-dipole} = \frac{\mu_1 \mu_2}{4\pi\epsilon_o r^3}(1 - 3cos^2\theta) \tag{3.8}$$

where μ_1 and μ_2 are the dipole moments. In case of succinate molecules on Cu(110), the dipole moments should be perpendicular to the copper surface. The centers of dipoles are in a plane, which is parallel to the surface ($\theta = \pi/2$). As a result, the dipole-dipole interaction energy of a molecule with its neighbors can be expressed by [83, 84]

$$E_{dipole-dipole} = \frac{2\mu^2}{4\pi\epsilon_o} \sum_i \frac{1}{r_i^3} \tag{3.9}$$

However, since the positive charge center and negative charge center of each dipole are close to each other, the dipole moment should be rather small. In addition, the dipole-dipole separations on Cu(110) are quite large, even the shortest distance is as large as 5.14 Å. As a consequence, the dipole-dipole interactions (equation (3.9)) can be neglected.

Trio interactions

Compared with the pair configurations, the trio structures occupy a much larger area on the Cu(110) surface. Due to the size limitation of the unit cell, only the condensed trio interactions have been calculated by DFT. Trio configurations are
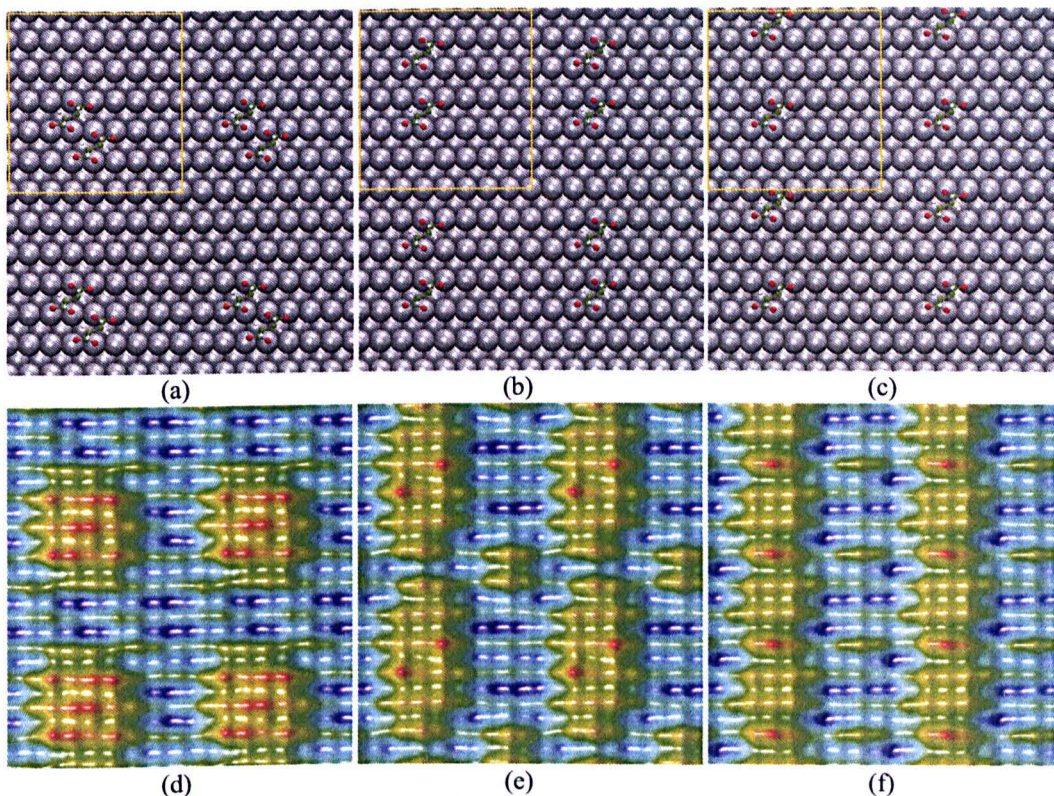
Figure 3.13: The atomic positions of the structures of P1_-1, P0_2, P0_3 and their local density of states. $V_{bias} = -0.21V$.

also characterized by the relative positions of the adsorbates. For instance, as shown in Fig. 3.16 (a), the succinate molecule in the middle is separated from the other two by two lattice vectors in the [110] and -[110] direction, respectively. This configuration is therefore called "T20_-20". Following the same naming rule, the other two structures in Fig. 3.16 are referred to as "T20_22" and "T2-1_-21" respectively. Nevertheless, in the trio structures, the starting point can be any of the three molecules. Consequently, every trio configuration can have three different names. Taking Fig. 3.16 (a) as an illustration, the other two names of this structure are "T20_40" and "T-20_-40" if the starting point is the left or right molecule, respectively. In the following, only one of the three names is used to refer to a particular trio configuration.

In addition, since the Cu(110) surface exhibits an atomic arrangement of C2 symmetry, the trio configuration "Tij_pq" represent the same structure as "T-i-j_-p-q". For instance, see Fig. 3.17, configuration T-1-2_2-1 is the same as T12_-21 after a rotation of 180° along the [110] direction. Symmetry significantly reduces the number of structures to be calculated with DFT.
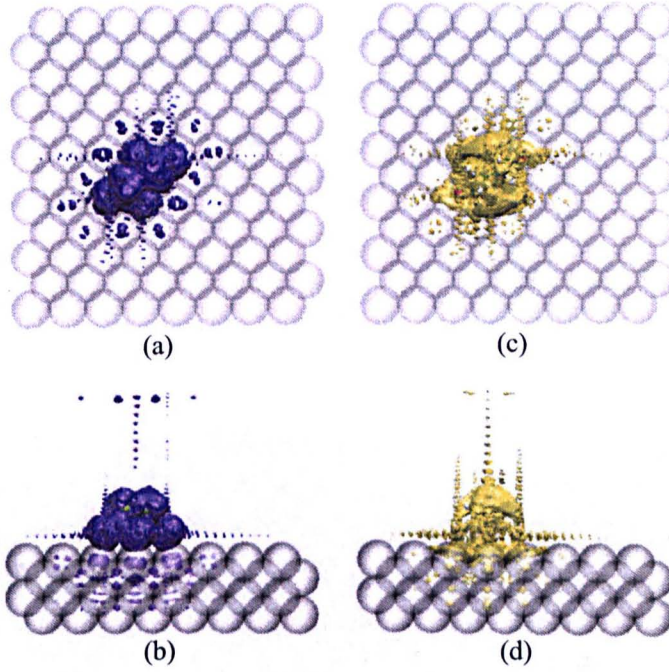
Figure 3.14: Electron density difference plot (EDD) for one succinate molecule on the Cu(110) surface. (a) Top view of the isosurface where the charge difference is 0.01e. (b) Side view of the isosurface where the charge difference is 0.01e. (c) Top view of the isosurface where the charge difference is -0.01e. (d) Side view of the isosurface where the charge difference is -0.01e.

The adsorption energy for each trio configuration is calculated with Equation 3.10 :

$$E_{ads}(Tij\_pq) = E(Pij\_pq) + 3E(H_2) - E(\text{clean surface}) - 3E(\text{succinic acid}).$$
(3.10)

The trio interaction energy is evaluated by :

$$V_T(ij\_pq) = \frac{1}{3}\left[E_{ads}(Tij\_pq) - 3E_{ads}(\text{Isolate}) - \text{pair interactions}\right]$$
(3.11)

We take the trio configuration T20_22 as an illustration, see Fig. 3.18. The unit cell is highlighted by a yellow frame, the three molecules are labeled by A, B, and C, respectively. Since the molecules in adjacent unit cells are far away from molecule A, B and C, the pair interactions included in this structure are P2_0 (A ↔ B), P2_2 (A ↔ C) and P0_2 (B ↔ C). Substituting the pair interactions into Equation (3.11), the trio interaction T20_22 is given by :

$$V_T(20\_22) = \frac{1}{3}\left[E_{ads}(T20\_22) - 3E_{ads}(\text{Isolate}) - 2V_P(2,0) - 2V_P(2,2) - 2V_P(0,2)\right].$$
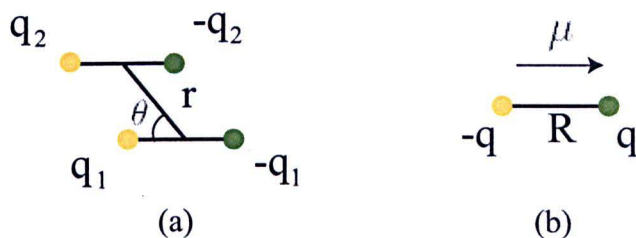(3.12)

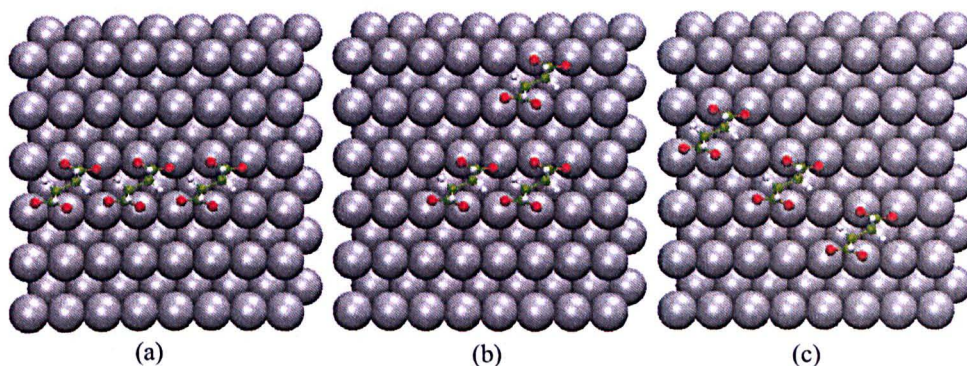Figure 3.15: Schematic diagram representing (a) two paralleled dipoles and (b) the dipole moment.

Figure 3.16: Schematic representations of trio interaction (a) T20_-20, (b) T20_22 and (c) T2-1_-21.

In total, 33 trio interactions have been evaluated with the method described above. Their adsorption energies and the values of interactions are listed in Table 3.5. Comparing the numbers in Table 3.4 with those in Table 3.5, we see that the average value of trio interactions is only about a third of the average value of pair interactions. In addition, the magnitude of most trio interactions are in the range of 10 meV. This suggests that the remaining higher-order configuration parameters should have even lower energy values. Thus, the expansion of all lattice configurations with pair interactions and small trio interactions is justified.

3.2   Derivation of the lateral energy parameters

The multi-scale sampling approach described in section 2.3 serves as a connection between the macroscopic thermodynamics and microscopic DFT calculations. The quality of this connection, however, is prominently dependent on the derived Lattice Gas Hamiltonian. In principle, a qualified LGH should be able to

Table 3.5: A list of the tio configurations calculated with DFT and the values of their interaction parameters

| Configurations	Adsorption energy(eV)	Trio interaction parameters(eV)
T20_-20	-3.816	0.003
T02_0-2	-4.457	-0.002
T22_-2-2	-4.367	-0.010
T2-2_-22	-4.314	-0.011
T12_-1-2	-4.339	0.004
T1-2_-12	-4.389	0.002
T2-1_-21	-4.327	-0.002
T20_02	-4.118	0.006
T12_-12	-4.118	0.011
T32_12	-4.112	0.009
T-32_-12	-4.102	0.003
T1-1_2-2	-3.942	0.025
T31_-3-1	-4.198	0.003
T32_-3-2	-4.411	-0.014
T20_22	-4.129	0.012
T20_42	-4.114	-0.004
T20_52	-4.083	0.004
T02_1-1	-4.206	-0.001
T02_2-1	-4.352	0.005
T02_-3-1	-4.319	-0.011
T02_30	-4.291	0.002
T02_-30	-4.320	0.004
T02_31	-4.270	0.005
T1-1_-2-2	-4.108	0.010
T1-1_-3-1	-4.068	-0.005
T1-1_12	-4.190	0.001
T1-1_22	-4.172	0.003
T12_-2-2	-4.320	0.006
T12_-3-1	-4.324	-0.019
T12_-21	-4.278	0.007
T31_-2-2	-4.309	-0.014
T32_51	-4.298	0.009
T32_2-1	-4.331	0.002
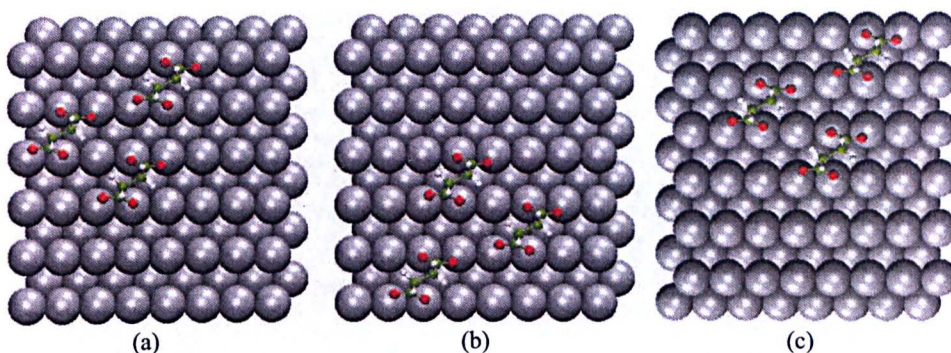
Figure 3.17: Schematic representations of trio interaction (a) T12_-21, (b) T-1-2_2-1 and (c) T-1-2_2-1 rotated by 180°.

accurately evaluate the most important interactions governing the phase behavior and the configurational entropy on phase stability of the system. Only a LGH of this kind is regarded to have the powder of prediction, and can be used to evaluate the new configurations different from those in the selection. Nevertheless, a standard process of deriving lateral interaction parameters from first principles calculations remains to date a delicate task.

3.2.1 The systematic approach

The detailed description of the systematic derivation of an optimal LGH is included in Appendix A. Here, only the basic principles of the method are introduced. First of all, the database of the configurations directly investigated with first-principles calculations is denoted as a data space. This data space is then partitioned into two subspaces: "training" and "test". The configurations in subspace "training" are used to derive a LGH, while the subspace "test" is used to evaluate the predictive property of that LGH. For each partitioning form, there will be a LGH derived from the "training" subspace. As listed in Tables 3.4 and 3.5, there are in total 49 lateral interactions calculated with DFT. The number of overall partitioning of this DFT database is therefore 2^{49}, because each configuration can either in the "training" or the "test" subspace. Since this number is astronomically large, a function optimizer is needed.

In general, the search of a qualified LGH from a given DFT database can be divided as three subtasks:

- Derive a LGH for any given "training" subspace.

- Evaluate a LGH with the corresponding "test" subspace and assign a "fitness value" to each LGH, so that the quality of different LGHs can be compared.

- A "training" subspace generator, which enables one to improve the quality of the derived LGH.
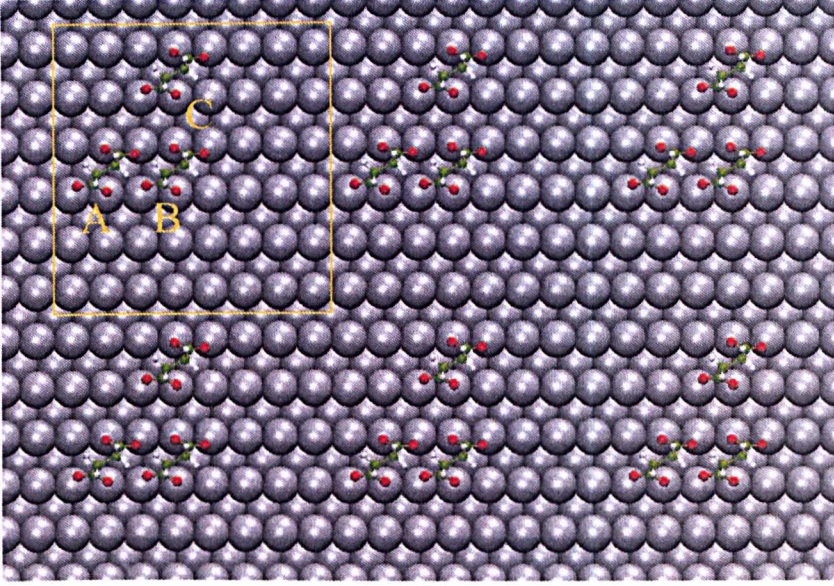
Figure 3.18: The schematic representation of trio interaction T20_22

For the first subtask, the energy parameters of a LGH are not derived using Equation 3.4 nor 3.11. Instead, each of the configurations in the given "training" subspace is expressed with the lattice gas hamiltonian.

$$\sum V_o + \sum V_{P_1} + \sum V_{P_2} + \cdots \sum V_{P_m} + \sum V_{T_1} + \sum V_{T_2} + \cdots + \sum V_{T_n} = E_{ads}$$
(3.13)

Here, V_o is denoted to the adsorption energy for an isolated molecule on a metal surface; while m and n represent the number of pair and trio interaction parameters, respectively.

Substituting Equation 3.13 into all interacting configurations contained in a "training" subspace, we obtain a large system of linear equations.

$$
\begin{aligned}
A_o V_o + A_1 V_{P_1} + A_2 V_{P_2} + \cdots A_m V_{P_m} + A_{m+1} V_{T_1} + A_{m+2} V_{T_2} + \cdots + A_{m+n} V_{T_N} &= E_{ads}(1) \\
B_o V_o + B_1 V_{P_1} + B_2 V_{P_2} + \cdots B_m V_{P_m} + B_{m+1} V_{T_1} + B_{m+2} V_{T_2} + \cdots + B_{m+n} V_{T_N} &= E_{ads}(2) \\
C_o V_o + C_1 V_{P_1} + C_2 V_{P_2} + \cdots C_m V_{P_m} + C_{m+1} V_{T_1} + C_{m+2} V_{T_2} + \cdots + C_{m+n} V_{T_N} &= E_{ads}(3) \\
&\vdots \\
Q_o V_o + Q_1 V_{P_1} + Q_2 V_{P_2} + \cdots Q_m V_{P_m} + Q_{m+1} V_{T_1} + Q_{m+2} V_{T_2} + \cdots + Q_{m+n} V_{T_N} &= E_{ads}(N)
\end{aligned}
$$
(3.14)

Here, "N" is the number of configurations included in the given "training" database. The system (3.14) can also be expressed in matrix notation:

$$AV = E.$$
(3.15)

Here, A is the $N \times (m + n + 1)$ coefficient matrix; while V and E are column vectors representing the unknowns and energetics, respectively.

$$\begin{bmatrix} A_o & A_1 & \cdots & A_{m+n} \\ B_o & B_1 & \cdots & B_{m+n} \\ \vdots & \vdots & \ddots & \vdots \\ Q_o & Q_1 & \cdots & Q_{m+n} \end{bmatrix} \begin{bmatrix} V_o \\ V_{P_1} \\ \vdots \\ V_{T_N} \end{bmatrix} = \begin{bmatrix} E_{ads}(1) \\ E_{ads}(2) \\ \vdots \\ E_{ads}(N) \end{bmatrix}$$

Since the elements included in matrix A and column vector E are all known, the matrix Equation (3.15) can easily be solved or best estimated with linear algebra methods (Appendix A).

When a LGH has been derived from a "training" database, its power of prediction is evaluated by a cross-validation (CV) error.

$$CV^2 = \frac{1}{M} \sum_{i=1}^{M} (E_{DFT}^i - E_{CE}^i)^2 \tag{3.16}$$

In Equation (3.16), M is the number of configurations in the whole DFT database, E_{DFT}^i is the energy of ith configuration obtained form direct first-principles calculations, while E_{CE}^i represents the energy obtained from "cluster expansion" with that lattice gas hamiltonian. The LGH with the lowest CV value is associated with the best predictive property.

For the last subtask, genetic algorithms (GA) are widely used for parameter optimizations. A typical implementation of the GA evolution in this kind of problem consists of five steps.

1. Randomly select a number of "training" subspaces, and encode each of them with a string of binary numbers with a fixed length of "L". The collection of these binary strings are referred to as the "first generation".

2. Since each string represents a particular "training " subspace, it can be used to derive a LGH.

3. The LGHs obtained from all strings are then evaluated by the cross-validation method so that the predictive property of the strings can be measured from the corresponding CV errors.

4. However, in genetic algorithms, the quality of strings need to be expressed by "fitness values", which measure how well the strings can solve the target problem. Of great importance, the value of fitness of a string has to be defined with respect to all other members of the current generation.

5. The new generation can then be produced by "mutation" and "crossover" from the current generation. However, one needs to note that the reproductive opportunity allocated to each string is not identical but proportional to its fitness value. Thus, the strings associated with higher "fitness values" have a better chance to produce "children" in the next generation.

6. Successively repeating step 2 to step 5 for a large number of generations, the quality of the obtained strings is expected to converge. The LGHs derived from the last generation are supposed to have reliable predictive power.

3.2.2 The non-systematic approach

However, the implementation of a systematic derivation described in 3.2.1 and Appendix A requires a large number of configurations, which have been investigated by direct DFT calculations. In case of succinate molecule on a Cu(110) surface, the number of configurations that can be properly evaluated is significantly reduced by the size of the unit cell. As a consequence, there are not enough configurations for a systematic derivation. Under this circumstance, however, all the configurations included in DFT database are used to derive the lattice gas hamiltonian. The method used for mapping DFT configurational energies onto LGH energy parameters has been described in section 3.1.6. Since we totally have 50 configurations evaluated with DFT, there are also 50 energy parameters contained in the derived LGH (see Tables 3.4 and 3.5).
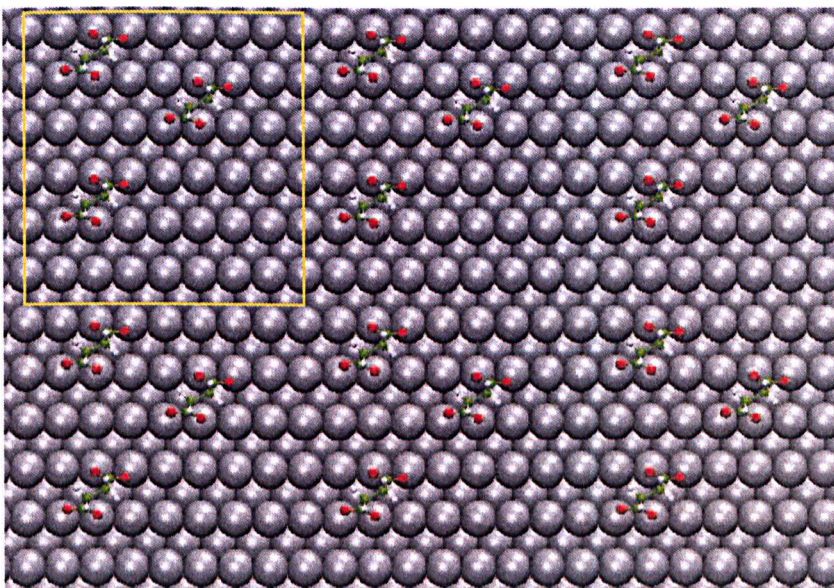

Figure 3.19: The schematic representation of trio interaction T03_32. The adsorption energy per unit cell is -4.320eV from direct DFT calculations and -4.314eV from the LGH prediction

In order to have a general idea about the predictive power of the generated LGH, we performed one more DFT calculation as the "test set", see Fig. 3.19. Surprisingly, the energy difference between the DFT result and cluster expansion prediction is only $6meV$. Although the accuracy of predictions for other configuration may not be as good as this one, it can be expected that the expansion for

other lattice configurations with this LGH will retain the general DFT accuracy.

Chapter 4

Thermodynamic properties of Succinate-Cu(110)

So far, the investigations have been focused on the zero temperature energy levels of the succinate-Cu(110) system. However, as stressed at the beginning of Section 2.3, only a statistical treatment of the obtained energy levels can result in a picture of the macroscopic physical properties of the system. In general, atoms or molecules chemically absorbed on a single crystal substrate tend to form some distinct two dimensional periodic structures which are uniform throughout the surface. This tendency is caused by a collective interplay of all elementary processes between the adsorbates. The formed ordered structures are referred to as the surface "phases". When a surface is in one of its phases, the electronic and structural properties of the surfaces are very well defined. Thus, the identification of these phases and the characterization of their chemical properties are of great importance to surface science and heterogeneous catalysis. If a phase is found to be particularly suitable for certain reactions, the external conditions can be tuned so that this phase can be stabilized during reactions. For the adsorbate-substrate systems, the stability of a surface state depends on the temperature T and the coverage Θ of chemisorbed species. One of the most succinct ways of presenting the changes of phases, a system can undergo, is its phase diagram, see Fig. 4.1.

4.1 Thermodynamic stability of surface phases

As can be seen in Fig. 4.1, a phase diagram shows the most stable state of a system for all combinations of coverage and temperatures. In thermodynamics, the stability of states is expressed by the direction of "spontaneous changes", i.e. the changes that do not require work to be done to bring them about. According to the Second Law of thermodynamics, the state function that identifies a spontaneous change is entropy: the entropy of an isolated system increases in the course of a spontaneous change.

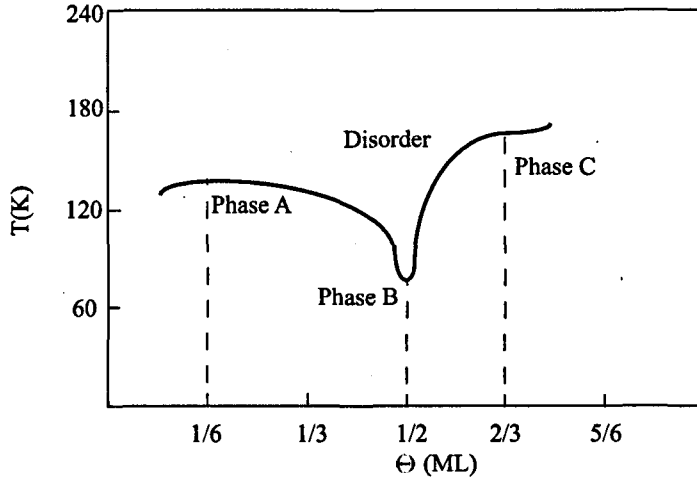$$\Delta S_{total} > 0 \qquad (4.1)$$

Figure 4.1: Schematic representation of an order-disorder phase diagram.

Here, ΔS_{total} represents the total entropy change of the system and the environment. Thus, for any spontaneous processes, we have

$$dS \geq \frac{dq}{T}, \tag{4.2}$$

where dq is the heat transferred between the system and environment during the changes. Equation (4.2) can be rewritten as

$$dq - TdS \leq 0 \tag{4.3}$$

Considering that a thermodynamic change occurs at constant volume and in the absence of non-expansion work, we can write $dq = dU$. As a consequence,

$$dU - TdS \leq 0. \tag{4.4}$$

If the condition of constant volume is replaced by constant pressure, then we get $dq = dH$. As a result,

$$dH - TdS \leq 0. \tag{4.5}$$

In thermodynamics, $U - TS$ is defined as the Helmholtz free energy F, while $H - TS$ is referred to as the Gibbs energy G. Consequently, the direction for thermodynamically spontaneous processes can be expressed by Equation (4.6).

$$\begin{cases} \Delta F \leq 0 & \text{constant volume} \\ \Delta G \leq 0 & \text{constant pressure} \end{cases} \tag{4.6}$$

For the succinate-Cu(110) system, the thermal expansion of the copper surface is negligibly small. The stability of the surface states are therefore dependent on their Helmholtz free energies.

84

At the temperature of $0K$, the Helmholtz free energy of a surface state is simply the internal energy $U(T = 0K)$. As the temperature increases, however, the Helmholtz free energy F is distributed into both the internal energy U(T) and the entropy S(T):

$$F = U - TS. \tag{4.7}$$

Since different thermodynamic states partition their free energy in different ways, one phase can be replaced by another phase if the external condition is properly changed. This kind of thermodynamic processes are the so called "phase transitions".

4.2   Phase transitions in two dimensions

In elementary physics, a phase is a set of states in which the chemical composition and physical properties are relatively uniform throughout the system. These phases correspond to the local minima of the free energy space. Under certain external conditions, the phases change. One of the classic examples of phase transitions is water. At 1 atm, ice is the most stable phase below $273.15K$. As the temperature goes up, liquid water becomes more stable. Above $373.15K$, all water molecules are in a gas phase. For molecules chemically adsorbed on the surfaces, the phases are commonly well defined two dimensional periodic structures. Correspondingly, a phase transition for these systems is a change of the two dimensional pattern.

Since the thermodynamic properties of each phase are encoded in the free energy (Section 2.3.2), a phase transition occurs when there is a singularity in the free energy or one of its derivatives. If there is a finite discontinuity in one or more of the first derivatives of the appropriate thermodynamic potential, the transition is termed first-order. Transitions of this kind are characterized by the coexistence of two phases at the critical temperature T_C. The most common examples are melting, freezing and vaporizing. If the first derivatives are continuous but the second derivatives are discontinuous or infinite, the transition is described as higher order, continuous, or critical. A transition of this kind is characterized by a divergent susceptibility, an infinite correlation length, and a power law decay of correlations [85].

When a phase transition occurs, it is often accompanied by a sharp change in the properties of the system. The properties used to characterize the whole phase transition process are called "order parameter". For different physical systems, the order parameter can be defined differently. For instance, in a liquid-gas transition, it can be the difference in densities between liquid and gas phases at the transition. In a ferromagnet, it can be the homogeneous magnetization.

In two dimensional surface structures, the order parameter most commonly used is the LEED diffraction intensity [86]. Fig. 4.2 shows a typical order-disorder transition of the surface structures. At low temperature, the value of the order
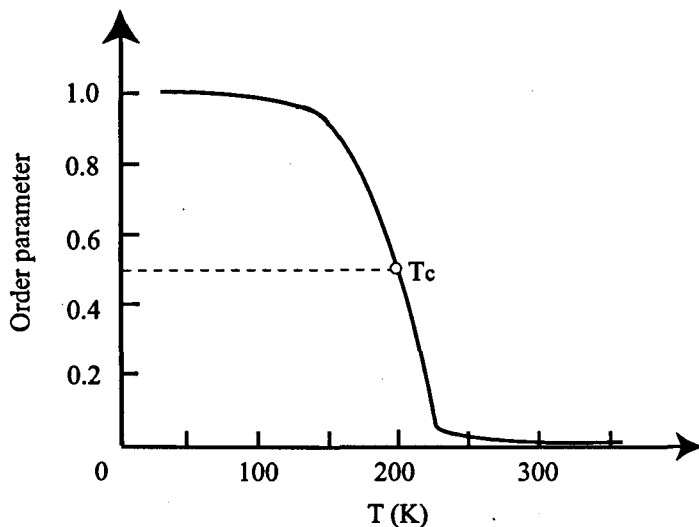
Figure 4.2: Schematic representation of an order-disorder phase transition at a constant coverage. The LEED diffraction intensity is plotted as a function of temperature.

parameter is defined to be 1. As the temperature increases to about $120K$, the original ordered phase starts to become disordered. Correspondingly, we see a sharp drop of the order parameter during $150K$ - $220K$. Above $250K$, the system is completely disordered. In a phase transition curve, the critical temperature T_C is denoted to the temperature, where the value of the order parameter is 0.5. The physical meaning of T_C is the temperature at which the two phases are in equilibrium and one phase is about to be replaced by the another.

In addition to the LEED diffraction intensity, the phase transition can also be characterized by an singularity of the heat capacity of the system. At constant volume, the thermodynamic definition of heat capacity for an isolated system is

$$C_V = \left(\frac{\partial U}{\partial T} \right)_{N,V} \tag{4.8}$$

During phase transitions, the value of C_V should significantly increase around the critical temperature, because the heat transferred to the system is mainly used for driving the change of phase rather than being stored as the internal energy. In the macroscopic scale, however, the results given by Equation (4.8) are not reliable because the determination of U is neither easy nor very accurate. Nevertheless, through statistical mechanics, the heat capacity can be related to

the spontaneous energy fluctuations of the system.

$$\left\langle (E- <E>)^2 \right\rangle = <E^2> - <E>^2$$
$$= -\left(\frac{\partial U(\beta)}{\partial \beta}\right)_V$$
$$= k_B T^2 C_V \qquad (4.9)$$

Here, $\beta = 1/K_B T$. Equation (4.9) opens a remarkable connection between the microscopic energy levels and the macroscopic state variables. Since the critical temperature of a phase transition is independent of the construction of the order parameter, both the LEED diffraction intensity and C_V can be used as order parameters for a Monte Carlo simulation.

4.3 Metropolis Monte Carlo Simulations

Once the Lattice Gas Hamiltonian has been determined, the energy for any lattice configuration can be predicted with the effective cluster expansion method. The quality of the phase diagram is then dependent on the method of numerical simulations. For molecules chemically adsorbed on surfaces a canonical ensemble is used for the numerical sampling, since experimental results are obtained at the conditions of ultra high vacuum and constant temperature. Correspondingly, the thermodynamic properties of the system are encoded in the partition function of the canonical ensemble

$$Z = \sum_i e^{-E_i/K_B T}. \qquad (4.10)$$

The important connection between the partition function Z and thermodynamic state variables is via the Helmoholtz free energy.

$$F = -K_B T ln Z \qquad (4.11)$$

If the value of the partition function is known, the probability that the system is in state i is given by the Boltzman distribution.

$$P_i = \frac{1}{Z} e^{-E_i/K_B T} \qquad (4.12)$$

The expectation value of any macroscopic property can then be calculated by

$$<A> = \sum_i P_i A_i. \qquad (4.13)$$

However, since only a small part of the state space are important in determining the expectation values of macroscopic state variables, Equation (4.13) can be approximately replaced by

$$<A> \approx A_M = \frac{1}{M} \sum_i^M P_i A_i \qquad (4.14)$$

in which most of the selected states are the "important" ones. As has been extensively discussed in Section 2.3, the Metropolis Monte Carlo method would enable a random but importance weighted walk in the whole configurational space. Given long enough "time", the sampled states should be able to construct an effective ensemble. The practical implementation of the Metropolis Monte Carlo algorithm is illustrated in Appendix B.

The main unit cell which is used for Monte Carlo simulations is Cu(110)-50×50. Since one succinate molecule occupies four copper surface atoms, the maximum number of adsorbates this unit cell can contain is 625. In this work, coverage of a surface is defined as

$$\Theta = \frac{\text{number of molecules}}{\text{number of surface atoms}}. \tag{4.15}$$

Therefore, the most condensed monolayer of adsorbates corresponds to a coverage of 25%. For all simulated systems, the initial configurations are randomly generated so that they are independent from each other. In total, six values of coverage have been selected for the ground state search. For each configura-

Ground state search

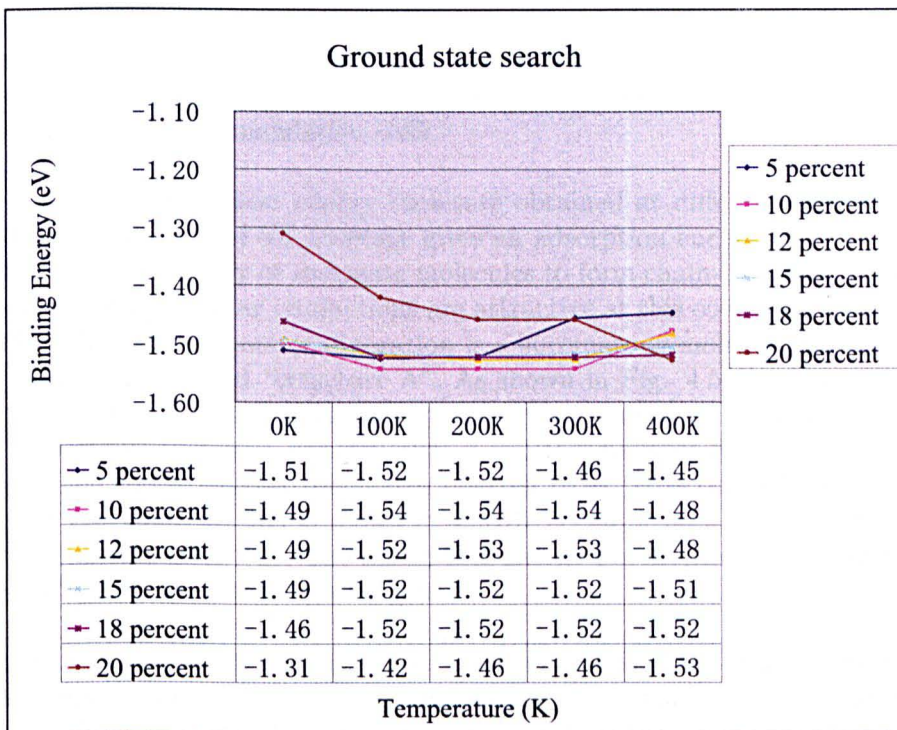|	0K	100K	200K	300K	400K
5 percent	-1.51	-1.52	-1.52	-1.46	-1.45
10 percent	-1.49	-1.54	-1.54	-1.54	-1.48
12 percent	-1.49	-1.52	-1.53	-1.53	-1.48
15 percent	-1.49	-1.52	-1.52	-1.52	-1.51
18 percent	-1.46	-1.52	-1.52	-1.52	-1.52
20 percent	-1.31	-1.42	-1.46	-1.46	-1.53

Figure 4.3: A plot of binding energy per molecule against temperature for 5, 10, 12, 15, 18 and 20 percent coverage configurations as the temperature increase from 0K to 400K.

tion, the growth of Markov chain stops when the averaged energy value for the

last 9 million "Monte Carlo steps" is independent of the "Monte Carlo time". In principle, the configurations obtained at $0K$ correspond to the ground state of the system. When the system is heated, more and more thermal motion of the molecules come into play, so that an decrease of binding energy is expected. However, as shown in Fig. 4.3, at $0K$ and $100K$ the binding energy per molecule is generally less than the values at $200K$. This is because at low temperatures, the acceptance probabilities in a Metropolis Monte Carlo simulation are so small that virtually no "movements" of molecules occurs for a very long "time" [88]. When the system is trapped in one of the local minima, it may never find its way to the ground state. For the systems with the coverage value of 5%, 10%, 12%, 15%, 18%, the lowest energy configuration occurs above $100K$. By contrast, the configuration of 20% coverage does not reach its ground state until the temperature increases to $400K$. This is also a common phenomena in Monte Carlo simulations, because 20% is very close to the maximum coverage value. Although the molecules repel each other due to the interactions, the number of empty sites is so limited that they experience an energy barrier to reach the ground state.

In order to make sure that the average energy values for each coverage at all temperatures are independent on their initial configurations or the size of the unit cell, 40 × 40 unit cell is used for the same simulations. We found that the energy values obtained from these two unit cell are very consistent. This means the results shown in Fig. 4.3, are reliable not only for the structures we simulated but also for any other simulation cells.

Fig. 4.4 shows the lowest energy structure obtained at different temperatures. At $0K$, the structure of 5% coverage gives an adsorption energy of $-1.51eV$ per molecule. The tendency of succinate molecules to form chain-like structures indicates that intermolecular interactions are attractive at this coverage. From $100K$ to $300K$ the ground state configuration is a succinate monolayer. In this work, this monolayer is called "structure A". As shown in Fig. 4.5(a), this monolayer corresponds to a coverage of 10 percent and gives a symmetry of (5 0, 3 2). When the temperature increases, the structure A vanishes between $300K$ to $400K$. The surface then shows another monolayer referred to as "structure B". An enlarged image of this configuration is shown in 4.5(b). It can be seen that structure B is of the symmetry (5 0, 3 1). The adsorption energy per molecule of this structure is 16 meV lower than that of structure A.

Comparing the three structures in Fig. 4.4, we found that the growth direction of some chain-like structures at 5% are very similar to the two monolayer structures. It seems that the molecules attempt to form well defined molecular chains, but the chain growth is disturbed by configurational interactions. In order to find the most stable chain structure in the 5% coverage configuration, the effective cluster expansion is performed for the configurations in Fig. 4.6. According to the simulated results, the binding energy per molecule for the structures in Fig. 4.6 are -1.496 eV, -1.379 eV, -1.490 eV and -1.398 respectively. They are all higher than the value obtained from the structure search, see Fig. 4.3. We thus con-
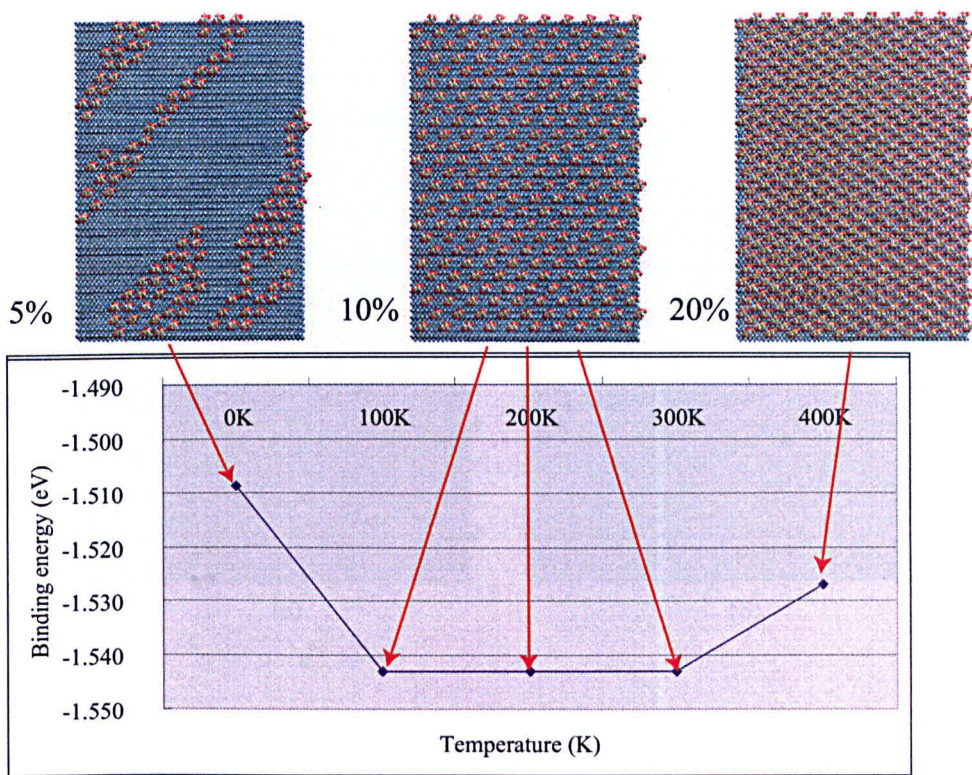
Figure 4.4: The lowest energy configurations of succinate molecules on the Cu(110) surface at different temperatures.

clude that the ground state at low coverage is very unlikely to be one-dimensional molecular chains. However, we also notice that some of the chain-like features show a tendency of 2D ordering (see Fig. 4.4). Thus, these ordered patterns were also evaluated by effective cluster expansion (see Fig. 4.6). However, neither of the configurations results in a higher binding energy than that obtained from structure search. Summing up all calculations, we conclude that the ground state structure for succinate-Cu(110) system at very low coverage is (or very close to) the configuration shown in Fig. 4.4.

As the coverage increase from 5% to 10%, the features of "structure A" first become manifest and then expand over the whole surface. Since this structure is associated with the lowest adsorption energy in our ground state search, it should be the configuration measured in the STM experiments. The thermodynamic stability of this configuration has also been evaluated with averaged binding energy and heat capacity, see Fig. 4.8. The critical temperature obtained from these two parameters are $390K$ and $430K$, respectively. Since the average energy is not as accurate as the heat capacity for determining critical temperatures, this small difference is acceptable.
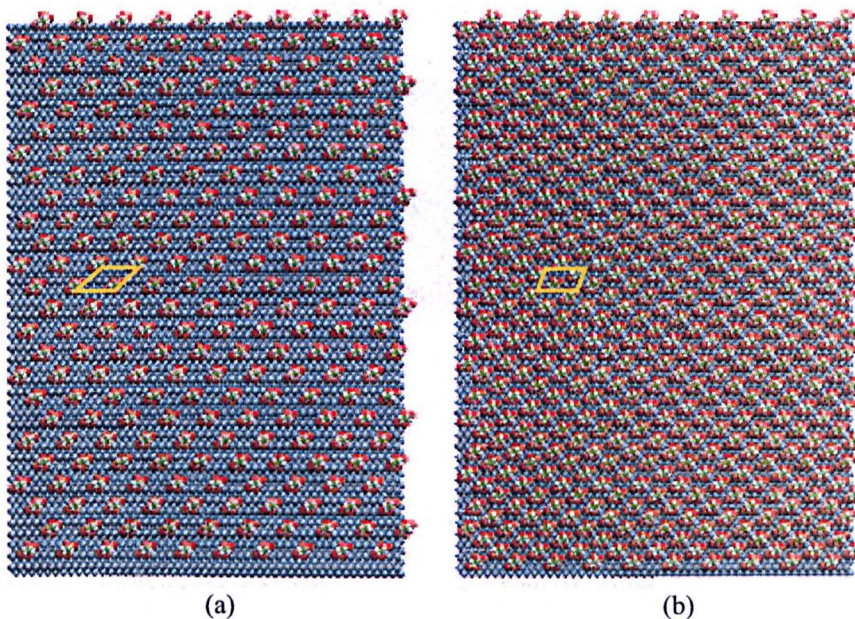
Figure 4.5: The enlarged structure of (a) monolayer of "structure A". (b) monolayer of "structure B".

When the coverage of succinate molecules exceed the threshold of 10%, "structure B" occurs. The ground state configuration in this case is a coexistence of the two structures. However, as the temperature is increased to around $400K$, "structure A" disappears. The surface is dominated by the islands of "structure B" only. Heating up the surface further, all ordered structures disappear, see Fig. 4.9. The critical temperature of the order-disorder transition depends on the coverage of molecules. Once a monolayer of "structure B" is formed, the molecules cannot "move" any more, because there are no "effective" sites left on the surface. As a consequence, the critical temperature tends to infinite.

In summary, with the standard Metropolis Monte Carlo simulations, we performed a ground state search for various coverages and temperatures. Two periodic ordered structures have been identified. The corresponding coverage values for these two structures are 10% and 20%, respectively. The most stable configuration is "structure A", which is associated with a (5 0, 3 2) symmetry. The adsorption energy of this molecular monolayer is $-1.543eV$ per molecule. At around $400K$, the (5 0, 3 2) monolayer undergoes a phase transition. The second monolayer structure is of (5 0, 3 1) symmetry, the adsorption energy per molecule for this configuration is 16meV lower than that of "structure A". Between the coverage 10% and 20%, the ground state is the coexistence of the (5 0, 3 2) and (5 0, 3 1) structures.
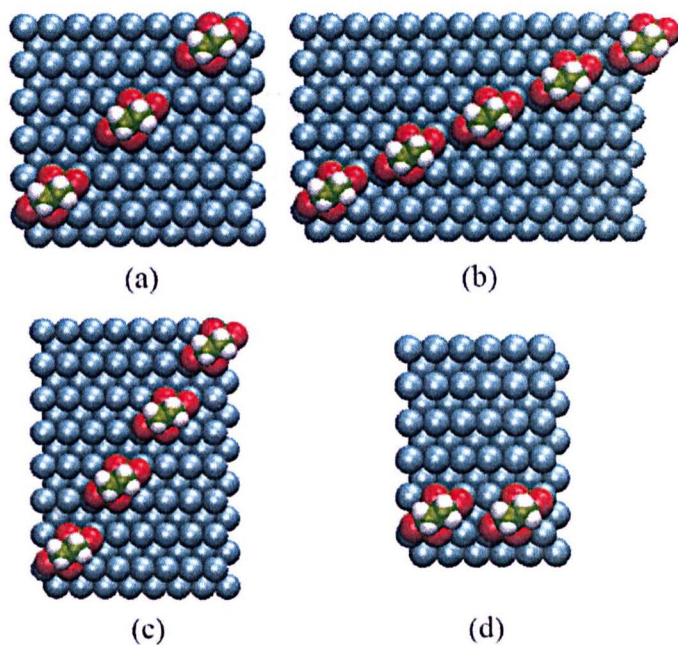
Figure 4.6: The molecular chain structures showed up in the 5% coverage configurations.
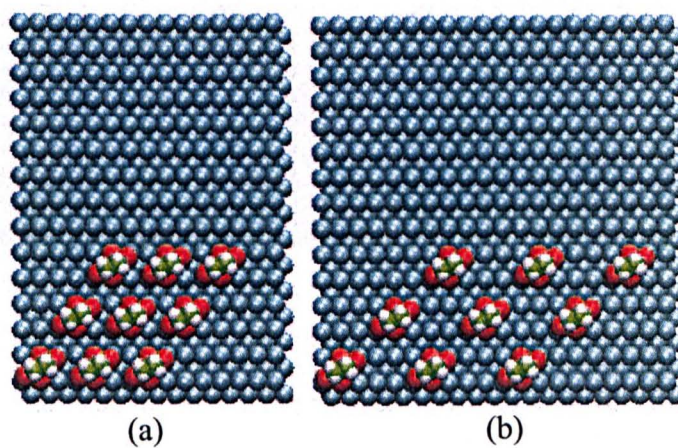

Figure 4.7: Two dimensional ordering structures showed up in the 5% coverage configurations.
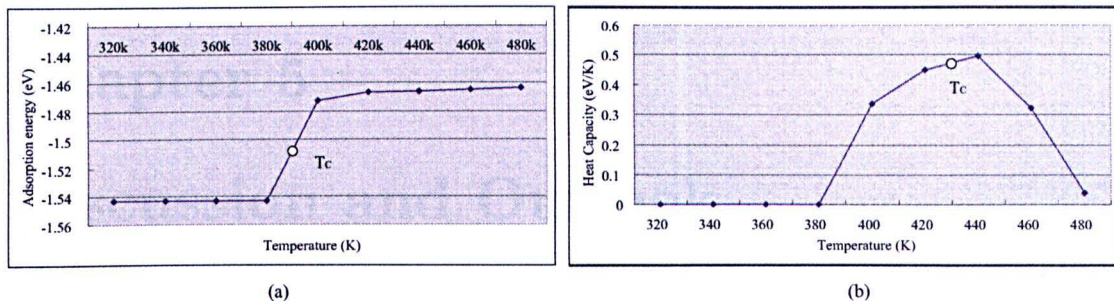
Figure 4.8: The phase transition diagram of the (5 0, 3 2) monolayer: (a) the averaged adsorption energy per molecule vs temperature. (b) heat capacity of the system vs temperature
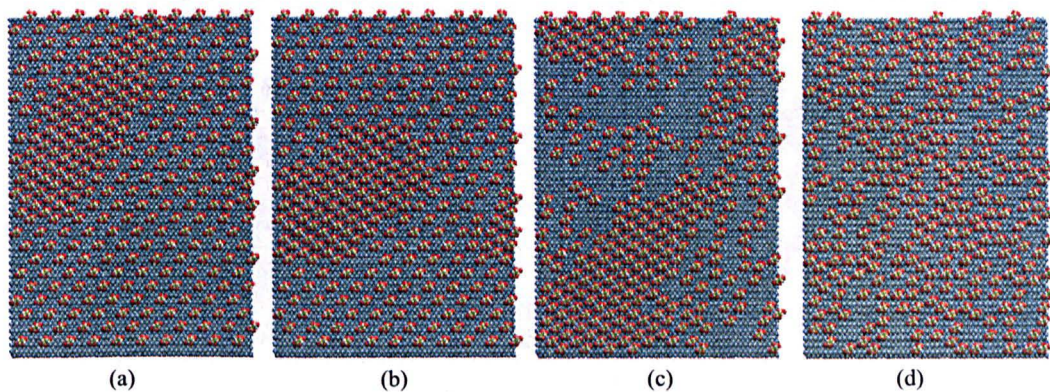

Figure 4.9: At the coverage of 12%, when the temperature is low, the ground state is a coexistence of two structure features. Above $400K$, the surface is dominated by islands of the "structure B". As the temperature goes higher, all ordered structures disappear. (a) $T = 200K$. (b) $T = 300K$. (c) $T = 400K$. (d) $T = 600K$.

Chapter 5

Discussion and Outlook

5.1   The (5 0, 3 2) structure

As illustrated in Section 3.1, the electronic surface landscape of the succinate-Cu(110) system depends prominently on the lateral interactions of the adsorbates. From the symmetry of the STM image alone, the corresponding surface structures can not be identified. Since the STM experiments were performed at low temperature, the (5 0, 3 2) ordered configurations may be the one we are looking for. The unit cell for (5 0, 3 2) monolayer is shown in Fig. 5.1(a). The adsorption energy per molecule obtained from the first-principles calculations is -1.440 eV, which is 0.1 eV higher than the predicted value. However, as this calculation is performed on a different unit cell, this energy difference may be caused by the non-comparable K-mesh samplings. In order to make the energy values obtained
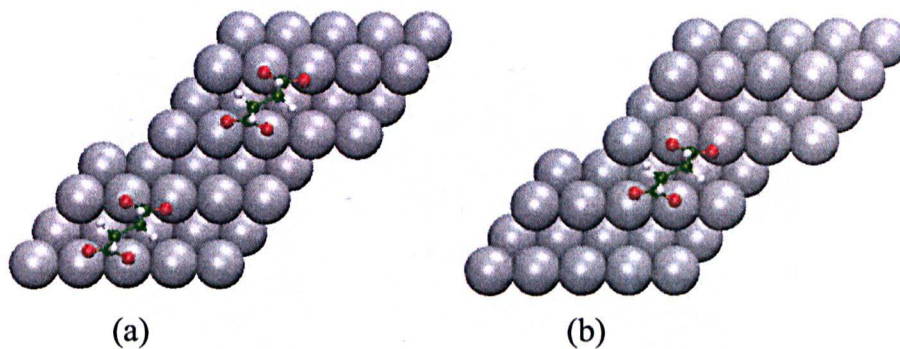

(a)   (b)

Figure 5.1: The unit cell of the (5 0, 3 2) monolayer used for DFT calculation. k-mesh is $2 \times 2 \times 1$. All the other calculation parameters are the same as the DFT calculations performed with the Cu(110)-$8 \times 6 \times 1$ unit cell.

from two unit cells be comparable, another isolated adsorption configuration is investigated, see Fig. 5.1(b). As shown in Table 5.1, the adsorption energy of the isolated succinate molecule in this unit cell is -1.473 eV, which is also about 0.1 eV higher than that in the Cu(110)-$8\times6\times4$ unit cell (1.572 eV). As a consequence,

the predicted intermolecular interaction energy on each succinate molecule is perfectly consist with direct DFT evaluations. Take into account that the relative

Table 5.1: Comparison of the energy values obtained from LGH predictions and direct DFT calculations. The lateral interactions per molecule is accurately predicted with the lattice gas hamiltonian.

	LGH predictions	DFT calculations
Adsorption energy per molecule in Fig. 5.1(a)	-1.543 eV	-1.440 eV
Adsorption energy per molecule in Fig. 5.1(b)	-1.572 eV	-1.473 eV
Intermolecular interaction in structure (5 0, 3 2)	0.029 eV	0.033 eV

energy differences in DFT calculations are very reliable and independent on the sampled k-meshes, we conclude that the lateral-interaction energy obtained from the effective cluster expansion fits the result of first-principle calculations. Thus, the ordered configuration (5 0, 3 2) is indeed the ground state when the coverage is 10%.
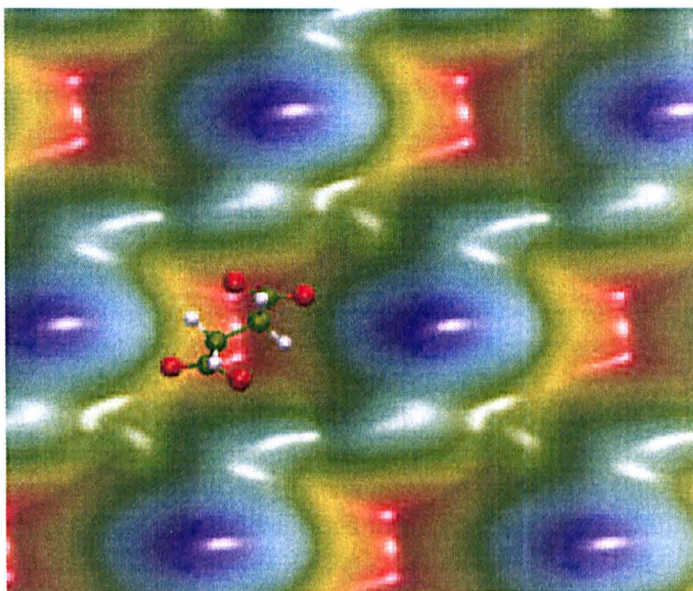

Figure 5.2: The simulated STM image of the (5 0, 3 2) monolayer with the Tersoff-Hamann method. $V_{bias} = -0.21V$.

The simulated STM image with the Tersoff-Hamann method is shown in Fig. 5.2. As can be seen, the symmetry of the electronic structure is consistent with the ordering of adsorbates. The adsorbed succinate molecules are imaged as protrusions with two small peaks. The shape of each protrusion, however, is very like

the features in Humblot's image [29], rather than the Liu's [30]. Nevertheless, since the Tersoff-Hamann method does not include any imaging effects of the STM tips, a different shape of the adsorbates could be obtained by including the STM tip in the simulation. Although the detailed features of a surface structure can be very different when the scanning conditions are changed, the growth directions of the adsorbates generally remains the same.

(a)   (b)   (c)
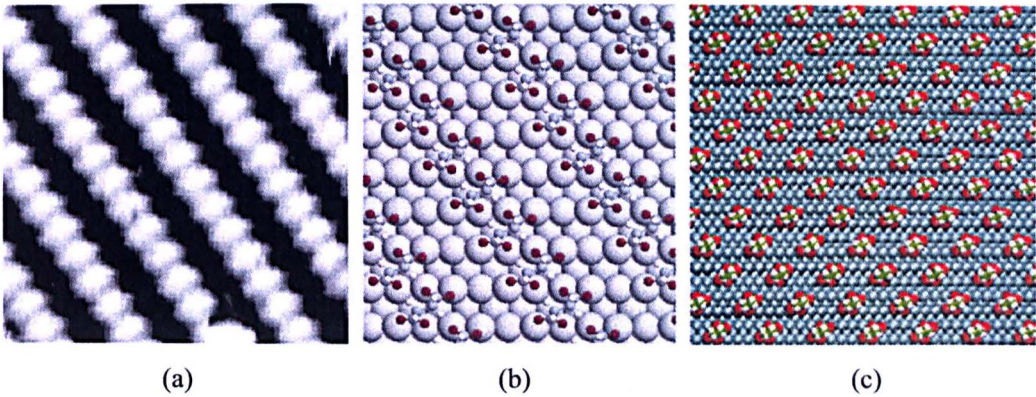
Figure 5.3: (a) The high resolution STM image of succinate molecules on Cu(110) surface [30]. (b) The theoretical model of the surface structure for image (a) [30]. (c) The structure of the (5 0, 3 2) monolayer.

As seen in Fig. 5.3, since the molecular growth directions in the (5 0, 3 2) monolayer is the same as that of Liu's image[30], it may also be the actual surface structure measured in Liu's STM experiment. However, we notice that the appeared coverage in the STM simulation is only half of that in the experimental result. The theoretical structure model in ref. [30] suggests that the ordering pattern of the imaged features in Fig. 5.3 (a) corresponds to a (5 0, 4 1) molecular overlayer, which doubles the density of adsorbates along the [1$\bar{1}$2] direction and leads to a surface coverage of 20%, see Fig. 5.3 (b). If this model is correct, the pair interaction and trio interaction in Fig. 5.4 should be attractive or at least not too repulsive. However, the calculations show that these two configurations are the most repulsive interactions in all pair and trio configurations. The interaction energy is 62 meV and 25 meV per molecule for Fig. 5.4(a) and (b), respectively. Comparing with this theoretical model, the (5 0, 3 2) configuration is thermodynamically much more stable. We thus suggest that the (5 0, 3 2) is the surface structure of the STM image Fig. 5.3(a). The growth directions of the molecular chains in this case are perfectly consistent. The difference of appeared coverage may be a result of the imaging effects of an STM tip.
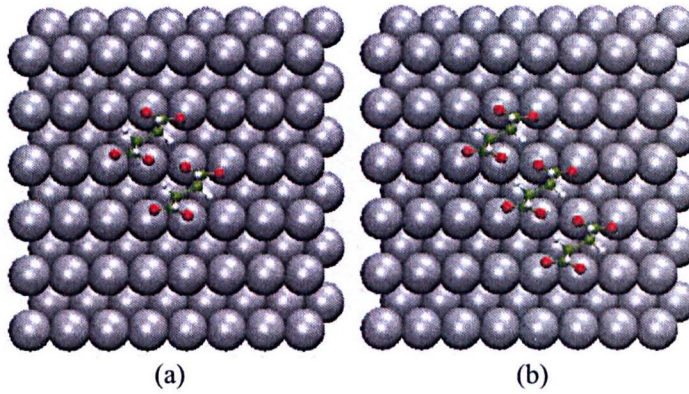
Figure 5.4: (a) The pair configuration and (b) trio configuration in the theoretical model of Ref. [30].

5.2 Monte Carlo simulation with Antonio model

Since the (1 1, -9 0) and (9 0, 1 -1) ordering pattern in Humblot's image has not been confirmed, we performed another set of DFT calculations based on Antonio adsorption configuration. The calculations were carried out in a smaller unit cell, see Fig. 5.5. In total, 32 configurations (18 pairs and 14 trios) have been
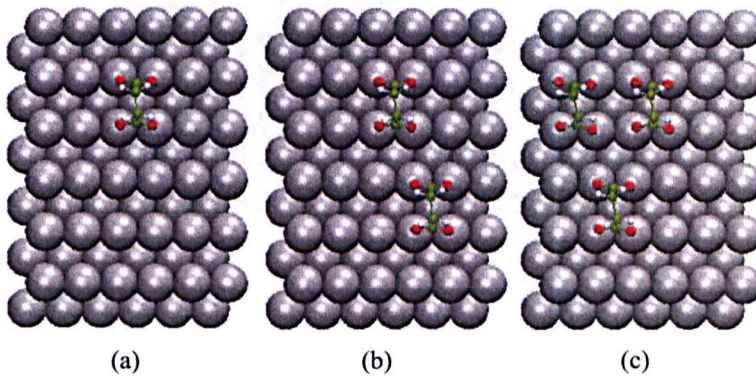

Figure 5.5: Schematic representation of some molecular structures calculated for Antonio adsorption configuration

directly calculated by DFT. The lattice gas hamiltonian is constructed with the same procedure described in 3.2. The Monte Carlo simulations are performed on a 40 × 40 unit cell. As shown in Fig. 5.6, the ground state is always the same ordered pattern. The stability of the ordered structures increases with the coverage. Nevertheless, the growth direction of this ordered structure is very different from the (1 1, -9 0) and (9 0, 1 -1) configurations in ref. [29]. Considering also that this adsorption configuration is less stable, we conclude that the Antonio model would not be correct.
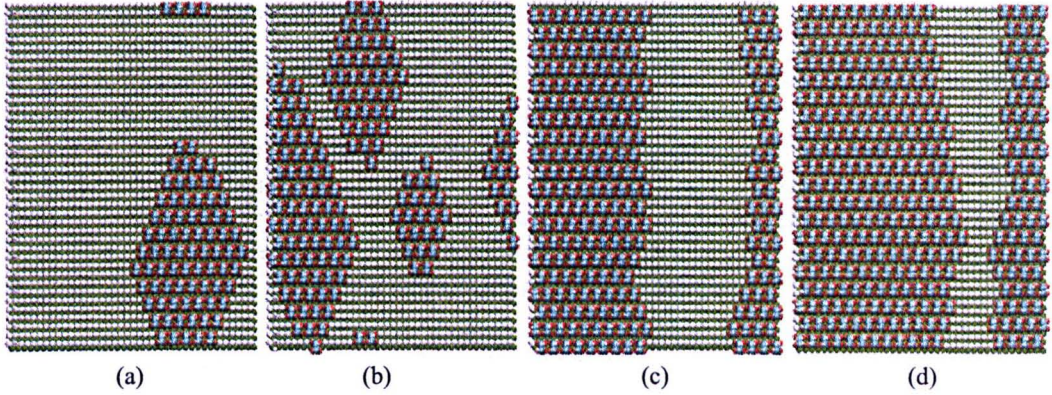
<div style="text-align:center">(a)   (b)   (c)   (d)</div>

Figure 5.6: Schematic representation of some molecular structures calculated for the Antonio adsorption configuration

5.3 A new adsorption configuration

Very recently, we found that if the original Antonio configuration is twisted, the obtained structure is found as stable as the Darling model. The adsorption

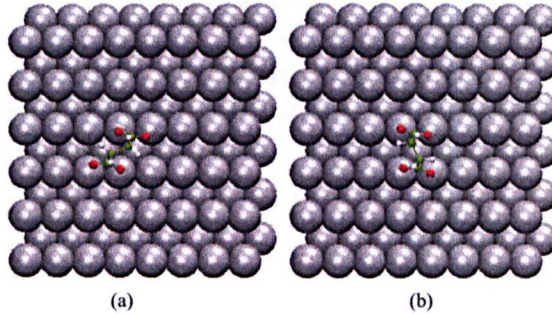<div style="text-align:center">(a)   (b)</div>

Figure 5.7: The adsorption energy of the Darling model (a) and the new model (b) differs by only 3meV.

energy of these two configurations differ only by 3meV. In order to ascertain this energy difference is not dependent on the number of k-points, the k-mesh has been increased to $2 \times 2 \times 1$. The relative energy difference remains the same. Since this configuration has only been found recently, the first-principles calculations have not been completed. However, as a first result we found that the interaction energies obtained from DFT are very different. Due to the time constrain of this PhD project, we could not complete a suitable set of first-principle calculations for this configuration. But, taking its stability into account, this configuration may yield a stable molecular structure which could account for the (1 1, -9 0) and (9 0, 1 -1) molecular ordering observed in some experiments [29].

5.4 Conclusions

In this work, we combined a microscopic first-principles description of intermolecular interactions with a macroscopic description of a large number of adsorbates. Based on the Darling adsorption configuration, two ordered structures have been identified for succinate molecules on Cu(110). The (5 0, 3 2) monolayer structure is found to be the ground state configuration accounting for the STM images in Ref. [30]. The mechanism of the self-assembly behavior can be attributed to the indirect interactions of the succinate molecules on the Cu(110) surface.

This multi-sampling approach opens a connection between DFT in the microscopic scale and Thermodynamics and Statistics in the macroscopic scale. The generated Monte Carlo code is able to statistically treat the information collected in the electronic regime and produce reliable predictions of the thermodynamic properties in the macroscopic regime.

Even though the multi-scale sampling method used in this work is one of the best approaches used to understand the nature of chemical processes, the limitations of this approach have also become evident. Differing from atomic adsorbates, organic molecules are associated with a large number of degrees of freedom; the ground state structure search for molecules on a surface thus plays a pivotal role in the simulated results. Since the determination of the ground state adsorption configuration is often ambiguous, the prediction of the ordered structures is not 100% reliable. In addition, a systematic generation of a lattice gas hamiltonian require a large number of DFT calculations. With today's computational resources into account, the success of this approach in determining the ordering of atomic adsorbates [87] cannot be easily transferred to molecular species. In future, more efforts have to be made, in my opinion, to the improve the sampling methods.

Appendix A

Derivation of LGH from a DFT database

The basic idea of the multi-scale molecular sampling method is to statistically treat the elementary processes of a large number of quantum particles. Since not all elementary processes can be described by dedicated DFT calculations, the cluster expansion approximation is introduced: use a set of energy parameters to describe the energetics of all configurations of the quantum particles. For a two dimensional lattice gas system in which all adsorbates are identical, the set of energy parameters is called "Lattice Gas Hamiltonian" (LGH). The energy of a lattice configuration is express by

$$
E = E(\text{clean surface}) + \overset{\text{all sites}}{\sum_{i}} \sigma_i E_o + \overset{\text{all site pairs}}{\sum_{i \neq j}} V_{P_{(i,j)}} \sigma_i \sigma_j + \overset{\text{all site trios}}{\sum_{i \neq j \neq k}} V_{T_{(i,j,k)}} \sigma_i \sigma_j \sigma_k + \cdots ,
$$

$$(A.1)$$

where the "sites" are corresponding to the local minima in the potential energy surface of the substrate-adsorbate system. The accuracy of predicted energy E in equation A.1 is therefore primarily dependent on the quality of the energy parameters.

A.1 Least-Squares Fitting of data

In multi-scale sampling, the energy parameters are derived from a database of DFT calculations. For a given set of calculated lattice configurations, the energetics obtained from DFT are expanded with equation (A.1); while the energy parameters are evaluated with linear algebra methods. To understand this point, we give an example. We start with the simplest lattice gas hamiltonian, which contains only two terms: the adsorption energy V_o, and the interaction energy between nearest neighbors V_P. The substrate is a metal (100) surface and the adsorption sites are exclusively the hollow sites.

Fig. A.1 shows three lattice configuration calculated with DFT. As can be seen, Fig. A.1(a) contains seven adsorbates and two pair interactions. The energy of
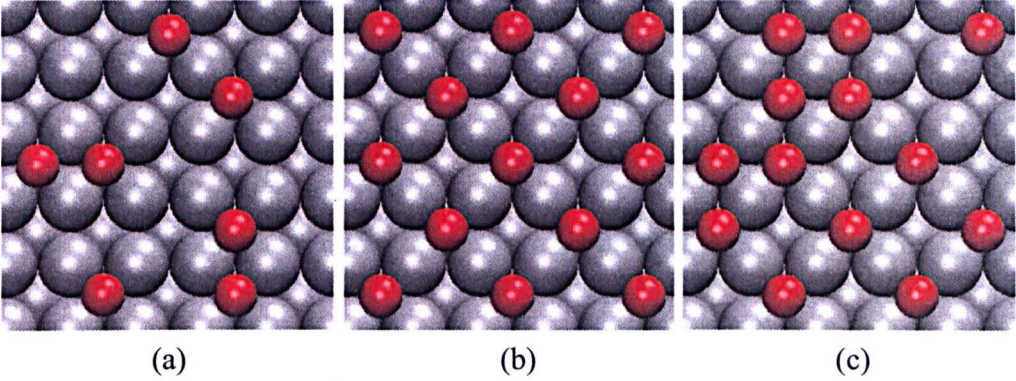
Figure A.1: Schematic representations of three lattice configurations.

this configurations is therefore expressed as

$$E_1 = 7V_o + 2V_P. \tag{A.2}$$

In Fig. A.1(b), there is no pair interactions but only 13 adsorbates. Thus, the energy expression for this structure is

$$E_2 = 13V_o + 0V_P. \tag{A.3}$$

In Fig. A.1(c), the number of pair interactions is 7 while the total number of adsorbates is 13. As a result, the energy for this configuration is expressed by

$$E_3 = 13V_o + 7V_P. \tag{A.4}$$

If we write the energy expressions of the three configurations together, we have

$$\begin{aligned}
E_1 &= 7V_o + 2V_P \\
E_2 &= 13V_o + 0V_P \\
E_3 &= 13V_o + 7V_P.
\end{aligned} \tag{A.5}$$

The obtained system of linear equations can then be expressed by the matrix notation:

$$Ax = b. \tag{A.6}$$

Here, A is the coefficient matrix, x is the column vector, which contains all unknown energy parameters in LGH; while b is the column vector representing the corresponding energies of the configurations.

$$
\begin{bmatrix}
7 & 2 \\
13 & 0 \\
13 & 7
\end{bmatrix}
\begin{bmatrix}
V_o \\
V_P
\end{bmatrix}
=
\begin{bmatrix}
E_1 \\
E_2 \\
E_3
\end{bmatrix}
\tag{A.7}
$$

Generally, the solutions of equation (A.6) can be considered in three situations.

- If there are more unknown variables than the independent equations, the equation system is under-determined. There will be infinite sets of solutions which satisfy the equation system.

- If the number of the unknown variables equals to the number of equations, x has exactly one solution.

- If the unknowns are less than the equations, the equation system is over-determined. As a consequence, there is no exact solution of the system.

In multi-scale sampling, the LGH cannot be derived when the equation system is under-determined. For the situation of equal numbers of equations and unknowns, x can always be solved by

$$x = \frac{A_T b}{A^T A}. \tag{A.8}$$

Here A^T represents the non-conjugate transpose of the coefficient matrix A. If the system is over-determined, the values of x can only be estimated by least squares fits.

$$\hat{x} = \frac{A_T b}{A^T A} \tag{A.9}$$

The difference between equation (A.8) and (A.9) is that the former results in an exact solution for x; while the latter gives only the best solution that minimizes the squared error E^2 (where $E^2 = \parallel b - Ax \parallel^2$). In the sample equation system (A.7), the squared error is

$$
\begin{aligned}
E^2 &= \parallel b - Ax \parallel^2 \\
&= (E_1 - 7V_o - 2V_P)^2 + (E_2 - 13V_o - 0V_P)^2 + (E_3 - 13V_o - 7V_P)^2 \tag{A.10}
\end{aligned}
$$

A.2 Cross-Validation method

Once a lattice gas hamiltonian has been derived, its predictive property has to be tested. Only the Hamiltonian providing the best prediction of configurational energy can be used for cluster expansions. Ideally, the predictive power of each lattice hamiltonian is tested by several extra first-principles calculations. If the predicted energy values are very close to the configurational energy calculated with DFT directly, the LGH can be regarded as a reliable predictor of energies of any other lattice configurations. Thus, the database of the DFT results should be partitioned into two sets: the "training set", where the LGH is derived and the "test set", which is used to evaluate the obtained LGH.

However, the first-principles simulations are very time consuming and the number of configurations for a particular unit cell is also limited. One thus needs a method to maximize the size of the two sets at the same time. Luckily, the cross-validation (CV) method [89, 90, 91] uses all available data for both training and testing purposes: the database is recursively partitioned into the "training set" and "test set" so that all possible combinations in training and testing are

included.

For a given "training set", the lattice gas hamiltonian can be derived with the linear algebra methods shown in section B.1. The cross-validation error can then be estimated by

$$CV^2 = \frac{1}{N} \sum_{i=1}^{N} (E_{DFT}^i - E_{CE}^i)^2,$$ (A.11)

where E_{CE} represents the configurational energies predicted with the derived LGH. The smaller the obtained CV value is, the better the predictive power of the derived LGH has. The importance of the cross-validation method is therefore to provide a quantitative criterium to identify the best LGH.

In contrast to the mean square error, the CV error does not monotonically decrease with increasing numbers of parameters included in the LGH. Instead, the CV values first decrease because more degrees of freedom are available to account for the variation of energy, and then increases due to an increase of the noise in the data fitting, i.e. over-fitting. The LGH we are looking for is therefore the best compromise of the two effects, i.e. the minimum CV value.

A.3   Genetic Algorithms

For the next step, a large number of training sets are going to be investigated in order to identify the best LGH. However, the number of all possible systems of an N-configuration DFT database is 2^N, because every configuration can either be in the "training set" or the "test set". Even for a moderate sized database, e.g. N = 40, the total number of possible training sets is already astronomically large. Of more importance, only very small partitions of the 2^N systems are capable of generating potentially important LGHs. Clearly, an optimized algorithm is needed to efficiently explore the equation-system space. Genetic algorithms (GA) are particularly suitable for this kind of problem.

In general, GA are a class of population based function optimizers, which use selection and recombination operators to generate new sample points in a search space. An implementation of a genetic algorithm begins with a population of randomly selected chromosomes. Traditionally, the chromosomes are represented with binary bit strings of length L. The encoded chromosomes are referred to as the "first generation". Once the initial population have been created, each string is evaluated by the "evaluation function" and then assigned a "fitness value". The fitness is a measurement of how well that chromosome can solve the targeting problem. Very importantly, the value of fitness of a string is defined with respect other members of the current generation. For instance, in the canonical genetic algorithm, the fitness value is defined by : $f_i/\bar{f}$, where f_i is the evaluation associated with the ith string and $\bar{f}$ is the average evaluation of the current population. The new generation is then produced by "mutation" or "crossover" from the first generation.

103

The reproductive opportunity of each string is allocated proportional to its value of fitness, i.e. those chromosomes with better fitness values have more chances to "reproduce" than those whose fitness values are poor. As a consequence, a large proportion of the second generation are the "children" of the fittest chromosomes in the first generation. The best "genes" can therefore be transmitted from one generation to the next. Over successive generations, the population evolves towards an optimal solution.

In this work, we aimed at an optimized lattice gas hamiltonian of predictive power for configurational energies. Thus, the corresponding chromosomes are the "training sets" partitioned from the DFT database. Since a "training set" can always be written in the format of a linear equation system (e.g. Equation 3.14), the chromosomes can be encoded into binary bit strings as following:

- The length of the bit string L equals the number of configurations included in the whole DFT database.

- All terms in DFT database are expressed by linear equations shown in Equation 3.14.

- The equations selected into the "training set" are marked as "1" the others are marked as "0". For instance, if there are 5 configurations calculated by DFT and the first two are selected in the "training set", the corresponding string for this partition is "11000".

When the initial population has been encoded, the generated strings are then evaluated by the cross-validation method (evaluation function). The fitness values of each "training set" should therefore be determined by the CV errors of all individuals in the first generation: a small CV error corresponds to a high fitness value. Based on the "reproductive ability" allocated to each string, the next generation is then produced from the current generation by

- Crossover: Select two strings as parents. Randomly choose a recombination point. Split the parent strings at that point and recombine the two parents to form two children, see Fig. A.2.

- Mutation: Apply random changes to individual parents to produce children.

$$1001|101 \xrightarrow{\text{crossover}} 1001|000$$
$$1111|000 \qquad\qquad\quad 1111|101$$
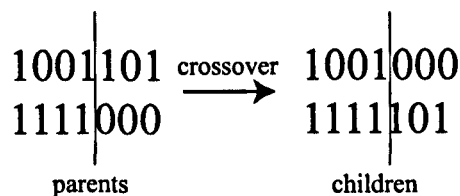
parents   children

Figure A.2: A schematic representation of the crossover method.

Repeating the processes described above over a number of generations. The CV errors of the produced chromosomes are expected to converge to the minimum. A predictive LGH can therefore be selected.

Appendix B

Implementation of the Metropolis Monte Carlo Algorithm

B.1 Set up a initial surface

The Monte Carlo code starts with constructing a Cu(110) surface of $M \times N$ sites. The adsorption sites are represented by a four dimensional array. In each row, the first two numbers indicate the positions of the sites and the last two numbers express the occupation state and energy level of that site. The occupation parameter takes only two values: 1 for occupied and 0 for empty. The energy parameter, however, is evaluated by effective cluster expansion. For any clean surfaces, the occupation and energy variables for all sites take the value of 0. When a certain coverage of molecules adsorb on the surface, the corresponding
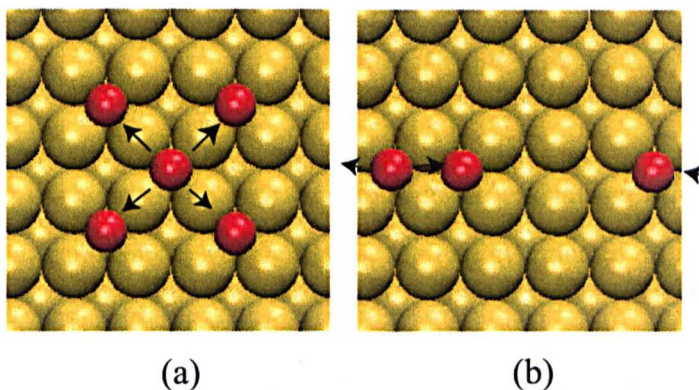

(a)   (b)

Figure B.1: Schematic representation of the local occupation environment: (a) the atom in the center has four pair interactions , which are of the same type. (b) Due to the periodic boundary condition the atom on the left has two pair interactions.

percentage of sites change their occupation value to 1 and the energy value of the sites are defined as following:

- Scan all the surface sites successively.

- If the site is empty, its energy value is assigned as zero.

- If the site is occupied, its neighboring sites are scanned. The energy value of this site is determined with the effective cluster expansion method, see Fig. B.1(a).

$$E = E_o + \sum V_{pairs} + \sum V_{trios} + \dots \qquad (B.1)$$

- In order to mimic an surface of infinite size, the boundary condition should be included in the energy calculation, see Fig. B.1(b).

When the scan is finished, the configurational energy of the surface can be calculated by summing up the energy values of all sites.

B.2 Move the molecules on a surface

Randomly select an occupied site A and an empty site B. Denote the energy value of site A as E_{old}. Swap the occupation values of the two sites. Calculate the site energy for B and denote its value as E_{new}.

- If $E_{new} < E_{old}$, this movement is accepted. As a result, the occupation value becomes 0 for site A and 1 for site B, respectively.

- If $E_{new} > E_{old}$, a random number, which is homogeneously distributed in the range of 0 to 1 should be generated. Here, we call this number as $rand$.

$$\text{Then, the movement is} \begin{cases} \text{accepted} & \text{if } rand \leq exp(-\Delta E/K_B T) \\ \text{rejected} & \text{if } rand > exp(-\Delta E/K_B T) \end{cases} \qquad (B.2)$$

where $\Delta E = E_{new} - E_{old}$. The occupation values of the selected sites are swapped in the case of acceptance and remain unchanged otherwise.

- After each movement of adsorbates, the energy value of all lattice sites are updated, so that the total energy of the system can always be expressed as the sum of the site energies.

- Repeat the steps described above for a large number of "Monte Carlo steps", the system should reach the thermodynamic equilibrium. The convergence can be checked with state variables such as the total energy and heat capacity.

- By tuning the values of temperature and coverage, this Monte Carlo code should produce a phase diagram of any adsorbate-substrate system.

B.3   Visualization of the lattice configurations

At contain coverage and temperature, in principle, the Metropolis Monte Carlo method can always lead to the ground state lattice configuration if the number of "Monte Carlo steps" are large enough. When the thermal equilibrium has been reached, the growth of the Markov chain should be terminated. Since the information of the lattice configuration is encoded in the four dimensional array, a visualization program is needed to convert the array to an image. In the case of succinate molecules on Cu(110), the adsorption site has been defined as the long-bridge site of the Cu surface. As a consequence, in the lattice configurational images, the long-bridge site is represented by a top-layer Cu atom and a second-layer Cu atom. When the clean surface has been properly constructed, a molecular structure is drawn above each occupied adsorption site.

B.4   Testing an Monte Carlo program

As has been stressed in the section 2.3, a Metropolis Monte Carlo code is used to study the physical properties of a thermodynamic equilibrium system. Since the state variables does not dependent the pathway how the equilibrium has been reached, the phase diagram should be independent on

- the random number generator, which determines the "growth direction" of the Markov chain.

- the size and shape of the unit cell used to represent an infinite adsorbate-substrate system in 2 dimensions.

- the initial configuration from which the Markov chain starts.

In this work, the thermodynamic properties of the succinate-Cu(110) systems have been investigated with a 50×50 unit cell, and the obtained results are then checked with a 40×40 unit cell. Since the two sets of simulations result in the same ordering patterns and other physical properties for a certain combination of temperature and coverage. We conclude that this Monte Carlo code is qualified to simulate the succinate-Cu(110) systems in equilibrium.

Bibliography

[1] S. Völkening, K. Bedürftig, K. Jacobi, J. Wintterlin and G. Ertl, Phys. Rev. Lett. 83, 2672 (1999)

[2] R. Whyman, "Applied Organometallic Chemistry and Catalysis" Oxford University Press, 2001.

[3] http://en.wikipedia.org/wiki/chirality

[4] P. Wothers, N. Greeves, S. Warren and J. Clayden, "Organic Chemistry", Oxford University Press, 2001.

[5] G. V. Smith and F. Notheisz, "Heterogeneous Catalysis in Organic Chemistry" Eds,. New York, 1999.

[6] D. Yu. Murzin, P. Mäki-Arvela and T. Salmi, Kinetics and Catalysis 44 No.3 323 (2003)

[7] H. U. Blaser and M. Müller, Stud. Surf. Sci. Catal., 59, 73 (1991)

[8] H. U. Blaser, F. Spindler and M. Studer, Appl. Catal. A. 221, 119 (2001)

[9] G. Ertl, H. Knözinger and J. Weitkamp, Eds., "Handbook of Heterogeneous Catalysis" Wiley-VCH, Weinheim, 1997.

[10] B. E. Nieuwenhuys, Surf. Rev. Lett. 50, 1869 (1996)

[11] M. A. Vannice, J. Catal. 50, 228 (1977)

[12] G. Binning, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. 49, 57 (1982)

[13] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. 50, 120 (1983)

[14] http://en.wikipedia.org/wiki/Scanning-tunneling-microscope

[15] W. A. Hofer, A. S. Foster, A. L. Shluger, Rev. Mod. Phys. 75, 1287 (2003)

[16] Y. Sun, H. Mortensen, S. Schar, A.-S. Lucier, Y. Miyahara, P. Grutter, and W. A. Hofer Phys. Rev. B 71 , 193407 (2005)

[17] F. Calleja, A. Arnau, J. J. Hinarejos, A. L. Vazquez de Parga, W. A. Hofer, P. M. Echenique, and R. Miranda Phys. Rev. Lett. 92, 206101 (2004)

[18] Z. T. Deng, H. Lin,W. Ji, L. Gao, X. Lin, Z. H. Cheng, X. B. He, J. L. Lu, D. X. Shi,W. A. Hofer, and H.-J. Gao Phys. Rev. Lett. 96, 156102 (2006)

[19] S. M. Barlow and R. Raval, Surf. Sci. Rep. 50, 201 (2003)

[20] V. Humblot, S. M. Barlow and R. Raval, Progress in Surface Science, 76 1-19 (2004)

[21] V. Humblot, M. O. Lorenzo, C. J. Baddeley, S. Haq and R. Raval J. Am. Chem. Soc. 126 6460 (2004)

[22] G. P. Srivastava, S. Mohan and Y. S. Jain, J. Raman Spectrosc. 13, 25 (1982)

[23] R. Bhattacharjee, Y. S. Jain, G. Raghubanshi and H, d. Bist, J. Raman Spectrosc 19, 51 (1988)

[24] J. Tabatabaei, B. H. Sakakini, M. J. Watson and K. C. Waugh, Catal. Lett. 59, 143 (1999)

[25] J. Tabatabaei, B. H. Sakakini, M. J. Watson and K. C. Waugh, Catal. Lett. 59, 151 (1999)

[26] T. Genger, O, Hinrichsen and M . Muhler, Catal. Lett.59, 137 (1999)

[27] G. Kresse and J. Furthmüller, Computational Materials Science 6 15 (1996)

[28] W. Koch and M. C. Holthausen, "A Chemist's guide to Density Functional Theory", Wiley-VCH, 1999.

[29] L. Antonio, M . M . Barbosa and P. Sautet, J. Am. Chem. Soc. 123 6639 (2001)

[30] N. Liu, S. Haq, G. R. Darling and R. Raval, "Direct Visualisation of Enantiospecific Substitution of Chiral Guest Molecules into Heterochiral MolecularAssemblies at Surfaces", to be published.

[31] W. A. Hofer, V. Humblot and R. Raval, Surf. Sci. 554, 141 (2004)

[32] M. Born and J. R. Oppenheimer and Ann. Physik, 84, 457 (1927)

[33] R. M. Martin, "Electronic Structure - Basic Theory and practical methods", Cambridge University Press, 2005

[34] D. R. Hatree, Proc. Camb. Phil. Soc. 24, 89, 111 (1928)

[35] K. Ohno, K.Esfarjani and Y. Kawazoe, "Computational Materials Science" Springer, 1999

[36] P. Honhenberg and W. Kohn, Phys. Rev. 136, B864 (1964)

[37] W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965)

[38] M. K. Harbola, Phys. Rev. B, 60, 4545 (1999)

[39] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. 45 566 (1980)

[40] S. J. Vosko, L. Wilk and M. Nusair, Can. J. Phys. 58, 1200 (1980)

[41] J. P. Perdew and Y. Wang, Phys. Rev. B 45 13244 (1992)

[42] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, Phys. Rev. B 46 6671 (1992)

[43] A. S. Foster and W. A. Hofer, "Scanning Probe Microscopes - atomic acale enginnering by forces and currents" Springer, 2006

[44] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. 50, 1988 (1985)

[45] J. Tersoff and D. R. Hamann, Phys. Rev. B 31, 805 (1985)

[46] J. Bardeen, Phys. Rev. Lett. 6 57 (1961)

[47] M. Bütticker, Y. Imry, R. Landauer, S. Pinhas, Phys. Rev. B 31 6207 (1985)

[48] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. 68 2512 (1992)

[49] G. A. D. Briggs and A. J. Fisher, Surf. Sci. Rep. 75, 1 (1999)

[50] D. Drakova, Rep. Progr. Phys. 64, 205 (2001)

[51] W. A. Hofer, J. Redinger, A. Biedermann and P. Varga, Surf. Sci. Lett. 466 L795 (2000)

[52] W. A. Hofer, A. J. Fisher, R. A. Wolkow and P. Grütter, Phys. Rev. Lett. 87 236104 (2001)

[53] C. J. Chen "Introduction to Scanning Tunneling Microscopy", New York, Oxford Univeristy Press (1993)

[54] D. Wasserfallen, M. Kastler, W. Pisula, W. A. Hofer, Y. Fogel, Z. H. Wang, K. Mullen JACS 128, 1334 (2006)

[55] S. G. Lemay, J. W. Janssen, M. Hout, M. Mooij, M, J. Bronikowski, P. A. Willis, R. E. Smalley, L. P. Wouwenhoven and C. Dekker, Nature (London) 412, 617 (2001)

[56] G. V. Nazin, X. H. Qiu and W. Ho. Science 302, 77 (2003)

[57] C. Joachim, J. K. Gimzewski and A. Aviram, Nature (London) 408, 541 (2000)

[58] E. G. Emberly and G. Kirczenow, Phys. Rev. Lett. 91, 188301 (2003)

[59] J. Kuntze, R. Berndt, P. Jiang, H. Tang, A. Gourdon and C. Joachim, Phys. Rev. B 65, 233405 (2002)

[60] J. Repp, G. Meyer, S. M. Stojković, A. Gourdon and C. Joachim, Phys. Rev. Lett. 94, 026803 (2005)

[61] C. Seidel, R. Ellerbrake, L. Gross and H. Fuchs, Phys. Rev. B 64 195418 (2001)

[62] H. Han, M. Hongying, C. Qiao, Y. Xinzheng, Q. Huiqin, Z. Jianhua, L. Haiyang, H. Pimo and B. Shining, Phytsica (Amsterdam) 352B, 36 (2004)

[63] K. Reuter, C. Stampfl and M. Scheffler, "Handbook of Material Modeling", Springer, Berlin (2005)

[64] K. Stowe, "An introduction to thermodynamics and statistical mechanics ", Cambridge University Press 2007

[65] R. P. Feynman, Phys. Rev., 56, 340 (1939)

[66] A.C. Hurley, Proc. Roy. Soc., 226A, 171 (1954)

[67] T. Guénault "Statistical Physics", Kluwer Academic Publishers (1995)

[68] J. M. Seddon and J. D. Gale "Thermodynamics and Statistical Mechanics" The Royal Society of Chemistry (2001)

[69] R. Bowley and M. Sánchez "Introductory statistical mechanics", Clarendon press, Oxford (1999)

[70] D. Frenkel and B. Smit "Understanding molecular simulation : from algorithms to applications", Academic Press (1996)

[71] A. R. Leach "Molecular Modelling : Principles and Applications", Pearson Education imprint (2001)

[72] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, J. Chem. Phys, 21, 1087 (1953)

[73] G. Kresse and J. Hafner, Phys. Rev. B 47, 558 (1993)

[74] G. Kresse and J. Hafner, Phys. Rev. B 49 14251 (1994)

[75] G. Kreese and J. Furthmüller, Comput. Mater. Sci. 6, 15 (1996)

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

[77] F. Birch, J. Appl. Phys. 9, 279 (1938)

[78] F. D. Murnaghan, Proc. Nat. Acad. Sci. 30, 244 (1944)

[79] C. -L. Fu and K. -M. Ho, Phys. Rev. B 28, 5480 (1983)

[80] Handbook of Chemistry and Physics, 76th edition,12-18, CRC press, (1995-1996)

[81] S. Kurth, J. P. Perdew and P. Blaha, International Journal of Quantum Chemistry 75 889 (1999)

[82] M. Scheffler and C. Stampfl, in Handbook of Surface Science: Electronic Structure, Vol. 2, edited by H. Horn and M. Scheffler, Elsevier, Amsterdam (2000)

[83] W. Kohn and K. H. Lau, Sol. State Commun. 18, 553 (1976)

[84] P. Atkins and J. Paula, "Atkins' Physical Chemistry" 8th edition, Oxford University Press, 2006

[85] J. M. Yeomans, "Statistical Mechanics of Phase Transitions", Clarendon Press, Oxford, 1992.

[86] K. Oura, V. G. Lifshits, A. A. Saranin, A. V. Zotov and M. Katayama, Surface Science: An Introdution, Springer, Berlin (2003)

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

[88] F. Wang and D. P. Landau, Phys. Rev. Lett. 86, 2050 (2001)

[89] A. van de Walle and G. Ceder, J. Phase Equilib. 23 348 (2002)

[90] D. M. Hawkins, S. C. Basak and D. Mills, J. Chem. Inf. Comput. Sci. 43 579 (2003)

[91] K. Baumann, Trends Ana. Chem.22, 395 (2003)