Università degli Studi di Sassari
Dipartimento di Chimica e Farmacia

Scuola di Dottorato in Scienze e Tecnologie Chimiche
Indirizzo Scienze Chimiche
Ciclo XXVI

Pros and Cons of three Approaches to the Study of Diffusion in Zeolites: Cellular Automata, Networks and second-order Markov Models

Tesi di Dottorato di
Alberto M. Pintus

Il Direttore:
Prof. Stefano Enzo

Il Supervisore:
Prof. Pierfranco Demontis

Anno Accademico 2012/2013

La presente tesi è stata prodotta nell'ambito della scuola di dottorato in Scienze e Tecnologie Chimiche dell'Università degli Studi di Sassari, a.a. 2011-2013 – XXVI °ciclo, con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività l.3.1".

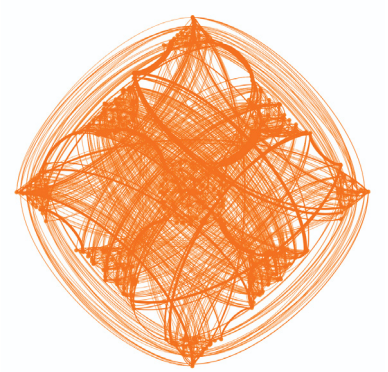
RINGRAZIAMENTI

Desidero ringraziare il mio supervisore prof. Pierfranco Demontis per la grande libertá intelletuale accordatami, il prof. Giuseppe Baldovino Suffritti, e il Dott. Federico Pazzona con il quale ho condiviso vari progetti durante questi tre anni. Un ringraziamento va al dott. Marcello Budroni per aver introdotto il germe delle reti nel nostro gruppo di ricerca e per utili discussioni su questo tema.

Un grazie assolutamente particolare lo devo al dott. Andrea Gabrieli per l'incondizionato, generoso, e spesso salvifico supporto informatico e non solo.

...É chiaro, quindi, che l'idea di un metodo fisso o di una teoria fissa della razionalità poggia su una visione troppo ingenua dell'uomo e del suo ambiente sociale. Per coloro che non vogliono ignorare il ricco materiale fornito dalla storia e che non si propongono di impoverirlo per compiacere ai loro istinti più bassi, alla loro brama di sicurezza intellettuale nella forma della chiarezza, della precisione, dell'"obiettività", della "verità", diventerà chiaro che c'è un solo principio che possa essere difeso in tutte le circostanze e in tutte le fasi dello sviluppo umano. E' il principio: qualsiasi cosa può andar bene.

Paul K. Feyerabend "CONTRO IL METODO"



Contents

Chapter 1

Introduction

The following pages constitute an attempt to rationalize a three years effort to find an original point of view on the problem of adsorption and diffusion in microporous materials. This effort was only partially successful and many questions regarding the various approaches here exposed remains unanswered. However some parts of this work remains interesting in my opinion and could deserve further investigations.

The fundamental topics on which this thesis work focuses, are the possibility of applying certain reductionistic approaches to the dynamics of sorbate molecules in zeolites, and similar materials, in order to obtain a coarse-grained model of it, and the development of tools for better understanding some of the fundamental mechanisms of this dynamics, inspired by the different points of view on the phenomena of interest that a meaningful coarse-graining could bring to light. In the following, the idea of coarse-graining is intended in a perhaps slightly different way from its most common use. It is in general related to the reduction of a complex problem to its really fundamental elements, by discarding all that is not strictly necessary, with the aim of making it easier to handle. But while in the study of proteins, for example, it refers to the limitation of the degrees of freedom of the macromolecule, by considering groups of atoms as single units interacting via a simplified force field (as in Go models), we deal here with a coarse-graining of space rather than of molecular structure. In a sense the molecular structure of the zeolite is coarse-grained in a first step, but this induces immediately a coarse-graining of the space available to the motion of sorbed molecules. Once the space is partitioned in a sensible way, dynamics can be treated on the basis of this discrete space.

Molecular Dynamics (MD) is now widely used to simulate adsorption and transport phenomena in zeolites [1–3]. It is a powerful tool, and its application has brought deep insights in the behaviour of molecules under confinement, shedding light on many interesting phenomena and providing an explanation to experimental results. However, despite the continuous

growth of available resources for massive calculation, MD simulations for large systems and/or long times are still very computationally demanding, and this justifies an effort to develop coarse-grained methods allowing a substantial extension of the range of systems that can be simulated.

Examples of coarse-graining in the field of zeolite studies [4] are Configurational-Bias Monte Carlo methods [5], Kinetic Monte Carlo [6,7] (KMC) techniques, the hierarchical approach proposed Tunca *et al* [8], and a lattice-gas Cellular Automaton (LGCA) developed in our group [9].

In section 2.2 a variant of this LGCA suitable for simulations of mixtures adsorbed in zeolites is presented, alongside with a still more reductionistic model of the same kind and a detailed study of the associated time correlation function.

Monte Carlo and Cellular Automata methods require a number of parameters, which must be properly tuned to reproduce the dynamics of interest. Definition of the sites, of the jumps between them and their probabilities, alongside with a meaningful definition of the time scale are all difficult issues, requiring usually an ad-hoc determination which is justified mainly by the agreement of simulation data with experimental or MD global results, such as diffusivities or adsorption isotherms. In a sense this appears to be a complex fitting procedure of the reference data where the unknown function is the algorithm itself with its parameters.

The choice of parameters used in the previous methods is often heuristically inspired by reasonable assumption regarding the system's microscopic nature, the involved interactions among its elements and additional information such as mesoscopic experimental data. It would be of course useful to find a systematic and reliable method to obtain them from some fine-grained trustworthy method such as Molecular Dynamics, and this is one of the main issues addressed in this thesis work.

In other words, given a certain knowledge of the system at a microscopic level, such as a reliable MD force field, is it possible to obtain a natural coarse-grained model of it? Or on the other hand, having a certain simple model already heuristically designed for mimicking a given system, is there a systematic way of making it realistic by tuning its parameters, starting from microscopic information?

The answer to any such question lies in our ability to recognize the general structure and key elements of the energy landscape associated with the system studied. The idea of studying the energy hypersurface of a microscopic system, which is a function of all its atoms coordinates, has gained increasing attention in the last decade. It is clear that all the information needed to understand and predict the system's behaviour is embedded in this object. However, while any simulation method performs in the end a sampling of this hypersurface, a direct exploration of it is not an easy task, and if the landscape is a rough one, with high barriers, the capability of a method like MD of sampling a significant part of it may be strongly reduced

or lost. In section 3.4 some direct analysis techniques are presented, which allows at least a partial mapping of the relevant critical points of the potential energy surface, however even knowing a significant part of the energy landscape web of minima and transition states, appropriate theoretical tools (such as a proper Transition State Theory, and a reliable way of taking into account entropic effects) are required if one wants to extract useful information from it. Two possible approaches are investigated here: the direct analysis of the energy landscape is the matter of chapter 3, and an implicit way of determining its geography, based on MD trajectories analysis, is exposed in chapter 4, alongside with a second-order Markov model developed on its basis.

The first approach is not particularly useful for the study of systems with weak intermolecular forces like those mainly considered here, but could be a valuable tool in the case of strongly interacting systems with high barriers. However a schematic map of the smooth energy landscapes characterizing our systems is presented that can help to rationalize other results.

The second approach gave interesting insights and brought to the development of a second-order Markov model of the systems studied. The reasons for using a second-order Markov chain approach rather than a first-order one, are discussed in sections 3.4 and 4.2. What is interesting to notice here, is that if a direct coarse-grained markovian model should rely on transition probabilities between regions in the discretized space, in our case calculated from MD trajectories, a second-order model is based on the transition probabilities between first-order transitions. In other words while in the case of a first-order Markov chain a possible system's story is a succession of positions in the discretized space, in the model described here it is considered as a succession of discrete events.

Once the event-event transition matrix is calculated from MD simulation data, it can be studied in itself via standard algebraic analysis, but a different point of view could also be interesting: as any other matrix, it can be seen as a network. In this case it represents a directed weighted network where each node correspond to a discrete event. A given pair of nodes a and b, is connected by a link $a \to b$ if the probability that the event b follows event a is non-null, with a weight proportional to this probability.

The reason for such a change of viewpoint is the expectation that the structure of this network of events depends on the underlying dynamics and could expose some interesting features of it, giving a hint of the basic mechanisms of diffusion, and of the way they change with different conditions, such as temperature, adsorbate density, its size, shape and electronic structure. However the particular kind of network we obtain i.e. both directed and weighted, has not to date received much attention in the impressively active field of network theory. Some of the basic concepts used in network analysis do not have a definite, useful formulation for this class, or are not considered in most of available network analysis packages. Moreover the

kind of system represented by our graph is very different from those usually considered in applications of network theory, so that even those measures that have been defined for directed weighted networks are often not suitable for our purposes. The ideas underlying this network theory approach to the event-event matrix are reported in chapter 5, together with the difficulties met and some related open questions.

1.1 Zeolites

The various ideas presented in this thesis, were all tested and tailored on a reference system: a portion of a zeolite ITQ-29 crystal (often referred to as ZK-4 in literature) hosting molecules of some adsorbate species. Zeolites are aluminosilicates displaying unique characteristics, on which depends their extraordinary variate and extensive utilization in many technological processes. This is the reason of the huge amount of experimental and theoretical studies focusing on these materials in the last decades. What makes

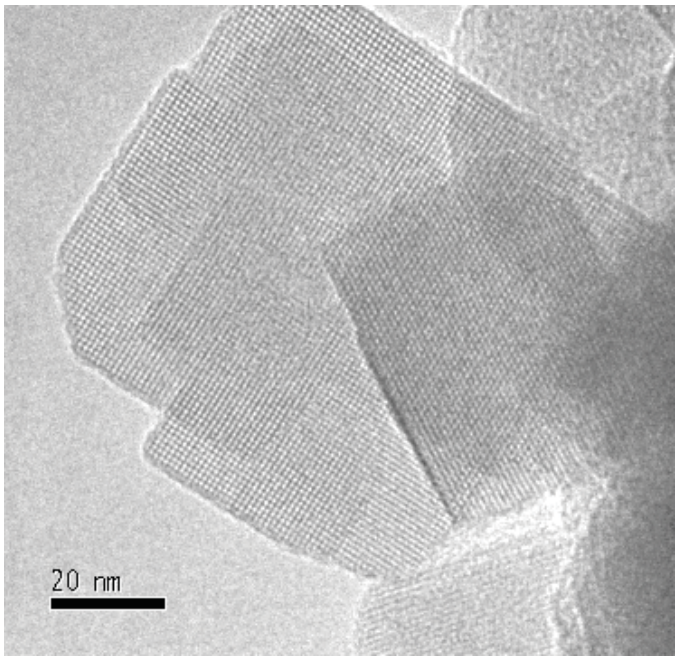


Figure 1.1: *Micro-crystals of zeolite LTA. The small openings of pores are recognizable on the surface.*

zeolites unique is the structure of their crystal lattice, showing nanometric pores and canals, capable of hosting a variety of small molecules, which are allowed to move *inside* the material. The strict confinement experienced by guest molecules, adsorbed on the impressively extended inner surface of the

zeolite affects strongly their behaviour, and is responsible for the catalytic properties of the material, as well as their capability of selectively trapping molecules, depending on their size, shape and polarity.

For this reasons zeolites are widely employed in petroleum industry as catalysts and molecular sieves; as ion traps; as dessicants, for their ability to absorb large amounts of water (the release of water under heating is the origin of the name zeolite, meaning 'boiling stone'); and in many other applications.

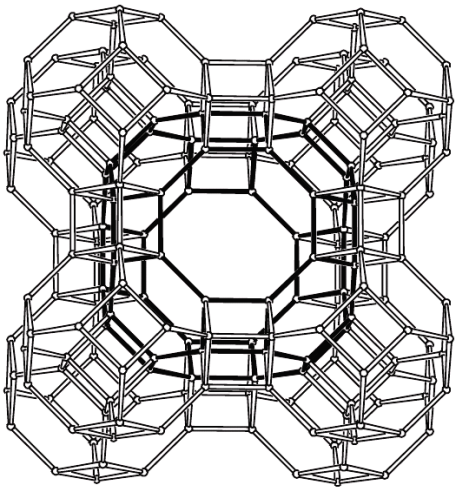


Figure 1.2: *The structure of zeolite ITQ-29 with the $\alpha-$cage highlighted. (Courtesy of IZA [10])*

Our focus was on ITQ-29, an all-silica zeolite of Linde type A (LTA), that is a cubic symmetry one, with neither aluminium nor extra-framework cations. Usually, for zeolites with small Si-Al ratio, cations are present to counter balance the positive charge defect of the lattice due to aluminium. Cubic symmetry and the absence of cations simplify the treatment of this system, but they are not essential in any way for the validity of the following discussion. In Fig. 1.2 a unit cell of ITQ-29 is shown, with the main pore, called α-cage, highlighted. This is the space where molecules resides, while the space inside the truncated octahedra (sodalite-cages) on its vertices, is not reachable due to the narrow openings. The α-cage is roughly 12.5 Å in length, and the octagonal windows, connecting it to the six neighboring cages, have a diameter of 4.8 Å.

Chapter 2

Cellular Automata models of diffusion in zeolites

2.1 Brief overview on Cellular Automata

A Cellular Automaton (CA) [11, 12] is a n-dimensional lattice of *cells*, each characterized by its *state*, with an associated evolution rule. This acts at discrete times on the whole space to generate a new configuration of it, depending only on the configuration at the previous time.

A Cellular Automaton is formally defined as a quadruple $\langle d, \Gamma, J, f \rangle$ where:

-d is a natural number representing the automaton space dimensionality

-Γ is a finite set called *state space*

-J is a finite subset of the integer numbers set Z^d, and is called *neighborhood index*. Being $|J|$ the number of elements in the subset, the cells belonging to the neighborhood of a given cell $\mathbf{i}$ are $\mathbf{i}+\mathbf{j}_1, \mathbf{i}+\mathbf{j}_2, \ldots \mathbf{i}+\mathbf{j}_{|j|}$

-f is a function $f : \Gamma^{|j|} \longrightarrow \Gamma$. Being $n_i(t)$ the state of the $i-th$ cell at time t the evolution rule of the cellular automaton is

$$n_i(t+1) = f(n_{i+j_1}(t), n_{i+j_2}(t), \ldots..n_{i+j_{|j|}}(t)) \tag{2.1}$$

In general a classical CA is:

-*discrete*, in time and space, which consists of an array of individual cells

-*omogeneous*, as it shows the same structure in the whole space and each cell has the same neighborhood

-*parallel*, as the states of every cell are updated simultaneously at each time step (this is valid at least inside each partition in a hierarchic automaton)

-*local*, as the evolution of each cell depends only on the state of its neighborhood (which can contain or not contain the cell itself)

There are several possible variations to the basic scheme previously described, such as:

-asynchronous automata, in which the evolution of some cells is randomly retarded for one or more time steps, and non-homogeneous automata. These can be inhomogeneous in time, obeying a certain evolution rule for some time steps and then switching to another one, or in space, obeying different rules in different regions of their space, possibly with different neighborhood definitions.

-hierarchical automata, with structured cells, each characterized by a state vector rather than a scalar state; and partitioned automata for which only some components of the state vector are relevant to the evolution rule.

-probabilistic automata, for which evolution from a given state may results in many different outcomes each with its own probability of realization. These are obviously very useful for thermodynamic studies.

The cellular automaton described in the next section is a hierarchical, probabilistic CA with a Margolous partition scheme.

2.2 A parallelizable Block Cellular Automaton for the study of diffusion of binary mixtures containing CO_2 in microporous materials.

The present section is an adaption of a paper appeared in The Journal of Chemical Physics [13]. Copiright 2011, American Institute of Physics

We applied a method based on a Block Cellular Automaton algorithm to the study of diffusion of various binary mixtures adsorbed in a model microporous material like zeolite ITQ-29. Our aim was to test the capability of our model to cope with systems in which more than one species is present, using a set of parameters based on heuristic considerations on the Molecular Dynamics results present in literature. A rigorous methodology for the assignment of suitable adsorption energies and diffusion activation barriers for our CA has not been developed yet, nonetheless the results were quite interesting at this stage and we obtained a good qualitative agreement with MD data in literature. The mixtures we investigated contain CO_2, which causes the so called *segregation-effect*, a strong suppression of self-diffusivity of co-adsorbed species. This effect gives rise to relevant problems in the application of some well established and robust methods while our model proved to be able to reproduce both the common features and the segregation anomaly in the trends of diffusion.

Among microporous materials, zeolites constitute a wide class of aluminosilicates displaying very interesting properties due to their microscopic structure. In the past decades their massive industrial utilization for many different applications (i.e. catalysis, molecular sieving, mixtures separation, detergents etc.) has led to a great theoretical and experimental effort in order to understand the underlying mechanisms of diffusive and reactive

processes taking place inside their porous structure. A wealth of Molecular Dynamics and Monte Carlo techniques have been used, which have given deep insights into the rich and various behaviors of these materials, with a particular focus on the intra-crystalline diffusion of adsorbate molecules [1–3]. Even if methods based on MD are invaluable for a detailed knowledge of the microscopic dynamics of these systems, they have severe restrictions both on the time and spatial extension of the simulations. In an ordinary MD experiment one cannot usually expect to study the evolution of systems containing more than about some thousands of atoms and for times longer than tens of nanoseconds, being thus unable to detect a wide range of behaviors, including for instance long range correlation effects on diffusion or the effect of concentration and temperature gradients. It would be in turn very interesting to understand this kind of properties for both practical and theoretical reasons. A possible solution to the aforementioned limitations is that of a strongly *coarse-grained* approach to the problem. This was the idea underlying the previous work of Demontis *et al.* [14–18] where a new Cellular Automaton model was presented. By neglecting most of the microscopic details of the zeolite-adsorbate system and retaining just the essential degrees of freedom needed in order to reproduce in a reliable way the main features of the diffusion processes of interest, this new model gave a good qualitative and quantitative agreement with the data obtained by other authors using MD, but with a strong reduction of CPU time. This means that by giving up to explore the microscopically detailed dynamics of the system, and by exploiting input data coming from other methods, the CA model is able to extend largely the time and spatial horizons of simulation. In the present work we present a slightly modified version of the algorithm proposed in [9, 14–18] and the extension of its application to the study of mixtures diffusion, a possibility that was not already tested, as in our previous works only pure species were considered. The results obtained studying three different mixtures in zeolite ITQ-29 are compared to those present in literature in order to test the reliability of our model and its ability to qualitatively reproduce, at this stage of development, data from well established and more sophisticated methods in three different systems displaying a non-trivial behavior.

2.2.1 The model

Our model consists of a probabilistic, hierarchic Block Cellular Automaton (BCA). The BCA space is a cubic-symmetry lattice of cells, each representing an α-cage of zeolite ITQ-29. Each cell is structured as an array of N_{tot} sites which correspond, at least to a first approximation, to the preferential adsorption sites for guest molecules inside the α-cage. There are two different types of site: N_{ex} exit-sites and N_{in} inner-sites for each cell, which, in the aforementioned approximation, can be seen as corresponding to regions

more and, respectively, less close to the windows connecting the α-cages. The state of each site, is determined by a collection of N_{sp} boolean variables $s^{X_1}, \ldots, s^{N_{\text{sp}}}$. N_{sp} is the number of guest species present in the system and a value of 1 for s^{X_n} means that the given site is occupied by a molecule of the n^{th} species. An exclusion principle holds such that each site can host a maximum of 1 molecule at any given time, so that

$$\sum_X s^X = 0 \vee 1 \tag{2.2}$$

Let us indicate as i any of the N_{tot} sites within each cell. Indicating as $S^X(\mathbf{r}, t)$ the number of molecules of species X in the cell with position $\mathbf{r}$ at time t,

$$S^X(\mathbf{r}, t) = \sum_{i=1}^{N_{\text{tot}}} s_i^X(\mathbf{r}, t), \tag{2.3}$$

where $s_i^X(\mathbf{r}, t)$ is the X-species occupancy of the i-th site of cell $\mathbf{r}$ at time t, we define the total occupancy of a cell as the sum $S(\mathbf{r}, t) = \sum_X S^X(\mathbf{r}, t)$.

All the cells are structurally equivalent so that the grid is homogeneous, and each cell can communicate only with its $2m$ (in this work, $m = 3$) first-neighbouring cells through a set, $\{\mathbf{e}\}$, of $2m+1$ direction listed in Table 2.1. In the present work inner-sites inside each cage are left indistinguishable, while exit-sites are oriented in space in order to link neighbouring cells according to the set $\{\mathbf{e}\}$. In our model, a molecule can leave a cell, say $\mathbf{r}$, to enter a neighbouring cell, say $\mathbf{r} + \mathbf{e}_x$, only by jumping from the particular exit-site of cell $\mathbf{r}$ indicated as $(\mathbf{r}, \mathbf{e}_x)$ to the exit-site $(\mathbf{r} + \mathbf{e}_x, \mathbf{e}_{-x})$ of cell $\mathbf{r} + \mathbf{e}_x$. Therefore, while the inner-sites can be identified by the coordinates $\mathbf{r}$ of the corresponding cell only, every exit-site needs also the specification of one of the $2m$ (in this work, 6) possible orientations specified in Table 2.1.

2.2.2 Block-partitioning scheme

The Automaton acts according to a Margolus block-partition [17, 19, 20]. Each *block* results from the alternative grouping of sites according to a definite rule. Several types of blocks can be defined depending on the number, the type and the connections of the sites they include, which can belong to one or more different cells. The set of all the blocks of the same type represents a *partition* of the whole lattice. A fundamental feature of the blocks is that *blocks belonging to the same partition do not communicate with each other*. In other words, every block is treated as a closed system. Let's indicate as $\mathbf{B}$ a particular block in the system. containing a number $N_{\mathbf{B}}$ of sites generally indicated as $I_1, \ldots, I_{N_{\mathbf{B}}}$. The state of the block $\mathbf{B}$ at time t is defined once the states $s_I^X(\mathbf{B}, t)$ of each site within it are defined. In the present work we consider two types of blocks. The first type is equivalent to a cell, and contains therefore all the N_{tot} sites (*inner* plus *exit*)

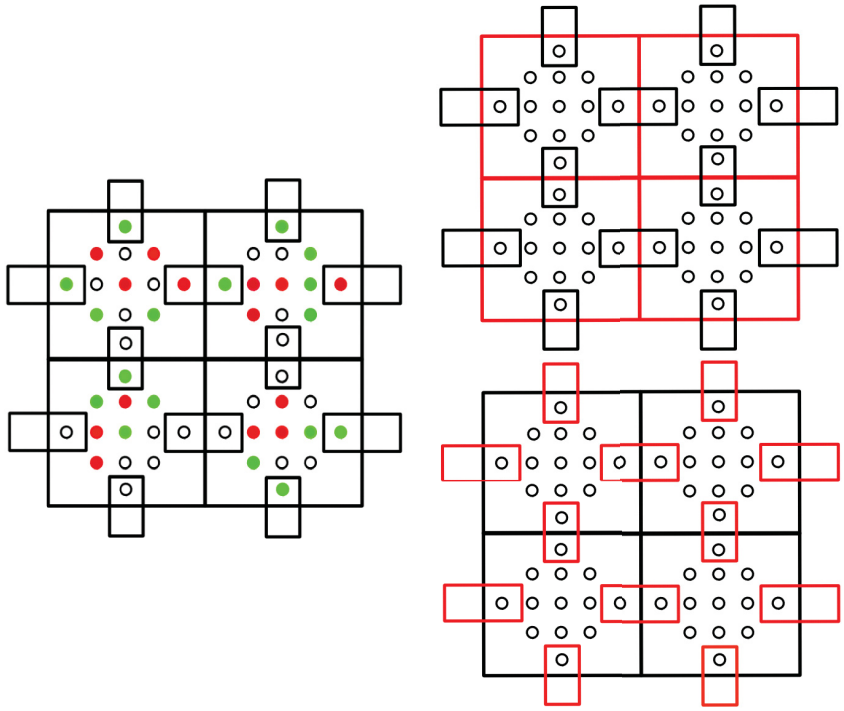


Figure 2.1: *(Left)A bidimensional section of a possible configuration of BCA is shown: squares encase blocks of the 'cage' type while small rectangles encase blocks of the 'window' type. The empty circles represent empty sites while red and green circles represent sites occupied by molecules of two different species respectively. (Right) The two Margolus partitions (in red) of the space of BCA. In the upper part of the figure the cage-blocks partition is shown, which covers the whole automaton space, while the window-blocks partition is shown in the lower part.*

belonging to it. We refer to these blocks as *cage-type blocks*, as they represents α-cages of the zeolite. We emphasize that, since the cells do not overlap and each cell communicates directly with its 6 neighbours with no intermediation of any extra-site besides their respective exit-sites, the first partition (containing all the blocks of the first type) covers the entire lattice. Given the exclusion condition in Eq. (2.2) one has that the total occupancy of a cage-type block cannot be greater than N_{tot}.

The second Margolus partition is the set of all the exit-sites of the system. Every block of the second type contains a pair of communicating exit sites. The type-2 blocks, can be classified in three categories depending on their orientation: an x-block is a block in the x direction, i.e. it includes the exit-sites $(\mathbf{r}, \mathbf{e}_x)$ and $(\mathbf{r} + \mathbf{e}_x, \mathbf{e}_{-x})$ (see Table 2.1). The same scheme is valid for

$\mathbf{e}_i$	$\mathbf{u}_x$	$\mathbf{u}_y$	$\mathbf{u}_z$	i
$\mathbf{e}_x$	λ	0	0	1
$\mathbf{e}_y$	0	λ	0	2
$\mathbf{e}_z$	0	0	λ	3
$\mathbf{e}_{-x}$	$-\lambda$	0	0	4
$\mathbf{e}_{-y}$	0	$-\lambda$	0	5
$\mathbf{e}_{-z}$	0	0	$-\lambda$	6
$\mathbf{e}_0$	0	0	0	$7, \ldots, N_{\text{tot}}$

Table 2.1: *The set of direction vectors for the lattice considered in the present work.* $\mathbf{u}_x$, $\mathbf{u}_y$, *and* $\mathbf{u}_z$ *are the three cartesian unit vectors, and* λ *is meant as the lattice spacing. In the fifth column, the sites within each cell,* i*, assigned to each direction are listed.*

directions y and z. We refer to these blocks as *window-type blocks* and the corresponding partition is the collection of all the windows of the lattice, which partly overlaps with the first partition. From the exclusion condition it results that the occupancy of a window-type block cannot be greater than 2. A schematic representation of the network and its partitions is reported in fig. 2.1

2.2.3 Interactions besides mutual exclusion.

Formally, besides mutual exclusion, interactions among the guest molecules within each cell of the automaton can be defined as follows: if we indicate as n_α^X the number of molecules of species X in the α-type sites of a cell, i.e.

$$n_{\text{ex}}^X = \sum_{i=1}^{6} s_i^X, \qquad n_{\text{in}}^X = \sum_{i=7}^{N_{\text{tot}}} s_i^X, \tag{2.4}$$

then the energy of a molecule of species X residing in a given site is

$$E_{\text{ex}}^X = \varepsilon_{\text{ex}}^X, \qquad E_{\text{in}}^X = \varepsilon_{\text{in}}^X + (n_{\text{in}}^X - 1)\phi^X, \tag{2.5}$$

for an exit and an inner site, respectively. In Eq. (2.5), $\varepsilon_\alpha^X < 0$ is the (invariant) adsorption energy binding a molecule of the species X to a site of type α, and ϕ^X is the molecule-molecule interaction energy within the inner sites (notice that the former equations are not general, but refer to the specifical choice we made in this work (see Section.III), for a general discussion on guest-guest interaction parameters in CA models of zeolites see [9]).

According to Eq. (2.5), only molecules in the inner sites can interact with each other, which is a reasonable approximation since in the particular zeolite framework investigated here (LTA-type), the interactions with the

host medium, for molecules close to the window connecting adjacent cages, can be considered more important than the interaction with the molecules located in the inner part of the cage.

Such an assumption is very convenient for optimization reasons, since the blocks of the window-type block partition can be considered as *independent* one of each other and therefore treated all simultaneously.

We can thus define the Hamiltonian of the system as the sum of the Hamiltonians of all the cells in the system, $\sum_{\mathbf{r}} H(\mathbf{r}, t)$, where

$$H(\mathbf{r}, t) = \sum_X n_{ex}^X(\mathbf{r}, t)\varepsilon_{\text{ex}}^X + \sum_X n_{in}^X(\mathbf{r}, t)\left[\varepsilon_{\text{in}}^X + (n_{\text{in}}^X - 1)\frac{\phi^X}{2}\right], \qquad (2.6)$$

where we remark that X runs over the values $X_1, \ldots, X_{N_{\text{sp}}}$.

An additional migration barrier, $\epsilon_{ij}^X = \epsilon_{ji}^X \geq 0$, can be introduced between two sites, i and j, within the same block, which only affects the model kinetics and not the system Hamiltonian.

2.2.4 Evolution rule.

The Automaton acts on each partition separately, starting from one of them chosen randomly at each time-step. In both partitions we have a synchronous updating of the state of each block indipendently from the state of others. The cage-blocks partition updating accounts for intra-cage dynamics of the host molecules, while within the windows-block partition we are considering inter-cage jumps. The evolution rule mimics the self-diffusion process as a series of activated Arrhenius-jump moves, with activation free energy $-E_\alpha^X$. Within each block all the molecules are invoked in a random sequence. Once a molecule succeeds in freeing itself from the binding site it can move towards any other site within its block. The move will be successful if the target site is free and if another activation barrier ϵ_{ij}^X is overcome. This barrier is related with the steric hindrance encountered by the molecule as it moves from site i to site j and with the distance between them. Once all the blocks in the partition have been updated, the algorithm switch to the other one. This is globally an Arrhenius-jump Monte Carlo scheme taking place for each molecule in every block (see fig. 2.2). Subdivision of the Automaton space in two Margolus blocks is necessary in order to ensure parallelizability of the algorithm. This choice marks the difference between our Cellular Automaton model and other Monte Carlo methods [4, 21, 22]. Using the Margolus approach ensures the evolution of every partition being in principle completely synchronous, as each single block belonging to it can be assigned to a different processor, if one wishes so, given that, with the above described partitioning scheme, no molecule can jump from one block to the other. Inside each singular block of the partition the algorithm is of course sequential.

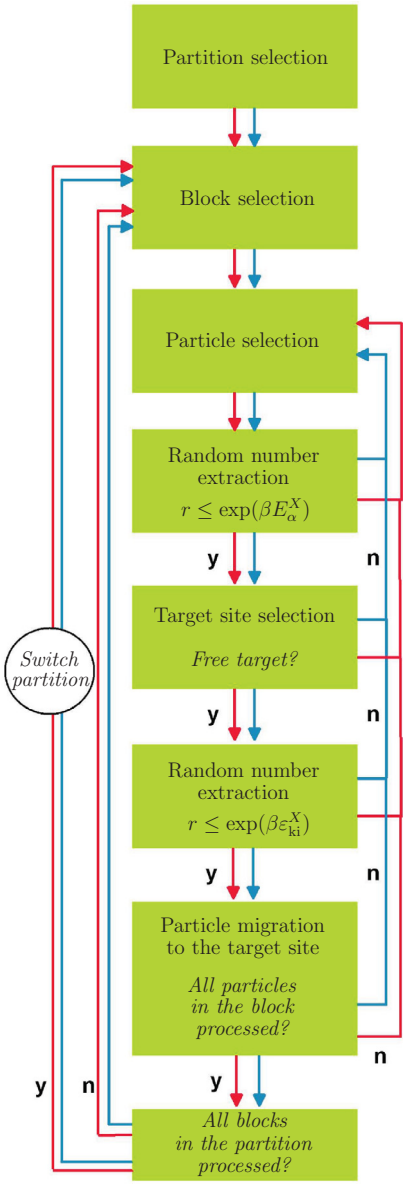


Figure 2.2: The simulation algorithm for a single time step. The scheme followed is the same for both the updating of the partition of the 'cage'-type (red arrows) and of the 'window'-type (blue arrows) blocks. The scheme represents the sequential implementation of the evolution rule used in the present work. However each block can evolve independently from the others. A parallelization of the algorithm is thus straightforward.

It should be pointed out that although the described algorithm is intended to study both statical and dynamical properties, it is unable (unlike classic and parallel kMC methods [7, 23–25]) to generate an autonomous physically meaningful time scale [26, 27]. Nonetheless it gives rise to a precise dynamical hierarchy among the various processes taking place, which is enough in order to reproduce, for instance, the different diffusive behavior of adsorbed species.

Attention should be paid in interpreting the elements of the model we described. Regarding the relationship between adsorption sites in the real zeolite and what we called sites in our automaton model for instance, we notice that the spatial localization of a region of a zeolite pore where interaction with a certain molecule is favorable (*adsorption site*) is not always well defined. In case of large rigid molecules (such as benzene) or if the zeolite shows cations inside the cavities where adsorption takes place, we can speak of real adsorption sites, but in case of small adsorbed molecules and/or *all-silica* zeolites the situation is less defined and the energetic landscape experienced by the molecules is smooth enough as to allow them to spend much of their time outside relative potential minima, being thus largely delocalized, even at room temperature [28]. Thus in our BCA model, sites should be interpreted as collectors of points in the phase space of an adsorbed molecule rather than as localized adsorption sites. The common feature characterizing the ensemble of situations symbolically represented by each site is the likelihood for a molecule within it to jump towards a certain window or not. In particular, inner-sites don't allow host molecules to reach any window, while each exit-site is connected to a particular window.

Some remarks are due concerning the achievement of equilibrium in our model. It is Markovian and it surely satisfies the *balance* condition as the Arrhenius Jump algorithm on which it is based ensures the Boltzmann distribution to be invariant under application of the evolution rule. It is ergodic also — apart from the extreme case of complete saturation of the whole lattice — as there are no traps in the space of configurations of the system, and nothing prevents it from visiting all of its possible states within a sufficiently long time. Following a weaker interpretation of *Deem et al.* [29] we believe that this suffices for our model to be a reliable one. Indeed according to these authors regularity rather than ergodicity should be satisfied. However this constraint is required only in order to avoid loops in the Markov Chain, as these can prevent a system from reaching equilibrium even when it is ergodic. Given the structure of the algorithm, we exclude the possibility that a similar behavior can arise in our model and assume ergodicity to be a sufficient condition. Furthermore if one looks at the BCA timesteps as cluster-moves even *detailed balance* is satisfied, given that every single molecule move is reversible and exactly the same series of moves of a certain timestep t can be performed in the reverse order at $t + 1$. This is because even if the algorithm is sequential, in the sense that no molecule is selected

for a jump twice before all other molecules have been selected, the order of interrogations is random (if interrogations were performed in the same order at every timestep as in usual sequential algorithms, Detailed Balance would be obviously violated [30]).

2.2.5 Simulations

We studied three different binary gas mixtures — CO_2/CH_4, CO_2/N_2, and CH_4/N_2 respectively — adsorbed in zeolite ITQ-29. The first two mixtures exhibit what is called *segregation effect* [31, 32]. This effect is due to the presence of CO_2 and can be related to its strong tendency to reside within windows connecting the α-cages of zeolite ITQ-29 as shown in MD simulations by *Krishna et al.* [31,32]. This fact explains the strong suppression of diffusion observed for the species co-adsorbed with CO_2 as the former experiences a kind of *segregation* within the inner part of micropores due to the crowding of CO_2 molecules at the windows. As an example, self-diffusion for CH_4, when CO_2 is present in equimolar mixture, is one order of magnitude smaller than that of pure CH_4 at the same overall concentration.

The energetic parameters E_α^X for our calculations were chosen on a heuristic basis. Following our idea of qualitatively reproducing the fundamental behavior of mixtures in zeolites — even in a non-trivial case like that of segregation-effect — we assumed the values reported in Tab. 2.2. These represent our mean free-energies guessing for molecules in the two possible situations we are considering in our hierarchical automaton, namely occupation of inner- and exit-sites. Our line of reasoning followed indications regarding the order of magnitude of typical values for small molecules in zeolites found in literature [28,37]. We further made few simple assumptions on the basis of the evidence for preferential adsorption between the inner part of the α-cage and regions closer to the windows. These are based on adsorption isotherms and self-diffusivity *vs.* adsorbate loading plots calculated by means of GCMC [31]. Even if a systematic method for calculating these parameters starting from first principles could be desirable, we have found that the heuristic values used in this work suffice to reproduce the fundamental self-diffusivity trend as obtained from MD calculations present in literature. Similar considerations led to the non-zero choice for the mean field intermolecular interaction for methane in the inner part of a cage $\phi^{CH_4} = 1.1$ kJ/mol : we assumed that the spherical form of methane molecules implies a certain degree of repulsion between them due to sterical hindrance, while the cylindrical N_2 and CO_2 molecules should be able to pack in a more efficient way. Finally the ϵ^X's (Tab. 2.2) were chosen in order to reproduce further potential barriers encountered by molecules while moving between the exit sites across two neighboring cells. In particular we neglected barriers between inner sites, as to a first approximation these are little influential for the overall process of diffusion (even if they could be important in order to

X	$\varepsilon^X_{\text{ex}}$	$\varepsilon^X_{\text{in}}$	ϵ^X(intercage)
CH_4	-3	-13	11
N_2	-7.5	-9.5	6
CO_2	-17	-12	0

Table 2.2: *Guest-host interaction parameters (kJ/mol).*

understand correlation effects due to intra-cage migration). For inter-cage jumps, we assigned the highest barrier value to methane, due to its size, an intermediate value for N_2 and no barrier for CO_2 as it tends to adsorb in the windows thus being able to move between two given supercages by just escaping the adsorption free energy well, which contain the cage-cage boundary.

2.2.6 Results

In figs. 2.3, 2.4 and 2.5 we report the plots for the three equimolar mixtures CO_2/CH_4, CO_2/N_2, and CH_4/N_2, showing the self-diffusivity as a function of total adsorbate loading $\langle n \rangle$ (left) and as a function of the mole fraction of one component at a fixed overall loading (right). These were obtained from simulations at 300 K on a cubic lattice with a side lenght of 9 cells with periodic conditions applied. The self-diffusion coefficients were calculated from the mean square displacement of tagged molecules using the Einstein relation

$$D_s = \frac{1}{6} \lim_{t\to\infty} \frac{d}{dt} \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle \tag{2.7}$$

Each cage-block contains 15 sites, which is in agreement — at least in normal conditions — with the reported maximum loading for methane in α-cages of ITQ-29 [38].

As a test for the model we compared our results to those obtained by Krishna *et al.* from MD simulations and a method based on Maxwell-Stephan theory of diffusion [31, 33–36].

These authors point out that while the MS approach is in good agreement with MD calculations for most mixtures, it fails to reproduce the self-diffusivity trend in case of mixtures displaying segregation effect.

For a clearer comparison we multiplied our values of D_s by a scaling factor in order to obtain the best resemblance with the MD curves. The scaling factor is the same for all three systems. Our results are in good qualitative agreement with MD, reproducing the overall trend of diffusion for the three mixtures, and despite our strongly reductionistic approach it is able to deal with a non trivial behavior like that of segregation, whereas other methods gives very poor quantitative and qualitative agreement.

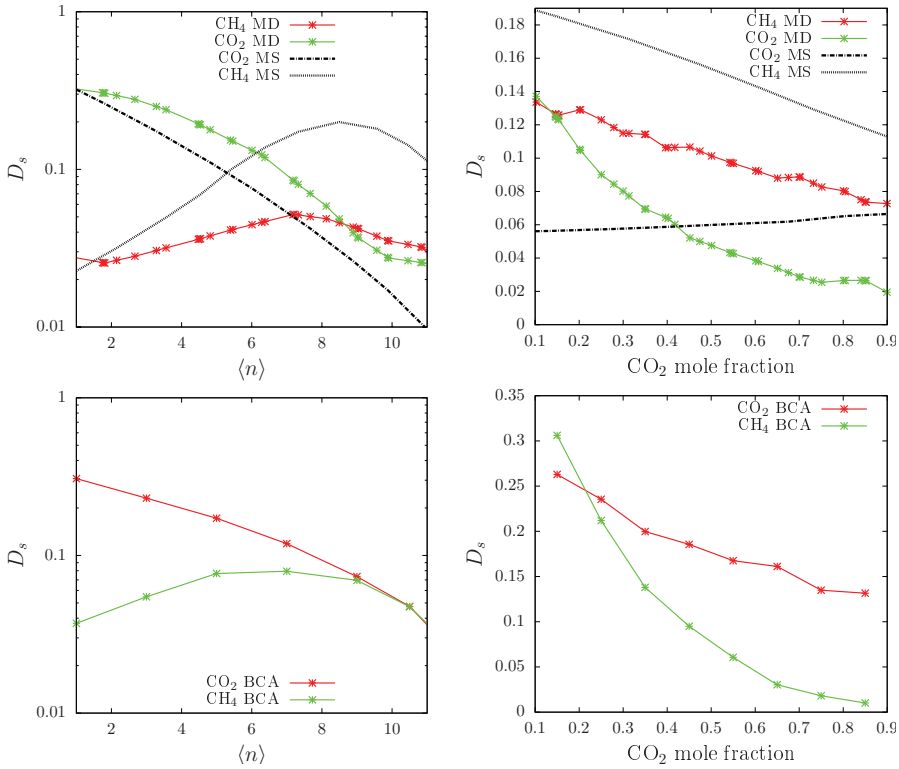


Figure 2.3: *(Left) Self-diffusivity ($D_s/10^{-8}m^2s^{-1}$) of CH_4 and CO_2 as a function of the average number of molecules per cage ($\langle n \rangle$) in equimolar mixture. Results from our model (bottom) are compared with those obtained by mean of MD and Maxwell-Stefan theory (MS) by Krishna et al. (top). (Right) Self-diffusivity ($D_s/10^{-8}m^2s^{-1}$) of CH_4 and CO_2 as a function of the mole fraction of CO_2 at $\langle n \rangle = 5.5$. Results from our model (bottom) are compared with those obtained by mean of MD and Maxwell-Stefan theory (MS) by Krishna et al. (top).*

2.2.7 Conclusions

We tested our Cellular Automaton method in the case of binary mixtures diffusing in zeolite ITQ-29 and compared the results with MD data, showing its ability to cope with more than one diffusing species giving good qualitative agreement with this well established and fully microscopically detailed technique. We stress that these results are based on heuristically determined parameters aiming mainly to explore the range of possibilities of the method and, nonetheless, in the case of non-trivial behavior such as the segregation-effect, we found that the CA gives a better agreement with MD than the robust and successful MS method used by Krishna and coworkers. It would be obviously desirable to have a rigorous algorithm for defining the

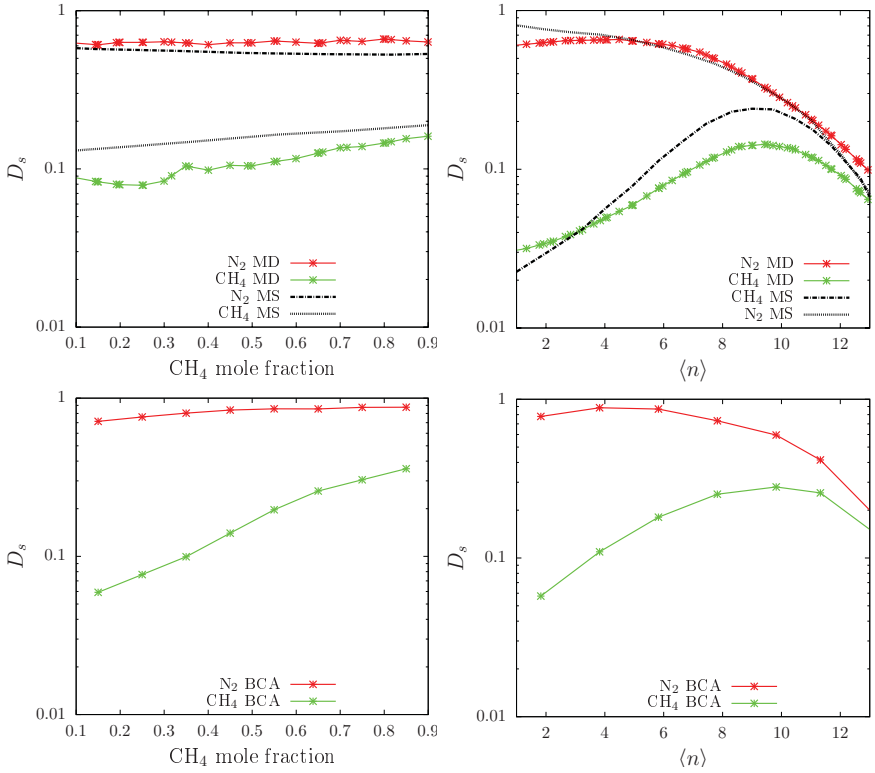


Figure 2.4: *(Left) Self-diffusivity ($D_s/10^{-8}m^2s^{-1}$) of CH_4 and N_2 as a function of the mole fraction of CH_4 at $\langle n \rangle = 5.5$. Results from our model (bottom) are compared with those obtained by means of MD and Maxwell-Stefan theory (MS) by Krishna et al. (top). (Right) Self-diffusivity ($D_s/10^{-8}m^2s^{-1}$) of CH_4 and N_2 as a function of the average number of molecules per cage ($\langle n \rangle$) in equimolar mixture. Results from our model (bottom) are compared with those obtained by mean of MD and Maxwell-Stefan theory (MS) by Krishna et al. (top).*

right input parameters for the Automaton, extracting them from atomistic simulation. But even at this stage our tests demonstrate that the model is able to display a range of interesting behaviors and to capture on a coarse-grained level the basic mechanisms of diffusion in zeolites. Validation via comparison with other theoretical approaches is important as this justify a further effort intended to exploit the intrinsic parallel nature of CAs for a drastic extension of the space and time scales usually accessible to the available simulation techniques for diffusion in zeolites. Using parameters obtained from MD and MC studies as an input, our BCA model is a promising tool for a future investigation of zeolite and ZIFs membranes and whole micro-crystal, both in and out of equilibrium.

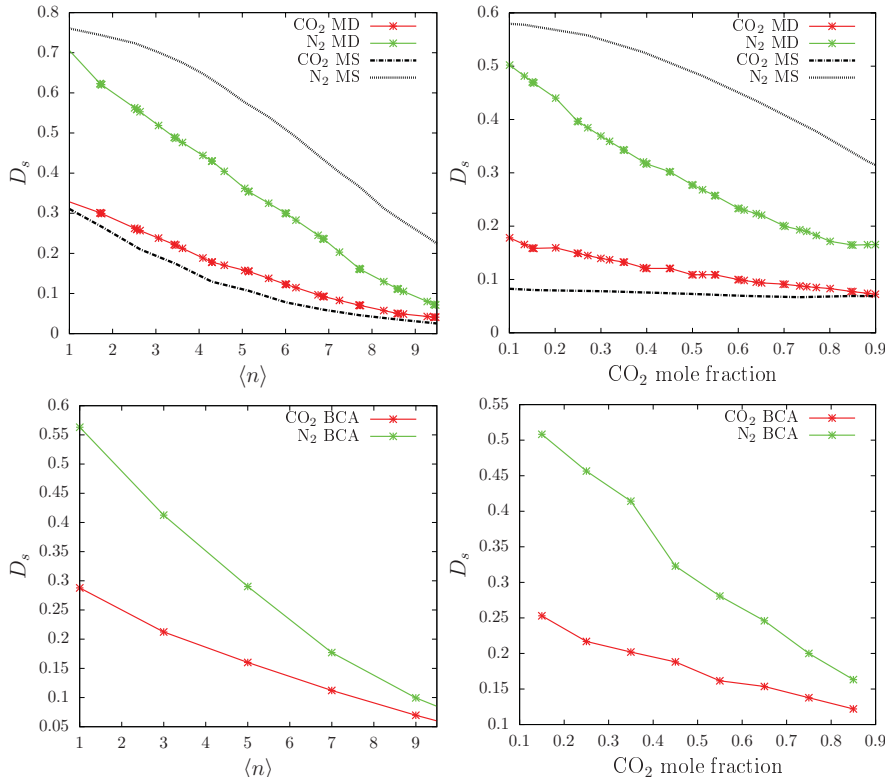


Figure 2.5: *(Left) Self-diffusivity ($D_s/10^{-8}m^2s^{-1}$) of CO_2 and N_2 as a function of the average number of molecules per cage ($\langle n \rangle$) in equimolar mixture. Results from our model (bottom) are compared with those obtained by mean of MD and Maxwell-Stefan theory (MS) by Krishna et al. (top). (Right) Self-diffusivity ($D_s/10^{-8}m^2s^{-1}$) of CO_2 and N_2 as a function of the mole fraction of CO_2 at $\langle n \rangle = 5.5$. Results from our model (bottom) are compared with those obtained by means of MD and Maxwell-Stefan theory (MS) by Krishna et al. (top).*

2.3 The Central Cell Model: a mesoscopic hopping model for the study of the displacement autocorrelation function.

The present section is an adaptation of a paper to which I collaborated, appeared in The Journal of Chemical Physics [40]. Copyright 2011, American Institute of Physics.

On the mesoscale, the molecular motion in a microporous material can be represented as a sequence of hops between different pore locations and from one pore to the other. On the same scale, the memory effects in the motion of a tagged particle are embedded in the displacement autocorrelation

function (DACF), the discrete counterpart of the velocity autocorrelation function (VACF). In this paper a mesoscopic hopping model, based on a lattice-gas automata dynamics, is presented for the coarse-grained modeling of the DACF in a microporous material under conditions of thermodynamic equilibrium. In our model, that we will refer to as *Central Cell Model*, the motion of one tagged particle is mimicked through probabilistic hops from one location to the other in a small lattice of cells where all the other particles are indistinguishable; the cells closest to the one containing the tagged particle are simulated explicitly in the canonical ensemble, whereas the border cells are treated as mean-field cells in the grand-canonical ensemble. In the present paper numerical simulation of the Central Cell Model are shown to provide the same results as a traditional lattice-gas simulation. Along with this a mean-field theory of self-diffusion which incorporates time-correlations is discussed.

The diffusive motion of molecules in a generic medium is usually affected by memory effects introduced by their interactions with each other and with the medium itself. This is especially true when the diffusing molecules are subjected to the confining action of a microporous material like a zeolite. [2, 41] In particular, the narrow windows of certain microporous materials can make the guest's diffusion profile (i.e., diffusivity *vs.* concentration at constant temperature) very different from what expected for the motion in a bulk phase as well as in any less strongly confining material.

Although the discreteness of the network of channels and cages of regular microporous materials suggests immediately an analogy with lattice-gas models, there is still no 'definitive' coarse-grained, lattice simulation method for molecules in zeolites which is able to play as a *cheaper* mesoscale version of classical Molecular Dynamics (MD). Several approaches are available depending on what specific properties of the host-guest system the simulator is interested in. As an example, Kinetic Monte Carlo (kMC) simulations are suitable for all the dynamical properties which do not explicitly involve correlations among different particles [25, 42, 43] (e.g. the self diffusion coefficient), whereas thermodynamic models can be successfully adopted for the study of static equilibrium properties (e.g. adsorption isotherm and local density distribution).

Due to their intrinsecally synchronous nature, the class of Lattice-Gas Cellular Automata (LGCA) can be thought of as the ideal candidate for a mesoscopic simulation of the collective properties. On the other hand, as a drawback of their synchronicity traditional LGCAs are much more difficult to handle than standard Monte Carlo (MC) models are. This makes it a hard task to *surely* achieve thermodynamic equilibrium,i.e. preserving both detailed balance and synchronicity, in the presence of explicit particle-particle interactions. To solve such a conflict, a partitioning technique has been proposed in our previous work, aimed to couple the LGCA computational framework with local MC (balanced) moves. [9, 14–18] The idea

underlying the resulting Partitioning Cellular Automaton (PCA), inspired by a heterogeneous model for surface diffusion by Chvoj *et al.* [44], is that the peculiar cage-to-cage dynamics of molecules under tight confinement is well-represented in a model lattice with heterogeneous adsorption locations inside each cage. According to this representation, in each zeolite cage we distinguish two types of locations: those close to the exit windows, termed *exit sites*, and the rest of the cage pictured instead as a set of *inner sites*. The exit sites in each cage are then access points to the neighboring cages, and differ from the inner sites in their statistical weight (i.e. the probability of being occupied). As recently confirmed by other simulation studies, [45] splitting the single cells into differently weighted locations provides a qualitatively correct mesoscopic representation of the problem (See Fig. 2.6).

Even though more work has still to be done to make cellular automata the 'definitive' environment for meso-simulations in micropores, our PCA approach captures many important aspects of adsorption and diffusion in zeolites, such as realistic (i.e. closely resembling those developed in MD simulations) density distribution, fluctuations, and *time-correlations*. Concerning the single-particle diffusion process (at arbitrary concentration), the *backscattering effect*, [46] a major source of time-correlation causing the self-diffusivity to be less than what expected, can be properly mimicked in the PCA approach since it allows the amount of memory lost in each cell during a single time step to be tuned.

Thus, our PCA can be taken as a starting point for further developments in many directions. The one explored in this work is the realization of a further simplified coarse-grained simulation of the hopping process of a tagged particle in a confined lattice system, where all the other guest particles are moving as well but they are kept indistinguishable. Our aim is to reproduce the memory effects affecting the particle motion in the PCA at the minimum cost possible. The strategy is to make the tagged particle 'feel' an environment very close to the one it would have experienced in the full automaton simulation. Since the model is constructed in such a way that the host cell of the tagged particle always results to be located exactly in the middle of the system, we called it *Central Cell Model* (CCM).

The lengthy PCA simulation of a large system is thus reduced to a small set of connected cells, a limited neighborhood of whose is simulated by the lattice-gas evolution rule in the canonical ensemble while the border cells are treated as mean-field cells. In any case, the CCM approach cannot be taken as substitutive of a full lattice-gas simulations. Collective dynamic properties, self-organization, and long-range phenomena arising in non-equilibrium conditions cannot be simulated directly through a CCM implementation of a lattice-gas rule. This approach is limited to the reproduction of the correlated motion of a *single* particle in a lattice-gas at arbitrary loading (i.e. concentrations of guest particles, also known as *coverage*), but under conditions of thermodynamic equilibrium, strictly local

μ	chemical potential
β	inverse temperature
$K_{\rm ex}, K_{\rm in}, K$	exit, inner, and total site number per cell
$\mathbf{s}$	micro-configuration of indistinguishable particles in a single cell
$n_{\rm ex}, n_{\rm in}, n$	exit sites, inner sites, and total occupancy of a single cell
$\mathbf{n} = (n_{\rm ex}, n_{\rm in})$	meso-configuration of the cell
$f^o_{\rm ex}, f^o_{\rm in}$	exit and inner site free-energy of adsorption in a singly-occupied cell (site deepness)
$\phi_{\rm ex}(n), \phi_{\rm in}(n)$	exit and inner site free-energy contribution due to the mutual interaction of n particles
$F(\mathbf{n}), F(\mathbf{s})$	cell free energy
$F^o(\mathbf{n}), F^o(\mathbf{s})$	cell free energy (non-interacting part)
$\Phi(\mathbf{n}), \Phi(\mathbf{s})$	cell free energy (interacting part)
$\epsilon_{\rm ki}(n, m)$	kinetic barrier to intercell migration from an n- into an m-occupied cell
C_{ab}	probability of targeting the site b from departure site a during randomization
$p_{\rm jump}$	acceptance probability for a single randomization jump
$\overline{s}_b\kappa(\mathbf{n}, \mathbf{m})$	acceptance probability for a jump from a cell with meso-conf. $\mathbf{n}$ into exit site b of a cell with meso-conf. $\mathbf{m}$
$p(\mathbf{n})$	equilibrium probability of a cell to be meso-configured as $\mathbf{n}$

Table 2.3: A list of the basic quantities involved in a numerical simulation with the Central Cell Model.

interactions, and absence of long-range correlations. When one or more of such conditions are not fulfilled, or if also the collective dynamics produced by some evolution rule need to be investigated, then a full lattice-gas simulation will be unavoidable to obtain reliable results. Nevertheless, the above mentioned conditions are fulfilled in many lattice-gas simulations of short-range interacting particles, so that for those cases the CCM will be the quickest way to retrieve the correct self-motion properties. This is of primary interest when, for example, one wishes to model the entity of memory effects in the single-particle motion (e.g. to mimic the diffusive behavior of some reference system in coarse-grained modeling) and therefore needs to check quickly how a particular setup of the parameters will affect the resulting diffusion isotherm.

The construction of the CCM version of a lattice-gas rule is a really direct way to uncover the basic mechanisms by means of which the tagged particle

preserves memory of its previous moves in time. In fact, it is straightforward to pass from the numerical CCM to a mean-field representation of the tagged particle's diffusion process at arbitrary loading, inclusive of the time-correlations. In this work, the CCM approach will be used to develop an approximated theory of self-diffusion for a lattice-gas automaton rule, based on a mathematical formulation of the displacement autocorrelation function (DACF), i.e. the key function embedding the memory effects of a generic diffusion process on the mesoscopic scale. The DACF plays the same role the velocity autocorrelation function (VACF) plays in atomistic simulations, although being more easily accessible for theoretical analysis. Earlier studies on LGCA emphasized the central importance of the discrete VACF in both the formulation of efficient computational schemes for the evaluation of transport properties, and the understanding of the entire self-diffusion process. [47–49] In the present case, the analysis of the DACF (we do not call it VACF since, differently from traditional LGCAs, in our approach no proper *velocity* vector is associated with the cell-to-cell migration) will lead to a closed mathematical formulation for the self-diffusion coefficient.

After a brief resumé of the lattice-gas model, the Central Cell Model will be presented. Then, we will describe the probabilistic analysis of the DACF leading to the mean-field formulation of self-diffusivity. Results of numerical tests will be presented throughout the paper, and discussed in a separated section.

2.3.1 Local randomization and propagation.

Here we will briefly outline the basic operations of the original automaton model. The interested reader can find a very detailed description in a previous work on this subject. [16, 18] The basic quantities that will be explicitly used in a simulation with the Central Cell Model are listed in Table 2.3.

In our approach, particles move within a three-dimensional network of structured points called *cells*. A single cell and a small cluster of connected cells of the automaton are sketched in Fig. 2.6a and 2.6b respectively. The total number of particles in the system, N, and the temperature, T (and so the inverse temperature, $\beta = (k_B T)^{-1}$ with k_B the Boltzmann's constant), are held fixed. The concentration $\langle n \rangle$ of the diffusing species in the lattice, termed *loading*, is the average number of particles per cell and is obtained just by dividing N by the total number of cells. Every cell is a discrete representation of a zeolite cage. It is made of K_{ex} exit sites and K_{in} inner sites, and every site can be free or singly occupied, thus giving a saturation occupancy of $K = K_{\mathrm{ex}} + K_{\mathrm{in}}$. As can be seen from Fig. 2.6b, every pair of neighboring cages are interfaced by a pair of connected exit sites. The system evolves in discrete time steps. Guest molecular species are represented via point particles whose migration mechanism at each iteration is performed in two substeps: a *randomization* changes the configuration of

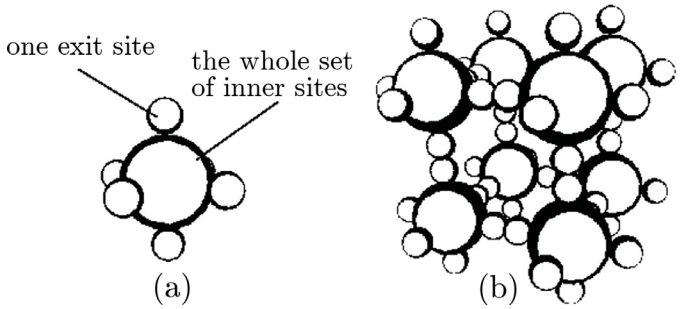


Figure 2.6: *A three-dimensional sketch of (a) a single cell, and (b) a small cluster of connected cells of the automaton. Every cell is representative of a single zeolite cage. When looking at the single cell, (a), small spheres represent the* exit sites*, i.e. the locations closest to the cage-to-cage connections in a real zeolite (e.g. an LTA zeolite), whereas the big sphere, named* inner site*, represents the set of all the remaining locations.*

guest particles on every cell according to a probabilistic scheme, and a *propagation* allows the particles in the exit sites to attempt to move into the respective neighboring cages.

The actual micro-configuration of (indistinguishable) particles in each cell has a primary importance, and is denoted as

$$\mathbf{s} = \{s_1, s_2, \ldots, s_K\}, \tag{2.8}$$

where the first K_{ex} and the next K_{in} entries are respectively the occupancies of the exit and of the inner sites (i.e. $s_i = 1$ if the i-th site of the cell is occupied, and 0 if empty). The cell occupancies are defined as the *exit site*, the *inner site*, and the *total* cell occupancies:

$$n_{\text{ex}} = \sum_{i=1}^{K_{\text{ex}}} s_i, \qquad n_{\text{in}} = \sum_{i=K_{\text{ex}}+1}^{K} s_i, \qquad n = n_{\text{ex}} + n_{\text{in}}. \tag{2.9}$$

Exit and inner site cell occupancies makes a *meso-configuration* of the cell, termed $\mathbf{n} = (n_{\text{ex}}, n_{\text{in}})$.

The static properties of each cell are determined by the adsorption (negative) free energy associated to every site, f^o_{ex} and f^o_{in} (also referred to as exit- and inner-site *deepness*), the actual cell occupancy n (i.e. the total number of particles in the cell), and an occupancy-dependent interaction term for every type of site, $\phi_{\text{ex}}(n)$, and $\phi_{\text{in}}(n)$. These parameters define the cell free energy function:

$$F(\mathbf{n}) = F^o(\mathbf{n}) + \Phi(\mathbf{n}), \tag{2.10}$$

with

$$F^o(\mathbf{n}) = n_{\text{ex}} f^o_{\text{ex}} + n_{\text{in}} f^o_{\text{in}}, \tag{2.11}$$

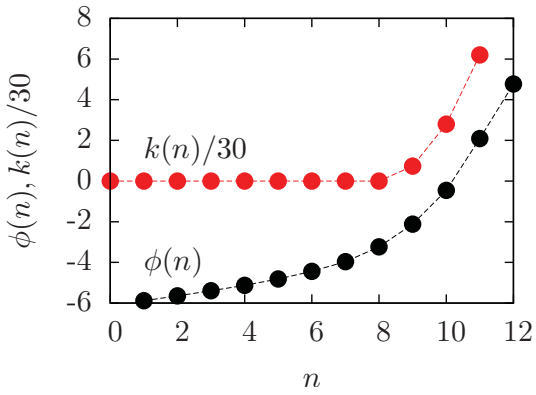


Figure 2.7: *The interaction parameter, $\phi(n)$ (in kJ mol^{-1}, defined in Section 2.3.1), for $0 \leq n \leq K-1$, and the parameter $k(n)$ (defined in Section 2.3.3) for the numerical tests we performed in this work. In all the simulations, $K_{\text{ex}} = K_{\text{in}} = 6$.*

and

$$\Phi(\mathbf{n}) = n_{\text{ex}}\phi_{\text{ex}}(n) + n_{\text{in}}\phi_{\text{in}}(n). \tag{2.12}$$

In the numerical simulation we performed as a test for the model, our choice for the interaction parameters was $\phi_{\text{ex}}(n) = \phi_{\text{in}}(n) = \phi(n)$, with a smoothly increasing repulsive effect as the occupancy increases. (see Fig. 2.7). We set the number of exit and inner sites as $K_{\text{ex}} = K_{\text{in}} = 6$.

Fixed adsorption free-energy f^o_{ex} and f^o_{in} take alternatively the values -50 and -40 kJ mol^{-1} in the various simulations.

2.3.2 Randomization.

The randomization can be performed in two ways. The *memoryless randomization* changes probabilistically the actual configuration of every cell while preserving its occupancy according to the probability distribution $P(n_{\text{ex}}|n)$ defined as

$$P(n_{\text{ex}}|n) = \binom{K_{\text{ex}}}{n_{\text{ex}}}\binom{K_{\text{in}}}{n-n_{\text{ex}}} e^{-\beta F(n_{\text{ex}}, n-n_{\text{ex}})}. \tag{2.13}$$

which is exactly the probability of an n-occupied cell to have n_{ex} particles in the exit sites (and consequently $n_{\text{in}} = n - n_{\text{ex}}$ in the inner sites). Such a choice causes no memory of the previous configuration(s) to be conserved (apart from the cell occupancy n, which is conserved).

In the *jump randomization* scheme instead all the n particles are invoked in a random sequence, and every particle is asked to perform a jump towards a randomly selected target site within the same cell. Therefore, the cell configuration is changed here in n steps instead of one (as it was for the

memoryless scheme). To illustrate the algorithm, let us take a single cell and store the identities of the n particles inside of it in the vector $\mathbf{I} = (I_1, \ldots, I_n)$. Let us then randomize the entries of $\mathbf{I}$, thus obtaining the random sequence of identities $\mathbf{I}^R = (I_1^R, \ldots, I_n^R)$. At this point, the following chain of jump events is realized:

$$\mathbf{s}^{(0)} \to \mathbf{s}^{(1)} \to \mathbf{s}^{(2)} \to \cdots \to \mathbf{s}^{(n)}, \tag{2.14}$$

where by definition $\mathbf{s}^{(0)} := \mathbf{s}$ is the first configuration of the chain, and $\mathbf{s}^{(k)}$ is the actual micro-configuration when the particle of identity I_k^R is invoked. Let us consider a transition $\mathbf{s}^{(k)} \to \mathbf{s}^{(k+1)}$ where $\mathbf{s}^{(k)}$ and $\mathbf{s}^{(k+1)}$ are two consecutive configurations in the chain (2.14). In this transition, the k-th particle in the random sequence of particles jumps from its departure site, say a, to the target site b chosen with a probability C_{ab}. The probability of such a jump to happen is then

$$\begin{aligned} p_{\text{jump}}(\mathbf{s}^{(k)} \to \mathbf{s}^{(k+1)}) =& C_{ab}\overline{s}_b^{(k)}\gamma e^{\beta f_a^o} \\ & \times e^{\beta\left\{\Phi(\mathbf{s}^{(k)}) - \max\left[\Phi\left(\mathbf{s}^{(k)}\right), \Phi\left(\mathbf{s}^{(k+1)}\right)\right]\right\}}, \end{aligned} \tag{2.15}$$

where $\overline{s}_b^{(k)}$ is the *non*-occupancy of the target site b in the actual micro-configuration $\mathbf{s}^{(k)}$, i.e. $\overline{s}_b^{(k)} = 1 - s_b^{(k)}$, and γ is a normalization constant aimed to further control the particles' mobility during randomization (this will affect correlations as well). In our simulations, we put $\gamma = \exp\{-\beta \max(f_{\text{ex}}^o, f_{\text{in}}^o)\}$. Such an algorithm preserves some memory of the previous configuration, since in the case of half/high cell occupancy n, the (locally) sequential jump criterion constrains the configuration not to vary too much in the chain shown in (2.14).

A few words about the choice for C_{ab}. In order to preserve detailed balance, it preferably should be symmetric, that is forward and reverse jumps should be chosen with the same probability. It is interesting to introduce several kinds of constraints (without violating symmetry) in the configuration path during randomization, to study their effects on correlations, and to check to which extent they can be predicted by a mean-field theory of diffusion. As an example we could decide, during randomization, to allow every particle to target any site with the same probability $1/K$, this giving a C matrix with all entries like

$$C'_{ab} = \frac{1}{K}, \qquad a, b \in [1, K], \tag{2.16}$$

or we could choose all targetings from an exit site towards a different exit site to be rejected. This would force the particles to spend some time in the inner site before changing direction of intercell migration. It would result in a C matrix like

$$C''_{ab} = \begin{cases} 0, & \text{if } a, b \in [1, K_{\text{ex}}] \text{ and } a \neq b \\ \dfrac{1}{K}, & \text{otherwise} \end{cases} \tag{2.17}$$

In the present work we will refer to the case of $C = C'$ in Eq. (2.16) as 'allowed ex-ex jumps', and to the case of $C = C''$ in Eq. (2.17) as 'forbidden ex-ex jumps'.

2.3.3 Propagation.

Once randomization changed the internal configuration of every cell independently one of the other (while preserving the cell occupancies), the propagation operation allow the cells to exchange the particles in their exit sites with their respective neighbors. In order to keep working with locally balanced Monte Carlo moves the propagation must be applied to every *pair* of communicating cells. Since some pairs can overlap, not all the pairs can be invoked at the same time. This is because of local interactions among the host-molecules of a given cell giving rise to different intercell migration barriers, depending on the loading of both departure and target cell. Therefore, either they have to be invoked in a random sequence, or they can be grouped into *partitions*, each containing the maximum possible number of non-overlapping pairs. Such a partitioning scheme, [16] originally known as Margolus' Neighborhood [19, 20] allows no conflict to arise during such a substep.

At every pair, the two cells communicate through two adjacent exit sites, say a and b. Provided a particle to be in a and site b to be empty, a jump from a to b is accepted with a probability $\kappa(\mathbf{n}, \mathbf{m})$ where the departure and destination cell are meso-configured respectively as $\mathbf{n}$ and $\mathbf{m}$:

$$\kappa(\mathbf{n}, \mathbf{m}) = \frac{\gamma e^{\beta f^o_{\rm ex}} e^{-\beta \epsilon_{\rm ki}(n,m)}}{1 + e^{\beta \Delta\Phi(\mathbf{n},\mathbf{m})}}, \tag{2.18}$$

where $n = n_{\rm ex} + n_{\rm in}$ and $m = m_{\rm ex} + m_{\rm in}$ are the actual occupancies of the departure and the target cell respectively, the quantity

$$\begin{aligned} \Delta\Phi(\mathbf{n}, \mathbf{m}) =& \Phi(n_{\rm ex} - 1, n_{\rm in}) + \Phi(m_{\rm ex} + 1, m_{\rm in}) \\ & - \Phi(n_{\rm ex}, n_{\rm in}) - \Phi(m_{\rm ex}, m_{\rm in}) \end{aligned} \tag{2.19}$$

is the difference in interaction free-energy between the outcoming and the incoming configuration of the pair of cells, and $\epsilon_{\rm ki}(n, m)$ is the kinetic barrier to intercell migration, given as the intersection energy, for $0 \leq x \leq 1$, between the two harmonics

$$E_{\rm dep}(x) \quad = \tfrac{1}{2} k(n-1) x^2 \tag{2.20}$$

for the departure cell, and

$$E_{\rm arr}(x) \quad = \tfrac{1}{2} k(m) (x-1)^2 \tag{2.21}$$

for the arrival cell. [17] The trend assigned to the parameter $k(n)$ in the numerical simulation performed in this work is quadratically increasing at the highest loading, as shown in Fig. 2.7.

2.3.4 Jumps and time-correlations.

Numerical simulations [18] have shown that correlation effects can be modeled (or excluded, if wanted) in our PCA. While every application of the memoryless randomization described in Sec. 2.3.1 pushes each cell straightforwardly towards a condition of local equilibrium, via an abrupt collective move, the configuration changes occurring by means of the jump randomization are much less marked, and slow down strongly the evolution towards equilibrium. This is because the output configurations available in the jump randomization are *much less* than in the memoryless randomization, thus causing memory effects to show up spontaneously as the system evolves in time.

Let us illustrate this in more details. The definition of *configuration*, **s**, given in Eq. (2.8) in Section 2.3.1 contains no information regarding the identity of the guest particles. In words, such a kind of identity-less configuration will be referred to as '*s*-configuration'.

Particles identities will be taken into account by the following σ-configuration instead:

$$\sigma(\mathbf{r}) = \{\sigma_{iI}\}, \qquad i = 1, \ldots, K \text{ and } I = 1, \ldots, N \tag{2.22}$$

where N is the number of guests, and σ_{iI} has value 1 if the I-th guest of the system is located at the i-th site of cell $\mathbf{r}$, and 0 otherwise. We will consider now a single, closed cell with configuration σ just *before* a randomization operation. The *memoryless* randomization will determine the output configuration by choosing it out of the whole set of $\Omega_\sigma = K!/(K-n)!$ possible arrangements of distinguishable particles in the cell. The *jump* randomization instead constrains the configuration path from σ to σ^R within a set of necessarily similar configurations, so that the number of possible output configurations, say $\Omega_\sigma^{\text{jump}}$, results smaller than Ω_σ with a discrepancy increasing as the cell occupancy increases. Such a discrepancy is the very origin of the memory effects in the self-diffusivity in the automaton [18] as well as in a host-guest system in general. Ideally, one should perform an infinite number of *jump* randomization cycles per time step in order to suppress it.

An analysis of the migration mechanism in the automaton will help a deeper understanding of the correlations introduced by the jump randomization. A low-cost study of correlations in the motion of a tagged particle induced by the local environment is the task of the Central Cell Model that we are about to introduce in Sec. 2.3.5 for the case of a discrete jump model.

2.3.5 The Central Cell Model.

In the model we present here, the lattice is constituted by (see Fig. 2.8)

(i) A *central cell*.

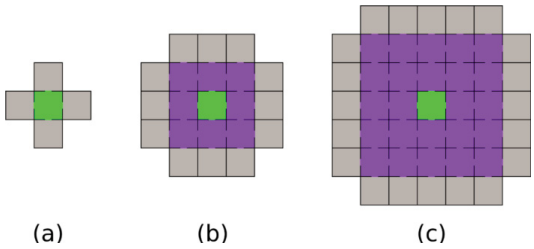


Figure 2.8: *The lattice space of the Central Cell Model. The central cell (in green), hosting the tagged particle, and the cells in the core shells (violet) are simulated through the prescribed lattice-gas rule in the canonical ensemble. The external cells (gray), instead, are mean-field. They maintain the whole system at thermodynamic equilibrium and work as a reservoir of particles coming in/out of the border core cells. Broken cell-to-cell boundaries are meant as cell-to-cell links. Figs. (a), (b), and (c) differ in the number of core shells, which is* $L^{\mathrm{sh}} = 0$ *in (a),* $L^{\mathrm{sh}} = 1$ *in (b), and* $L^{\mathrm{sh}} = 2$ *in (c).*

(ii) A finite number, N^{sh}, of cells surrounding the central one, organized into L^{sh} *shells*. Central cell and surrounding shells constitute the *core* of the system.

(iii) A casing of N^{mf} border *mean-field* cells enclosing the core. Mean-field cells are small grand-canonical systems, working for the core cells as a reservoir of particles and keeping the whole system in equilibrium at the desired value of chemical potential.

The cell-to-cell connections are established as follows: every cell in the core is connected with all the available first-neighboring cells in the system, so that if we consider a cubical arrangement of cells (so as to mimic the LTA zeolite topology, as an example) every of the core cells is then connected to six first-neighbors. Core cells need not to be connected with each other only:

cells at the borders of the core happen to have one or more mean-field cells in their neighboring list. Every of the mean-field cells instead are supposed to be connected with one cell at the border of the core only. No connection is assumed to exist between mean-field cells.

Since the mean-field cells exchange particles with an ideal reservoir, a chemical potential, μ, has to be selected first. This gives access to the absolute probability, $p(\mathbf{n})$, of a meso-configuration $\mathbf{n}$ defined as

$$p(\mathbf{n}) = [\Xi(\mu)]^{-1} \binom{K_{\mathrm{ex}}}{n_{\mathrm{ex}}} \binom{K_{\mathrm{in}}}{n_{\mathrm{in}}} e^{\beta\mu n} e^{-\beta F(\mathbf{n})}, \tag{2.23}$$

where the normalization factor $\Xi(\mu)$ is the grand-canonical partition function of a single cell:

$$\Xi(\mu) = \sum_{\mathbf{n}} \binom{K_{\text{ex}}}{n_{\text{ex}}} \binom{K_{\text{in}}}{n_{\text{in}}} e^{\beta\mu n} e^{-\beta F(\mathbf{n})}. \tag{2.24}$$

Occupancies n_{ex}, n_{in} and n in Eqs. (2.23) and (2.24) are meant as the occupancies of the exit sites, the inner sites, and the whole cell respectively when the meso-configuration is $\mathbf{n}$. Such a notation will be used throughout the whole paper.

The average occupancy (often referred to as the *loading*) is then $\langle n \rangle = \sum_{\mathbf{n}} np(\mathbf{n})$.
The probability distribution in Eq. (2.23) will be used to update the state of the mean-field cells at each time iteration.

Generating the initial configuration. The initial configuration is constructed by randomly assigning each cell a meso-configuration according to the distribution $p(\mathbf{n})$ (see Eq. (2.23)). Such a meso-configuration is then converted into a micro-configuration $\mathbf{s}$ of indistinguishable particles, randomly chosen out of those satisfying the meso-configuration itself. Whereas not needed by the other cells, the central cell must contain at least one particle, that will be 'tagged' thus allowing us to follow its dynamical path.

Time evolution.

Once the initial configuration of the system is ready, the system evolves in discrete time steps, $t_0, t_0 + \tau, t_0 + 2\tau, \ldots$, each of physical duration τ (see our previous work [16] for a discussion about the time step). At each time step (say, t):

(i) A jump randomization is performed at each cell.

(ii) The pairs of connected cells are chosen in a random sequence, and a propagation operation is performed at every pair. Until now, the whole lattice has preserved its total number of particles.

(iii) The move performed by the tagged particle is stored. If it has left the central cell, then the system has to be re-centered so that the newly occupied cell becomes the central cell. Such an operation is performed by simply transforming the coordinates of all the cells. If the tagged particle made a cell-to-cell jump, then the coordinates of the cells are transformed as follows:

$$\mathbf{r}(t + \tau) = \mathbf{r}(t) - \delta\mathbf{r}(t), \tag{2.25}$$

where $\delta\mathbf{r}(t)$ is the distance vector between the arrival and the departure cell. Due to this operation, the mean-field cells happening to fall outside of the lattice space are destroyed, whereas those resulting not configured at all will be assigned a new configuration in the next operation.

(iv) The mean-field cells are randomly assigned a new micro-configuration according to the same procedure of generation of the initial one (applied to the mean-field cells only though).

The update strategy described above is sketched in Fig. 2.9. In Fig. 2.10 we compare the self-diffusivity resulting from a simulation of the canonical $9 \times 9 \times 9$ lattice-gas with the one computed from a CCM simulation on an increasing number L^{sh} of shells around the central cell.

We can see clearly that increasing L^{sh} improves the matching between the two types of simulations, and that two shells are enough to obtain a reasonable agreement.

2.3.6 Analysis of the self-diffusion process: the displacement autocorrelation function

The mean-field analysis is carried on in terms of the possible jump sequences a tagged guest can perform during the diffusion process, treated as a Markov chain, where jumps are meant as site-to-site migrations and can be categorized into (i) jumps within the same cell, and (ii) jumps between neighboring cells.

$\mathbf{e}_0 = (0,0,0)$
$\mathbf{e}_1 = (\lambda,0,0)$ $\quad \mathbf{e}_2 = (0,\lambda,0)$ $\quad \mathbf{e}_3 = (0,0,\lambda)$
$\mathbf{e}_4 = (-\lambda,0,0)$ $\quad \mathbf{e}_5 = (0,-\lambda,0)$ $\quad \mathbf{e}_6 = (0,0,-\lambda)$

Table 2.4: The set of direction vectors (cubic lattice).

Each jump category has a certain probability to occur which is *dependent* on the actual position of the guest itself and of the surrounding particles. Due to the complexity of such a multi-body problem, a mean-field approach must be used to derive readable equations linking correlations in the self-motion to some macroscopic quantities (e.g. densities, total transfer rates, etc.).

A dynamical quantity of major importance in the analysis of the diffusion process in a mesoscopic lattice is the *instantaneous cell-to-cell displacement*, $\delta\mathbf{r}(t)$, of the tagged guest at time t, introduced in Eq. (2.25). The instantaneous displacement can take values in the set of lattice vectors $\{\mathbf{e}_j\}$, with $j = 0, \ldots, K_{\mathrm{ex}}$, listed in Table 2.4 for the case (considered in this work) of a cubic lattice.

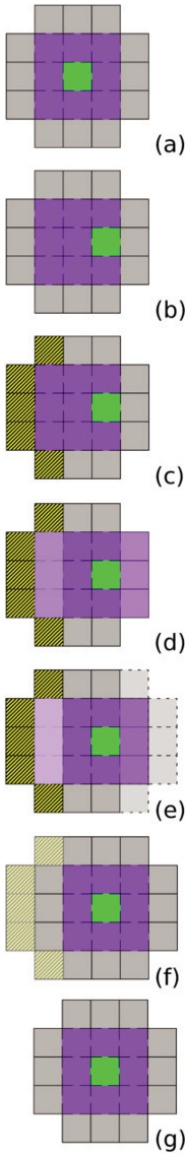


Figure 2.9: *A schematic representation of the update strategy when the particle leaves the central cell to move in the right neighboring cell (a,b). In (c) the mean-field cells at the very left are destroyed. Then (d) the* core *cells at the very left become mean-field, whereas the mean-field cells at the right retain their actual configurations and enter the new core. Finally (e) new mean-field cells are created from scratch at the very right to complete the mean-field casing, the proper cell-to-cell connections are established (f), and the system is ready to undergo the next randomization-propagation cycle (g).*

The displacement autocorrelation function (DACF), given by $\langle \delta\mathbf{r}(z\tau) \cdot \delta\mathbf{r}(0) \rangle$ (where $z \geq 0$ is an integer and τ is the duration of a time step),

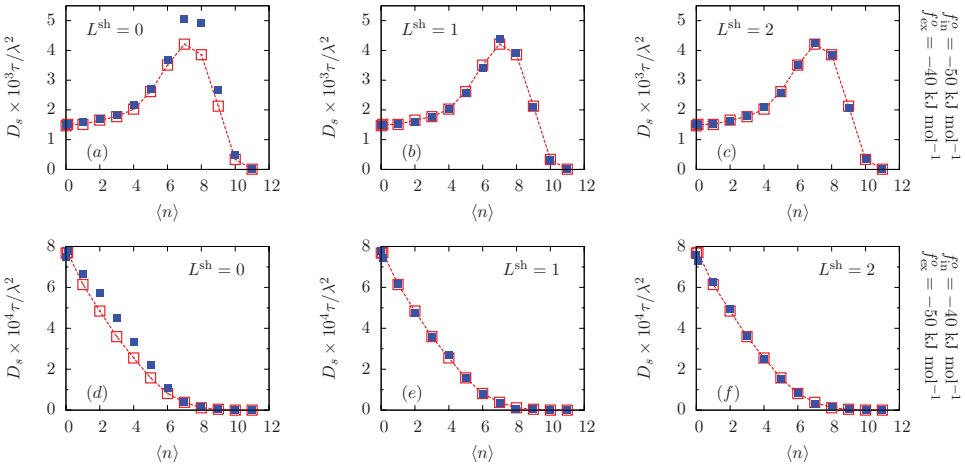


Figure 2.10: *The self-diffusivity, D_s, resulting from numerical simulations (in the canonical ensemble) of the traditional lattice-gas automaton model for a* closed *test system of $9 \times 9 \times 9$ cells, in comparison with the results of (grand-canonical) simulations of the Central Cell Model with increasing number of shells L^{sh}. In the first row, the inner sites have been set as deeper than the exit sites, and vice-versa in the second row. Ex-ex jumps are allowed.*

correlates in time the cell-to-cell displacements. It is related to the self-diffusivity via the Green-Kubo formula: [18]

$$D_s = \frac{1}{2d\tau}\left[\langle \delta\mathbf{r}(0)\cdot\delta\mathbf{r}(0)\rangle + 2\sum_{z=1}^{\infty}\langle \delta\mathbf{r}(z\tau)\cdot\delta\mathbf{r}(0)\rangle\right]. \tag{2.26}$$

where $d = 3$ is the number of dimensions of a cubic lattice. The peculiarity of the DACF in a regular lattice is that it is strictly connected to the jump probability. It is the aim of this Section to reconstruct the terms appearing in Eq. (2.26) starting from the list of the possible movements of the tagged particle.

Contribution at the initial time.

First of all the contribution at $t = 0$

$$D_0^{mf} = \frac{1}{2d\tau}\langle \delta\mathbf{r}(0)\cdot\delta\mathbf{r}(0)\rangle, \tag{2.27}$$

that is the uncorrelated diffusivity, proportional to the DACF at time zero, turns out to be also proportional to the escape probability of the guest from the host cell. The escape event will be indicated with the symbol $\Diamond$. In terms of the randomization-propagation dynamics, such an event can be rewritten as:

◊ The guest reaches any of the K_{ex} exit sites of the current cell during randomization, and then the propagation step lets it migrate to the corresponding neighboring cell during propagation.

Since at the initial time $\delta\mathbf{r}(0)\cdot\delta\mathbf{r}(0)$ equals λ^2 if the guest migrates to a neighboring cell and 0 otherwise, then Eq. (2.27) can be rewritten as

$$D_0^{\mathrm{mf}} = \frac{1}{2d}\frac{\lambda^2}{\tau}p(\lozenge), \tag{2.28}$$

where $p(\lozenge) = \lambda^{-2}\langle\delta\mathbf{r}(0)\cdot\delta\mathbf{r}(0)\rangle$. is the escape probability.

Contribution after one iteration: a probabilistic interpretation of the normalized DACF.

Now, let us suppose that at time zero the particle escaped its host cell along a generic *non-null* direction $\mathbf{e}_j$ picked out of the set of direction vectors, listed in Table 2.4 for a cubic lattice. This is the starting point for the listing of all the subsequent events along with their respective probabilities, represented as a Markov Chain. In this approach the choice of a (hyper)cubic topology turns out to be the most convenient, since $\delta\mathbf{r}(t')\cdot\delta\mathbf{r}(t)$ is non-zero if and only if the displacements at the times t and t' are parallel and non-null. More specifically, it is positive if the displacement direction are the same, and it is negative if they are equal but opposite. Therefore the normalized DACF, $\langle\delta\mathbf{r}(z\tau)\cdot\delta\mathbf{r}(0)\rangle/\langle\delta\mathbf{r}(0)\cdot\delta\mathbf{r}(0)\rangle$, represents the conditional probability of a guest to migrate at time $z\tau$ in the same direction of displacement at time 0, given that at time 0 the displacement was not null, *minus* the conditional probability of a migration in the opposite direction.

We will proceed now with the listing of the basic in-cage and cage-to-cage jump events at the time $t = \tau$, given a successful propagation at the previous time. Every event will be associated a symbol, ς, taking values in the following set

$$S = \{\Rightarrow, \rightarrow, \Leftarrow, \leftarrow, \Updownarrow, \updownarrow, \bigcirc\}, \tag{2.29}$$

meaning respectively, for a given direction of motion (say the x axis), ($\Rightarrow$) successful and ($\rightarrow$) unsuccessful step forward, ($\Leftarrow$) successful and ($\leftarrow$) unsuccessful step backwards, ($\Updownarrow$) successful and ($\updownarrow$) unsuccessful step out of the direction of motion, $\bigcirc$ no attempt of leaving the cell.

The main approximation in the mean-field analysis is a factorization of the joint probability, $p(\lozenge,\varsigma)$, of an escape event ($\lozenge$) followed by the event ς at the next time step:

$$p(\lozenge,\varsigma) = p(\lozenge)p(\varsigma|\lozenge) \tag{2.30}$$

For the sake of clarity, in the list that follows we will give a short description of the events mentioned in Eq. (2.29). Those events are also sketched in Fig. 2.11.

$\Rightarrow$ A step forward. The randomization moves the particle from the exit site into the opposite one. After this, the propagation is successful and the particle migrates in the corresponding neighboring cell. This happens with conditional probability $p(\Rightarrow|\Diamond)$.

$\Leftarrow$ A backscattering event. At the end of randomization the particle find itself in the same exit site it entered by the event $\Diamond$. The propagation is successful and the particle jumps back into the cell it occupied before event $\Diamond$. (Conditional probability: $p(\Leftarrow|\Diamond)$).

$\Updownarrow$ A change of direction. The particle performs a migration jump whose direction is not parallel to the direction of the jump performed during the event $\Diamond$. (Conditional probability: $p(\Updownarrow|\Diamond)$).

$\bigcirc$ The guest reaches an inner site of the current cell during randomization. (Conditional probability: $p(\bigcirc|\Diamond)$).

Single arrows, i.e. $\rightarrow$, $\leftarrow$, and $\updownarrow$, differ from $\Rightarrow$, $\Leftarrow$, and $\Updownarrow$ respectively in the fact that the propagation event is unsuccessful. Let us now introduce the quantity $\chi(\varsigma|\Diamond)$, returning a value 1 if the cell-to-cell displacement at time t (represented by the symbol ς) has equal sign of the displacement at time 0, a value -1 if the sign is opposite, and 0 in all other cases:

$$\chi(\varsigma|\Diamond) = \begin{cases} 1, & \text{if } \varsigma = \Rightarrow \\ -1, & \text{if } \varsigma = \Leftarrow \\ 0, & \text{otherwise} \end{cases} \tag{2.31}$$

Therefore, since the process is Markovian one can define

$$\langle \delta\mathbf{r}(\tau) \cdot \delta\mathbf{r}(0) \rangle = \lambda^2 p(\Diamond) \sum_{\varsigma \in S} \chi(\varsigma|\Diamond) p(\varsigma|\Diamond) \tag{2.32}$$

where the set S has been defined in Eq. (2.29), which returns

$$\langle \delta\mathbf{r}(\tau) \cdot \delta\mathbf{r}(0) \rangle = \lambda^2 p(\Diamond) \left[p(\Rightarrow|\Diamond) - p(\Leftarrow|\Diamond) \right]. \tag{2.33}$$

Contribution after several iterations.

Since we are assuming the migration process to be Markovian, the conditional migration probabilities for $t = 2\tau$ will depend only on the outcome at time $t' = \tau$. Relations between the conditional probabilities after two steps and those after one step are listed in Table 2.5. It should be noted that a guest starting from an inner site or from an exit site not pointing towards the direction $\mathbf{e}_j$ nor $-\mathbf{e}_j$ will have equal probability to reach those sites during randomization.

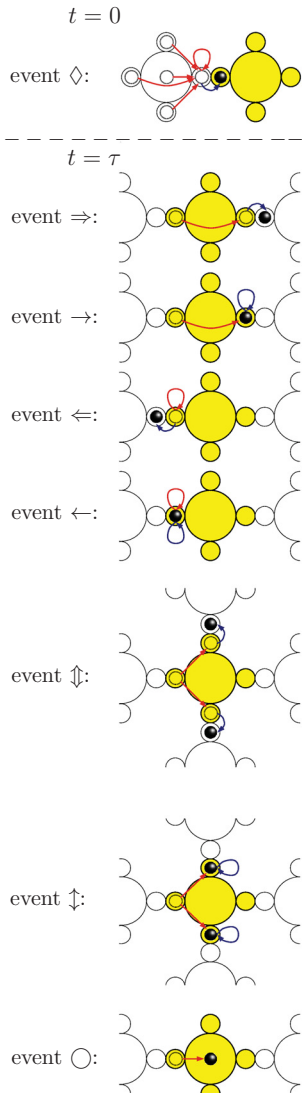


Figure 2.11: *A graphical 2-d representation of the main events contributing to the diffusive motion of a single particle in the automaton. The events pictured here for times $t = \tau$ are assumed to echange their role in time with the event for $t = 0$ according to a Markov chain. For each event, the black 3-d sphere represent the actual position of a tagged particle (other guest particles eventually present are omitted), while the empty circles represent its possible(s) position(s) at the immediately preceding timestep. Red and blue arrows represent respectively the possible randomization and propagation outcomes.*

$p(\varsigma \mid \Rightarrow)$	
$p(\Rightarrow \mid \Rightarrow) = p(\Rightarrow \mid \Diamond)$	$p(\rightarrow \mid \Rightarrow) = p(\rightarrow \mid \Diamond)$
$p(\Leftarrow \mid \Rightarrow) = p(\Leftarrow \mid \Diamond)$	$p(\leftarrow \mid \Rightarrow) = p(\leftarrow \mid \Diamond)$
$p(\Updownarrow \mid \Rightarrow) = p(\Updownarrow \mid \Diamond)$	$p(\updownarrow \mid \Rightarrow) = p(\updownarrow \mid \Diamond)$
$p(\bigcirc \mid \Rightarrow) = p(\bigcirc \mid \Diamond)$	
$p(\varsigma \mid \rightarrow)$	
$p(\Rightarrow \mid \rightarrow) = p(\Leftarrow \mid \Diamond)$	$p(\rightarrow \mid \rightarrow) = p(\leftarrow \mid \Diamond)$
$p(\Leftarrow \mid \rightarrow) = p(\Rightarrow \mid \Diamond)$	$p(\leftarrow \mid \rightarrow) = p(\rightarrow \mid \Diamond)$
$p(\Updownarrow \mid \rightarrow) = p(\Updownarrow \mid \Diamond)$	$p(\updownarrow \mid \rightarrow) = p(\updownarrow \mid \Diamond)$
$p(\bigcirc \mid \rightarrow) = p(\bigcirc \mid \Diamond)$	
$p(\varsigma \mid \Leftarrow)$	
$p(\Rightarrow \mid \Leftarrow) = p(\Leftarrow \mid \Diamond)$	$p(\rightarrow \mid \Leftarrow) = p(\leftarrow \mid \Diamond)$
$p(\Leftarrow \mid \Leftarrow) = p(\Rightarrow \mid \Diamond)$	$p(\leftarrow \mid \Leftarrow) = p(\rightarrow \mid \Diamond)$
$p(\Updownarrow \mid \Leftarrow) = p(\Updownarrow \mid \Diamond)$	$p(\updownarrow \mid \Leftarrow) = p(\updownarrow \mid \Diamond)$
$p(\bigcirc \mid \Leftarrow) = p(\bigcirc \mid \Diamond)$	
$p(\varsigma \mid \leftarrow)$	
$p(\Rightarrow \mid \leftarrow) = p(\Rightarrow \mid \Diamond)$	$p(\rightarrow \mid \leftarrow) = p(\rightarrow \mid \Diamond)$
$p(\Leftarrow \mid \leftarrow) = p(\Leftarrow \mid \Diamond)$	$p(\leftarrow \mid \leftarrow) = p(\leftarrow \mid \Diamond)$
$p(\Updownarrow \mid \leftarrow) = p(\Updownarrow \mid \Diamond)$	$p(\updownarrow \mid \leftarrow) = p(\updownarrow \mid \Diamond)$
$p(\bigcirc \mid \leftarrow) = p(\bigcirc \mid \Diamond)$	
$p(\varsigma \mid \Updownarrow)$	
$p(\Rightarrow \mid \Updownarrow) = p(\Rightarrow \mid \updownarrow)$	$p(\rightarrow \mid \Updownarrow) = p(\rightarrow \mid \updownarrow)$
$p(\Leftarrow \mid \Updownarrow) = p(\Leftarrow \mid \updownarrow)$	$p(\leftarrow \mid \Updownarrow) = p(\leftarrow \mid \updownarrow)$
$p(\Updownarrow \mid \Updownarrow) = p(\Updownarrow \mid \updownarrow)$	$p(\updownarrow \mid \Updownarrow) = p(\updownarrow \mid \updownarrow)$
$p(\bigcirc \mid \Updownarrow) = p(\bigcirc \mid \updownarrow)$	
$p(\varsigma \mid \updownarrow)$	
$p(\Rightarrow \mid \updownarrow) = p(\Leftarrow \mid \updownarrow)$	$p(\rightarrow \mid \updownarrow) = p(\leftarrow \mid \updownarrow)$
$p(\varsigma \mid \bigcirc)$	
$p(\Rightarrow \mid \bigcirc) = p(\Leftarrow \mid \bigcirc)$	$p(\rightarrow \mid \bigcirc) = p(\leftarrow \mid \bigcirc)$

Table 2.5: Possible guest jumps after two time steps for the case where during the jump randomization each guest can select any of the K sites in the cell as target sites.

This means that when the starting position is $\bigcirc$, or $\updownarrow$, or $\Updownarrow$, the net average displacement is null. Therefore, only the moves $\Rightarrow, \rightarrow, \Leftarrow, \leftarrow$ do contribute in the general formula for the mean-field DACF:

$$\begin{aligned}\langle \delta\mathbf{r}(z\tau) \cdot \delta\mathbf{r}(0)\rangle =& \lambda^2 p(\Diamond) \sum_{\varsigma_1 \in S} \cdots \sum_{\varsigma_z \in S} \chi(\varsigma_z|\Diamond) p(\varsigma_1|\Diamond) \\ & \times \prod_{j=1}^{z-1} p(\varsigma_{j+1}|\varsigma_j), \qquad z \geq 1 \end{aligned} \tag{2.34}$$

Therefore, general mean-field expressions can be formulated for both the

DACF and the self-diffusivity, Eq. (2.26):

$$\langle \delta\mathbf{r}(z\tau)\cdot\delta\mathbf{r}(0)\rangle = \lambda^2 p(\Diamond)\Big[p(\Rightarrow|\Diamond)-p(\Leftarrow|\Diamond)\Big]\left\{p(\Rightarrow|\Diamond)-p(\Leftarrow|\Diamond)-\Big[p(\rightarrow|\Diamond)-p(\leftarrow|\Diamond)\Big]\right\}^{z-1}. \tag{2.35}$$

$$D_s^{\mathrm{mf}} = \frac{\lambda^2}{d\tau}p(\Diamond)\left\{\frac{1}{2}+\frac{p(\Rightarrow|\Diamond)-p(\Leftarrow|\Diamond)}{1+p(\rightarrow|\Diamond)-p(\leftarrow|\Diamond)-\Big[p(\Rightarrow|\Diamond)-p(\Leftarrow|\Diamond)\Big]}\right\}. \tag{2.36}$$

Eq. (2.36) is a quite general approximated equation. The terms in it can be obtained straight from a numerical simulation of the Central Cell Model. One can proceed as follows: for evaluating $p(\Diamond)$ it is enough to store the number of cell-to-cell jumps, $N_\Diamond$, of the tagged particle, and then dividing it by the number of time iterations (say, N_τ):

$$p_{\mathrm{num}}(\Diamond) = \frac{N_\Diamond}{N_\tau}, \tag{2.37}$$

where the subscript 'num' denotes that the quantity has been evaluated from a numerical simulation.

For evaluating the conditional probability, it will be enough to store the jump direction every time the tagged particle performs a cell-to-cell jump. At the next time

(i) if the particle performs another jump in the same direction as before, the quantity $N_\Rightarrow$ is increased by one,

(ii) if the particle fails a jump attempt towards the same direction as before, the quantity $N_\rightarrow$ is increased by one,

(iii) if the particle performs a jump towards the *opposite* direction, then the quantity $N_\Leftarrow$ is increased by one,

(iv) if the particle fails a jump attempt towards the opposite direction, then the quantity $N_\leftarrow$ is increased by one.

Then the conditional probabilities are obtained as

$$p_{\mathrm{num}}(\varsigma|\Diamond) = \frac{N_\varsigma}{N_\Diamond}, \qquad \varsigma\in\{\Rightarrow,\rightarrow,\Leftarrow,\leftarrow\} \tag{2.38}$$

Results of the numerical mean-field evaluation of Eq. (2.36) will be compared with the self-diffusivity obtained by explicit calculation of the DACF from the output of the simulations in the Results and Discussion section.

2.3.7 Mean-field DACF: Theoretical prediction of self-diffusivity.

In this Section we derive an approximate mean-field expression for the DACF. We will first apply the general mean-field DACF formula in Eq. (2.35) to the limiting case of infinite dilution. Then, we will propose further approximations to apply Eqs. (2.35) and (2.36) to the case of diffusion at arbitrary loading.

2.3.8 Exact DACF in the limit of infinite dilution.

When the motion of a lone particle in an empty system is considered, correlations with the motion of other particles are absent and an exact mathematical formula for the DACF can be written. In this limit the migration probability during propagation if the particle stays in an exit site is

$$J_{\rm prop} = \frac{1}{2}\gamma e^{\beta[f^o_{\rm ex} - \epsilon_{\rm ki}(1,0)]} \tag{2.39}$$

and $p(\Diamond)$ is given by

$$p(\Diamond) = p_{\rm ex} J_{\rm prop}, \tag{2.40}$$

where

$$p_{\rm ex} = \frac{K_{\rm ex} e^{-\beta f^o_{\rm ex}}}{K_{\rm ex} e^{-\beta f^o_{\rm ex}} + K_{\rm in} e^{-\beta f^o_{\rm in}}} \tag{2.41}$$

is the equilibrium probability of the lone particle to occupy an exit site. The other terms in Eq. (2.35) can be determined by properly weighting every possible randomization jump. They are listed in Table 2.6 for both the case of allowed and forbidden ex-ex jumps [i.e., use of $\{C'_{ab}\}$ or $\{C''_{ab}\}$ matrix, Eqs. (2.16) and (2.17), during the randomization procedure].

Allowed ex-ex jumps
$p(\Rightarrow \vert \Diamond) = (1/K) J_{\rm ex-ex} J_{\rm prop}$
$p(\rightarrow \vert \Diamond) = (1/K) J_{\rm ex-ex} (1 - J_{\rm prop})$
$p(\Leftarrow \vert \Diamond) = (1 - \gamma_{\rm ex} - \gamma_{\rm in}) J_{\rm prop}$
$p(\leftarrow \vert \Diamond) = (1 - \gamma_{\rm ex} - \gamma_{\rm in}) (1 - J_{\rm prop})$
Forbidden ex-ex jumps
$p(\Leftarrow \vert \Diamond) = [1 - \gamma_{\rm in}] J_{\rm prop}$
$p(\leftarrow \vert \Diamond) = [1 - \gamma_{\rm in}] (1 - J_{\rm prop})$

Table 2.6: Probability values for events of jump starting from initial condition $\Diamond$ at time 0 for the case of jump randomization with allowed (upper part) and forbidden (lower part) ex-ex jumps, where $\gamma_{\rm ex} = [(K_{\rm ex} - 1)/K] J_{\rm ex-ex}$ is the probability of the guest to jump into an exit site different from the departure one, and $\gamma_{\rm in} = (K_{\rm in}/K) J_{\rm ex-in}$ is the probability to jump to an inner site.

In the infinite dilution limit the quantities $J_{\mathrm{ex-ex}}$ and $J_{\mathrm{ex-in}}$ mentioned in the formulas of Table 2.6 have the same value:

$$J_{\mathrm{ex-ex}} = J_{\mathrm{ex-in}} = J_{\mathrm{ex}} := \gamma e^{\beta f^{o}_{\mathrm{ex}}} \tag{2.42}$$

Since its value depends only on the departure (exit) site, we simply called it J_{ex}.

$$\begin{aligned}\lim_{\langle n\rangle\to 0} \langle \delta\mathbf{r}(z\tau)\cdot\delta\mathbf{r}(0)\rangle = &-\lambda^2 p_{\mathrm{ex}} J^2_{\mathrm{prop}} \left(1-2J_{\mathrm{prop}}\right)^{z-1} \\ &\times \left(1-J_{\mathrm{ex}}\right)^{z},\end{aligned} \tag{2.43}$$

for the case of allowed ex-ex jumps, and

$$\begin{aligned}\lim_{\langle n\rangle\to 0} \langle \delta\mathbf{r}(z\tau)\cdot\delta\mathbf{r}(0)\rangle = &-\lambda^2 p_{\mathrm{ex}} J^2_{\mathrm{prop}} \left(1-2J_{\mathrm{prop}}\right)^{z-1} \\ &\times \left(1-\frac{K_{\mathrm{in}}}{K}J_{\mathrm{ex}}\right)^{z},\end{aligned} \tag{2.44}$$

for the case of forbidden ex-ex jumps We remark that Eq. (2.43) is independent of the number of exit/inner sites in the cell, while Eq. (2.44), where jumps between different exit sites are forbidden, shows an explicit dependence on the number of sites constituting the cell.

Therefore the accessibility of the adsorption sites plays a fundamental role in determining the entity of correlations.

2.3.9 Approximated mean-field DACF and self-diffusivity at arbitrary loading.

At arbitrary loadings the tagged particle is likely to share its host and neighboring cells with other particles. This means that, during randomization, the variety of sequences in which the particles can be invoked to attempt a jump have an effect on the probability of the tagged particle to reach an exit site, as well as they affect the tendency of the cell to keep memory of its previous configurations from time to time. Since we are interested in improving our understanding of the self-diffusion process by obtaining a *readable* equation,

(i) We will treat as a mean-field the other guests sharing the cell with the tagged particle. That is, we assume that when the tagged guest is invoked to attempt a jump during the randomization process, the other guests in the cell are distributed according to the equilibrium distribution. This is equivalent to approximating the jump randomization scheme with a different local operation where, just before the tagged guest is invoked, all the other guests in the cell undergo a memoryless randomization (see Sec. 2.3.2). Such an approximation will become

less accurate the more binding are the sites and the more restricted the dynamics is, since given these conditions the cell reaches local equilibrium more slowly.

(ii) We will treat mean-field randomization and propagation separately. In other words, the probability of jumping towards some direction will be factorized into probability of reaching some exit site during randomization and probability of performing a successful propagation, treated as independent one of the other.

The DACF at $t = 0$ is not affected by time-correlations and can be well approximated with

$$\begin{aligned}\langle \delta\mathbf{r}(0) \cdot \delta\mathbf{r}(0)\rangle =& \lambda^2 \frac{1}{\langle n\rangle} \sum_{\mathbf{n}} \sum_{\mathbf{m}} n_{\mathrm{ex}} \left(1 - \frac{m_{\mathrm{ex}}}{K_{\mathrm{ex}}}\right) \\ & \times p(\mathbf{n})\, p(\mathbf{m}) \kappa(\mathbf{n}, \mathbf{m}),\end{aligned} \tag{2.45}$$

where λ is the lattice spacing, and $\langle n\rangle$ is the loading (average number of occupied sites in a cell). The relations among D_0^{mf}, $\langle \delta\mathbf{r}(0) \cdot \delta\mathbf{r}(0)\rangle$ and $p(\Diamond)$ are given in Eqs. (2.27) and (2.28).

As we can see in Eqs. (2.35) and (2.36), the probabilities of interest refer to jumps starting from an exit site position. Thus, when evaluating the DACF terms for $z \geq 1$, one has to consider the conditional probability of the tagged guest *already located in an exit site* to stay in a cell with meso-configuration $\mathbf{n}$, rather than the absolute probability of $\mathbf{n}$ itself. Therefore we introduce $g_{\mathrm{ex}}(\mathbf{n})$, that can be re-interpreted as the conditional probability of a cell with an occupied exit site to be meso-configured like $\mathbf{n}$, i.e. to have $n_{\mathrm{ex}} - 1$ of the remaining $K_{\mathrm{ex}} - 1$ exit site and n_{in} of the K_{in} inner sites filled.

$$g_{\mathrm{ex}}(\mathbf{n}) = \frac{n_{\mathrm{ex}} p(\mathbf{n})}{\sum_{\mathbf{n}'} n'_{\mathrm{ex}} p(\mathbf{n}')}, \tag{2.46}$$

where the quantity

$$\frac{n_{\mathrm{ex}}}{K_{\mathrm{ex}}} p(\mathbf{n}) = [\Xi(\mu)]^{-1} \binom{K_{\mathrm{ex}} - 1}{n_{\mathrm{ex}} - 1} \binom{K_{\mathrm{in}}}{n_{\mathrm{in}}} e^{\beta\mu n} e^{-\beta F(\mathbf{n})} \tag{2.47}$$

is the total probability of one particular exit site, $n_{\mathrm{ex}} - 1$ of the remaining exit sites, and n_{in} inner sites to be occupied in a cell.

Mean-field jump randomization.

Once defined the probability distribution g_{ex} in Eq. (2.46), it is straightforward to derive mean-field expressions for the probability that, once the

tagged particle has targeted another exit site, it reaches it:

$$J_{\mathrm{ex-ex}} = \gamma e^{\beta f^{o}_{\mathrm{ex}}} \sum_{\mathbf{n}} \left(1 - \frac{n_{\mathrm{ex}} - 1}{K_{\mathrm{ex}} - 1}\right) g_{\mathrm{ex}}(\mathbf{n}), \tag{2.48}$$

This is the average acceptance of an exit-to-exit jump during randomization. Similarly, the average acceptance of an exit-to-inner jump is

$$\begin{aligned} J_{\mathrm{ex-in}} = &\gamma e^{\beta f^{o}_{\mathrm{ex}}} \sum_{\mathbf{n}} \left(1 - \frac{n_{\mathrm{in}}}{K_{\mathrm{in}}}\right) g_{\mathrm{ex}}(\mathbf{n}) e^{\beta \Phi(\mathbf{n})} \\ &\times e^{-\beta \max[\Phi(n_\alpha - 1, n_\nu + 1), \Phi(\mathbf{n})]} \end{aligned} \tag{2.49}$$

where γ has been defined when illustrating Eq. (4.4).

Mean-field propagation.

The mean-field propagation probability, that is the probability that during propagation a guest located in an exit site effectively migrates into the corresponding neighboring cell (this is sometimes referred to as *transmission coefficient*), can be formulated as

$$J_{\mathrm{prop}} = \sum_{\mathbf{n}} \sum_{\mathbf{m}} \left(1 - \frac{m_{\mathrm{ex}}}{K_{\mathrm{ex}}}\right) g_{\mathrm{ex}}(\mathbf{n}) p(\mathbf{m}) \kappa(\mathbf{n}, \mathbf{m}) \tag{2.50}$$

Mean-field jump probabilities.

We are now ready to write down mean-field expressions for the conditional probabilities included in Eqs. (2.35) and (2.36), for both the case of allowed and forbidden ex-ex jumps. These are listed in Table 2.6. Including them into Eqs. (2.35) and (2.36) gives:

$$\langle \delta\mathbf{r}(z\tau) \cdot \delta\mathbf{r}(0) \rangle = -2d\tau D_0^{\mathrm{mf}} J_{\mathrm{prop}} \left(1 - 2J_{\mathrm{prop}}\right)^{z-1} \left[1 - \frac{K_{\mathrm{ex}}}{K} J_{\mathrm{ex-ex}} - \frac{K_{\mathrm{in}}}{K} J_{\mathrm{ex-in}}\right]^{z} \tag{2.51}$$

$$D_s^{\mathrm{mf}} = D_0^{\mathrm{mf}} \left\{ 1 - 2J_{\mathrm{prop}} \frac{1 - \frac{K_{\mathrm{ex}}}{K} J_{\mathrm{ex-ex}} - \frac{K_{\mathrm{in}}}{K} J_{\mathrm{ex-in}}}{1 - (1 - 2J_{\mathrm{prop}}) \left[1 - \frac{K_{\mathrm{ex}}}{K} J_{\mathrm{ex-ex}} - \frac{K_{\mathrm{in}}}{K} J_{\mathrm{ex-in}}\right]} \right\}, \tag{2.52}$$

for allowed ex-ex jumps, and

$$\langle \delta\mathbf{r}(z\tau) \cdot \delta\mathbf{r}(0) \rangle = -2d\tau D_0^{\mathrm{mf}} J_{\mathrm{prop}} \left(1 - 2J_{\mathrm{prop}}\right)^{z-1} \left[1 - \frac{K_{\mathrm{in}}}{K} J_{\mathrm{ex-in}}\right]^{z} \tag{2.53}$$

$$D_s^{\mathrm{mf}} = D_0^{\mathrm{mf}} \left\{ 1 - 2J_{\mathrm{prop}} \frac{1 - \frac{K_{\mathrm{in}}}{K} J_{\mathrm{ex-in}}}{1 - (1 - 2J_{\mathrm{prop}}) \left[1 - \frac{K_{\mathrm{in}}}{K} J_{\mathrm{ex-in}} \right]} \right\}, \qquad (2.54)$$

for forbidden ex-ex jumps. The series

$$\sum_{z=1}^{\infty} A^z B^{z-1} = \frac{A}{1 - AB} \qquad (2.55)$$

has been used to perform the summation of the correlated part.

Unlike the more general mean-field formulas in Eqs. (2.35) and (2.36), the various quantities in Eqs. (2.51) to (2.54) do not depend on whether the ex-ex jumps are allowed or forbidden in the randomization algorithm. As can be seen, forbidding the ex-ex jumps has the only effect of dropping the term $-\frac{K_{\mathrm{ex}}}{K} J_{\mathrm{ex-ex}}$ out of the mean-field formulas.

Although the formulas above lead to a qualitatively correct representation of correlations, they do not always match quantitatively with the results of numerical simulations. Nevertheless, the obtained discrepancies are of great help in understanding the correlation mechanism, as we will discuss in Sec. 2.3.10.

2.3.10 Discussion of the mean-field results

In Figure 2.12 we plot the results of numerical simulations of the Central Cell Model applied in the cases where the deepest sites are the inner or exit ones respectively, each studied with two different levels of time-correlation entity. The values calculated explicitly from the numerical simulations through the Green-Kubo formula, Eq. (2.26), are reported as black squares, whereas general mean-field values and mean-field theoretical predictions are reported as blue circles and solid lines respectively.

As expected, the self-diffusion coefficient when the inner sites are the deepest ones increases from low to intermediate loadings as a consequence of the increasing probability of the tagged particle to occupy an exit site (thus being able to attempt a cage-to-cage jump), and starts decreasing at higher loadings when the exit sites tend to be saturated so that each pair of adjacent exit sites of communicating cells is more likely to be saturated, this leading the cells to exchange no particles. When the exit sites are the deepest ones instead, the pairs of exit sites tend to be saturated from the beginning (i.e. at low loadings), this leading to the expected decreasing diffusivity.

The escape probability, and thus D_0^{mf}, does not vary depending on whether or not the ex-ex jumps are allowed (see Figs. 2.12a and d). This

is because there are no correlations to be taken into account. As a consequence, both the general mean-field equation (2.36) and the more specific one obtained through the DACF value in Eq. (2.45) perfectly match with the explicit numerical value of D_0^{mf}.

The functions constituting the specific mean-field equations, Figs. 2.12b and f, give some insights about the migration probability of the individual processes. The way the average jump acceptances $J_{\mathrm{ex-ex}}$, $J_{\mathrm{ex-in}}$, and J_{prop} behave w.r.t. loading is the basis of the mean-field treatise of correlations. They are strictly connected to the choice of the difference between the site adsorption free-energies f_{ex}^o and f_{in}^o. As we described above, when the inner sites are the deepest ones the exit sites are poorly occupied. The accep-

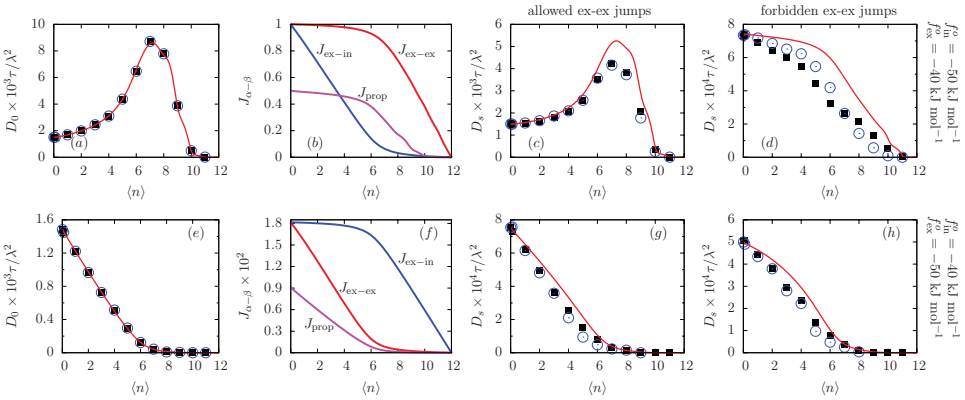


Figure 2.12: *Comparison between diffusivity obtained from numerical simulations of the Central Cell Model and from the mean-field theory. Black squares are obtained from the trajectory data outcoming from numerical simulations through the Green-Kubo formula, Eq. (2.26). Blue circles are obtained by applying on the same trajectory data the more general mean-field approximation, Eq. (2.36). Solid lines are theoretical prediction values obtained from the more specific mean-field approximations in Eqs. (2.52) and (2.54). In the first row, Figs. (a), (b), (c), and (d) the inner sites are 10 kJ mol^{-1} deeper than the exit sites, and vice-versa for the second row, Figs. (e), (f), (g), and (h). In the first column, Figs. (a) and (e), the zero-time diffusivity, Eq. (2.27) is shown. In the second column, Figs. (b) and (f), the trends of the macroscopic quantities $J_{\mathrm{ex-ex}}$, $J_{\mathrm{ex-in}}$, and J_{prop} constituting the theoretical mean-field approximation are shown. In the third and the fourth columns, respectively Figs. (c), (g) and (d), (h) the case of allowed and forbidden ex-ex jumps are considered.*

tance of ex-in jumps, $J_{\mathrm{ex-in}}$, starts to decrease from low loadings whereas the acceptance of (allowed) ex-ex jump, $J_{\mathrm{ex-ex}}$, is almost unity and does not decrease significantly as long as the inner sites are not close to saturation, around $\langle n \rangle \approx K_{\mathrm{in}}$. The behavior of J_{prop} is similar to $J_{\mathrm{ex-ex}}$: it remains

almost constant (about 1/2) until the loading becomes high enough so that the exit sites start being filled. Inverting the site depths $f^o_{\rm ex}$ and $f^o_{\rm in}$ exactly inverts the respective behaviors.

The average jump acceptances are combined together by Eqs. (2.52) and (2.54) to give approximated values for the correlated self diffusivity $D_s^{\rm mf}$ (see solid lines in Figs. 2.12c, d, g, and h). Although the theoretical prediction are qualitatively correct, they are close to the simulation values especially at low loadings, while usually they fail at higher loadings. The more general (numerical) diffusivity equation (2.36) gives a better appoximation than the theoretical prediction. This is because the numerical evaluation of $D_s^{\rm mf}$ through Eq. (2.36) does not suffer from the separation of mean-field randomization and propagation operations, which was the leading assumption when we derived the theoretical diffusivity formulas in Section 2.3.9. However, the general diffusivity equation becomes less accurate in situations where the memory of the previous local configurations is lost slowly, as for the case shown in Fig. 2.12d. When discussing about the amount of memory locally lost during each randomization step, it is interesting to find out the main sources of correlations, and to identify which of the cases above is the most memory-preserving.

Memory preserved in exit and inner sites. Since the cell-to-cell migrations occur via the exit sites, and their connectivity from one cell to the other determines the topology of the whole grid of cells, all events involving them will introduce more correlation than the events occurring in the inner sites, which instead are structureless so that they can be considered as the less memory-preserving part of the cell.

Memory-preserving backscattering. When a tagged guest migrates from cell to cell during propagation, the probabilities related to every next move *do* depend on the configuration of both cells before the propagation occurred. In other words, the assumption in Eq. (2.30) is strong and this is especially true when correlation effects are particularly evident, such as in the case of forbidden ex-ex jumps shown in Fig. 2.12d. In that case, (i) forbidding the ex-ex jumps gives the backscattering contribution a major role in the production of correlations (this because the randomization will produce only very small changes in the local configuration), and (ii) cage-to-cage jumps are infrequent because $f^o_{\rm in} < f^o_{\rm ex}$, so that the configuration of the exit sites tends not to change significantly from step to step. Due to these two facts, a backscattering particle which has left the cell $\mathbf{r}$ at time t and backscatters into it at time $t+\tau$ is very likely to find $\mathbf{r}$ just little changed or no changed at all. If the exit sites are the deepest instead, even though ex-ex jumps are forbidden one has that propagation events are more likely to occur at low-intermediate loadings than what expected when the inner sites were the

deepest. This causes the memory-preserving attitude of the exit sites to be less marked when the migration events are frequent. Therefore, as it can be seen from Fig. 2.12, the d case (deepest inner sites and forbidden ex-ex jumps) is the more affected by time-correlations in the self-diffusion process. The approximation in Eq. (2.30) becomes then less accurate, whereas in all the other cases it is acceptable.

2.3.11 Conclusions

In this work we laid down the basis of a simple computational framework, the Central Cell Model (CCM), aimed to be specific for the study of the motion on the mesoscopic scale of a single particle in a system of connected cavities in the presence of other diffusants, in conditions of thermodynamic equilibrium. Our model is local and discrete in both space and time, and in the numerical applications we have shown here it has been constructed starting from the algorithm of a lattice-gas model for diffusion in microporous material. We have shown that, although being not possible for the CCM to sample all the information obtainable by a full lattice-gas, a CCM simulation provides an accurate reproduction of the memory effects in the self-diffusion (and thus, of the diffusion isotherm) at a minimum computational cost.

The way the CCM is constructed suggested how to carry on a mean-field study of the self-diffusion process produced by the particular evolution rule adopted. This has led to two approximated mathematical expressions for self-diffusion. The first one, more general, can be applied with data coming straight from the CCM simulation. The second one, more case-specific and derived by assuming fast local equilibration, is theoretical and yelds a more accurate approximation the weaker the correlations and the lower the loadings are. Interpretation of the discrepancies between the self-diffusivity trends obtained from the numerical simulations and their two different mean-field approximations helped to understand how, and how strongly, memory effects can emerge depending on the very general features of the model parametrization.

The obtained results suggest the CCM approach to be suitable for other theoretical studies, e.g. the time-correlations in the local density, [50, 51] as well as for direct applications in the field of the molecular coarse-graining. For example, the CCM approach could be further extended to the sampling of both the adsorption and the self-diffusion isotherm through a single simulation when the lattice-gas rule includes an explicit cell-to-cell interaction potential which makes (in principle) impossible do derive the equilibrium probability distribution of states *a priori*. This could be done by performing a grand-canonical Monte Carlo on the border cells while keeping the core evolving with the prescribed dynamic lattice-gas rule in the canonical ensemble. Also, an even more intriguing extention of the CCM approach could be made in the field of hybrid MC-MD schemes aimed to realistically mimic

the bulk effects in the motion of a tagged guest in an atomistic simulation.

Chapter 3

Microscopic systems and Energy Landscapes

3.1 Potential and Free Energy Surfaces

A very interesting and profound point of view on atomic and molecular systems is that offered by looking at their associated potential energy surface (PES). The geography of these multidimensional *energy landscape* contains, in principle, all the relevant information relative to the system [52, 53]. In order to be able to say something on the thermodynamics and kinetics, one has to develop tools to find and envision the significant feature of this object, and to interpret them in a suitable theoretical framework, so as to build a bridge between the landscape and the experimental observables.

The PES is in general a very complex multidimensional object (for a system consisting of N atoms it depends in general on 3N coordinates), and the complete knowledge of its structure is impossible apart from very simple systems. A first step in the exploration can be made by reducing the number of degrees of freedom which we want to consider. A typical example is given by the construction of the Ramachandran plot, which consist of the projection of the energy landscape of a small peptide on a plane, by considering only two collective degrees of freedom, namely a pair $[\phi, \psi]$ of dihedral angles. Although the PES is a function of the three-dimensional coordinates of all the atoms except one (because in general we are only interested to the potential energy associated to a given configuration of the molecule, regardless of the position in space of its center of mass), it turns out that considering the possible states as defined by couples of values of ψ and ϕ, gives a satisfactory description of the molecule's behaviour at a greatly reduced cost. This is a typical way of coarse-graining used in the study of great molecules such as peptides.

Whenever we project out some degree of freedom from our description of a system, and start to ignore a part of it, entropy starts to play its role. So

free energy is needed if one wants to treat properly most aspects of equilibrium and dynamics. Nonetheless it is still possible to build a coarse-grained PES, by considering at each point of the reduced space of configurations the potential energy of the minimum energy configuration corresponding to the collective coordinates defining that point. If we want a more reliable description of the system we should use the free energy, taking into account the whole space of configurations compatible with the independent collective coordinates specified. We can thus build a free energy surface (FES), which is not uniquely defined for a system, but depends on the particular coarse-graining choice applied. In the reminder we refer to the energy landscape as a synonimous of PES unless otherwise specified.

Not all features of the energy landscape are equally relevant for predicting the behaviour of a system: if one is interested to the equilibrium thermodynamics it is in general enough to sample the minima of the PES, while for the study of kinetics at least first order saddle points must also be considered. The depth of the minima and approximate shape of their *attraction basins*, i.e. the subset of points for which a direct minimization leads to a given minimum, determine the equilibrium distribution. A system will in general spend most of its time in the lower regions of the PES. In the limit of 0K it will be confined to the global minimum (unless a too fast cooling traps it in some local minima), and as the temperature rises it is allowed to explore larger and larger portions of the PES, until, as $T \to \infty$ it can be found everywhere in the configuration space with equal probability. A first survey of the energy landscape should in general consist of a sampling of the relevant minima, possibly starting from the global one and going up depending on the temperature of interest. At temperatures different from 0K not only the depth of minima determine their associated population, but also the width of their attraction basins, on which the configurational entropy depends. Once a relevant portion of the minima is known the saddle points connecting them are needed for a definition of the kinetics. While higher order saddle points are rarely of interest, first order ones represents the *transition states*, determining the rates at which the system moves from one attraction basin to the other.

Unfortunately the number of minima grows more than exponentially as the number of atoms in the system, or the relevant degrees of freedom grows, and the growth of the number of saddle points is even faster. So a method for directly exploring the energy landscape, aiming to predict something of the system on the basis of this underlying structure, must be able to find its way in a maze of uninteresting local minima and saddles of various order.

We will see different methods for performing this search in the next sections and chapters.

However all of the simulation methods in computational chemistry can be seen as implicit PES sampling techniques.

3.2 Disconnectivity Graphs and archetypal Energy Landscapes

A useful way of schematically representing energy landscapes is offered by *disconnectivity graphs* (DS) [52,73]. Usually there is a huge number of direct minimization attraction basins (one for each minimum), but the direct minimization map of the EL to all of its minima is in general overwhelmingly rich and in most cases it is more useful to consider grouping schemes giving *super-basins*, i.e. groups of basins separated by low barriers, which gives a much more clear representation of the relevant features of the landscape, and are a first step in the direction of a coarse-grained treatment of the system. There are two principal ways of mapping the EL to a collection of super-basins, depending of the problem of interest: a microcanonical mapping, based on the absolute height of the barriers separating minima, and a canonical mapping, based on the relative height of the barriers. Given a certain threshold value (an absolute energy for the canonical map or a small multiple of kT for the microcanonical map), all basins separated by barriers lower than this threshold are lumped together. In canonical mapping two minima are considered as belonging to the same super-basin if the height of the TS connecting them, with respect to the deepest one, is lower than the chosen threshold. By iterating this lumping procedure several times, at different threshold values, a DS is built in the following way. We consider as

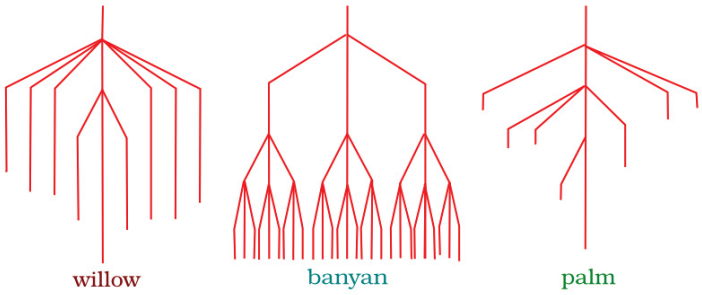


Figure 3.1: *Three examples of paradigmatic disconnectivity graphs.*

an example a microcanonical mapping: starting from an energy $E \longrightarrow \infty$ we have a single super-basin comprising the whole configuration space of the system; by moving down in energy, the space starts to break in various distinct, non communicating super-basins, until, when $E \longrightarrow 0$, we obtain the direct minimization map again. We can keep track of the way in which greater super-basins split to smaller ones as energy (or temperature) is lowered, by drawing a graph in which basins in the PES are represented as points arranged in layers, each corresponding to the mapping at a different energy

(or temperature), and connecting with arrows the parent super-basins to its children. What we obtain in this way is a disconnectivity graph, which is probably the best way we have to visually represent the multidimensional energy landscape and grasp an immediate idea of its structure, hierarchy and connectivity. By analyzing the appearance of these microcanonical DS it is easy to recognize some recurrent model in their global look, each corresponding to an archetypal energy landscape class. The first DS in fig. 3.1, resembling in some way a *willow-tree*, is representative of a typical rough energy landscape, with very high barriers connecting deep minima. This kind of EL is characteristic of glass forming systems, because every minimum is a kinetic trap, and finding the global minimum is very hard, so that the system is most of the time found in a metastable state different from the global minimum at low temperatures, unless a very slow cooling allows it to find the bottom. The second DS is similar in some way (we can call it a *banyan-tree*), as it has many low lying minima, all on a similar energy level, separated by high barriers, but in this case every branch keeps on branching at all levels, giving a hierarchy of basins and barriers. This is in general the worse scenario for finding the global minimum. The third DS, resembling a *palm-tree*, is typical of good 'structure-seeker' systems, such as proteins and regular crystals. It corresponds to a funnel-shaped energy landscape in which shallow minima lie on the slopes of the global minimum super-basin, so that, wherever the system starts from, it is naturally and rapidly led to the global minimum. The funneled EL of proteins is now widely recognized as the solution to the Levinthal paradox, which points to the fact that in a 'golf-course' model of the PES, where all minima, irrespective of their depth, have the same basin width and their barriers all lie at the same level, the time a protein, with the enormous number of minima on its PES, would take to find its native, biologically active state, would be many times the age of the universe. The fact that proteins fold on the scale of ms-μs can be explained by the 'focusing' nature of the underlying energy landscape.

However in most cases the energy landscape has a mixed overall character, it can show more than one well defined funnel, or many branches each with a given characteristic aspect.

3.3 Canonical Disconnectivity Graphs and kinetics

The microcanonical DS is useful to envision at a glance the shape and topography of potential energy surfaces, while the canonical DS is useful to describe the kinetics, both qualitatively and quantitatively. We can define a *transition* as a motion along a path that crosses a barrier higher than the thermal energy. Since in a canonical disconnectivity graph states belonging to the same basin are separated by barriers lower than kT, there are no

formal transitions between these states, which are in rapid equilibrium. The canonical DS allows the study of transitions between super-basins rather than states (or minima), and this level of coarse-graining, explored with various techniques, will be the focus of the reminder of this thesis.

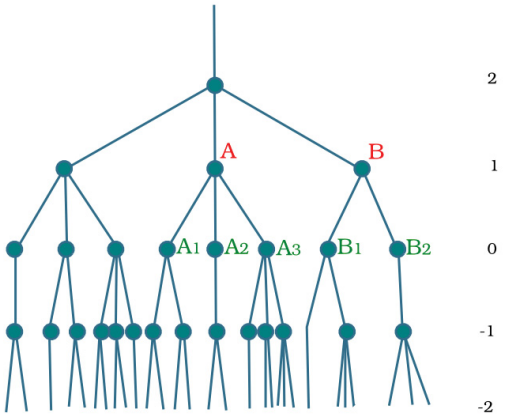


Figure 3.2: *At the temperature corresponding to level 0, all transitions between nodes under this level are fast, and details of the connectivity of the nodes should be irrelevant. On the contrary it is necessary to know the connectivity of upper levels. In general the higher is the least upper join between two nodes, the slower is the corresponding transition.*

Let us consider the canonical DS in fig. 3.2. We number the levels of the graph relative to the system temperature, which is assigned the index 0. All nodes on the 0-level of the graph represent actual super-basins at this temperature. These are denoted as B_1, B_2, A_1, etc., and in many cases are connected to several sub-basins, which in turns are in rapid equilibrium. The transition rate between the 0-level vertices of the graph is related to the location of their least upper join on it, which in turn is proportional to the height of the lowest barrier separating the super-basins. The lower is the barrier the faster is the transition. As an example, in fig 3.2 a transition from A_1 to A_2, having a join at level 1, will in general be faster than a transition from A_1 to B_1, having their least upper join at level 2. If the barrier between two super-basins is high enough we are not interested to the connectivity among the different basins they contain, and the system is markovian i.e. equilibrium is reached inside a given super-basin long before a jump to another super-basin can occur, thus making the system dynamics uncorrelated. But if this is not the case, the basin connectivity may start to play a not negligible role. In other words, if the kinetic hierarchy is not strongly separated, so that rates for transitions within super-basins are not

well lower than those for transitions between super-basins, we must take into account entropic effects arising from the inner structure of super-basins, and the system is not markovian anymore, as thermalization inside a super-basin, and memory loss, are not ensured before a new transition out of it can occur. This does affect dynamics, and the path and rate for reaching equilibrium, but the equilibrium itself is not affected. A complete understanding of a system showing not strongly separated dynamic hierarchy, requires a basins connectivity graph beside the DS.

Let us consider a system whose configuration space is divided in n states, which can be basins or super-basins for example. Its evolution can be described by a master equation, which is a gain-loss equation for the probability of finding the system in a state i. Its basic form is

$$dp_i/dt = \sum_J [W_{ij}p_j(t) - W_{ji}p_i(t)] \tag{3.1}$$

where W_{ij} is the transition probability from state j to state i. Eq. 3.1 can be rewritten in matrix form as

$$d\mathbf{p}(t)/dt = \mathbf{W}\mathbf{p}(t) \tag{3.2}$$

where $\mathbf{W}$ is a $n \times n$ transition matrix with elements

$$\mathbf{W}_{ij} = W_{ij} - \delta_{ij}(\sum_k W_{ki}), \tag{3.3}$$

and $\mathbf{p}$(t) is the n-dimensional probability vector at time t.
The formal solution of Eq. (3.2) is

$$\mathbf{p}(t) = e^{t\mathbf{W}}\mathbf{p}(0). \tag{3.4}$$

A convenient way of solving this equation is by expanding the probability vector $\mathbf{p}(t)$ in terms of the eigenvectors and eigenvalues of the transition matrix as

$$\mathbf{p}(t) = \mathbf{p}^{eq} + \sum_{\lambda_i<0} C_i\mathbf{s}_i e^{t\lambda_i} \tag{3.5}$$

where $C_i = \langle \mathbf{s}_i|\mathbf{p}(0)\rangle$ is the projection of the starting probability vector on the i-th eigenvector $\mathbf{s}_i$. The previous equation describes the relaxation to equilibrium in terms of the decay to zero of all components of the probability vector on eigenvectors different from $\mathbf{s}_0$, which corresponds to the only non-negative eigenvalue $\lambda_0 = 0$, which in fact represents the equilibrium distribution.

As $\mathbf{W}$ is not symmetric, it cannot be guaranteed that it has a complete set of eigenvectors spanning its space, and thus that the eigenexpansion in Eq. 3.5 is possible. However if the detailed balance condition $W_{ij}p_j^{eq} =$

$W_{ji}p_i^{eq}$ is satisfied, there is a transformation that transforms $\mathbf{W}$ into a real symmetric matrix $\mathbf{V}$

$$\mathbf{V} = [\mathbf{P}^{eq}]^{-1/2}\mathbf{W}[\mathbf{P}^{eq}]^{1/2} \tag{3.6}$$

where $\mathbf{P}^{eq}$ is a diagonal matrix with the equilibrium vector $\mathbf{p}^{eq}$ along its diagonal. Eigenvalues κ_i and eigenvectors $\mathbf{u}_i$ of $\mathbf{V}$ can be easily obtained with standard methods, and given the nature of the transformation the $\mathbf{V}$ and $\mathbf{W}$ matrices have the same eigenvalues, and their eigenvectors are related by

$$\mathbf{u}_i = [\mathbf{P}^{eq}]^{-1/2}\mathbf{s}_i. \tag{3.7}$$

Applying the previous transformation to the probability vector $\mathbf{p}(t)$ gives a transformed probability vector $\mathbf{q}(t) = [\mathbf{P}^{eq}]^{-1/2}\mathbf{p}(t)$, which obeys the master equation, Eq. 3.2 with the symmetric transition matrix $\mathbf{V}$:

$$d\mathbf{q}(t)/dt = \mathbf{V}\mathbf{q}(t). \tag{3.8}$$

Rewriting the solution in terms of the eigenexpansion as in 3.5 we obtain

$$\mathbf{q}(t) = \mathbf{q}^{eq} + \sum_{\kappa_i<0} C_i'\mathbf{u}_i e^{t\kappa_i} \tag{3.9}$$

with $C_i' = \langle \mathbf{u}_i|\mathbf{q}(0)\rangle$. By solving this equation, and transforming back to Eq. 3.5 we obtain the time evolution of the original probability vector $\mathbf{p}$(t) as

$$\mathbf{p}(t) = \mathbf{p}^{eq} + \sum_{\kappa_i<0} C_i'[\mathbf{P}^{eq}]^{1/2}\mathbf{u}_i e^{t\kappa_i} \tag{3.10}$$

where we can equivalently redefine C_i' as $C_i = \langle \mathbf{u}_i|[\mathbf{P}^{eq}]^{-1/2}\mathbf{p}(0)\rangle$. In the new form of Eq. 3.5 the coefficients are given by $C_i = \langle \mathbf{s}_i|[\mathbf{P}^{eq}]^{-1}\mathbf{p}(0)\rangle$, where $[\mathbf{P}^{eq}]^{-1}$ is the normalization factor of the vectors $\mathbf{s}_i$.

The PES disconnectivity graph contains only information about the super-basins bottom minima and transition states energies, so in order to calculate the transition probability W_{ij}, that is the rate constants k_{ij} for the transition from super-basin j to super-basin i we must adopt some transition state theory (TST) assumption. A general TST expression for the rate is

$$k_{ij} = W_{ij} = \frac{kT}{h}\frac{Q_{ij}^{\dagger}}{Q_j}e^{-E_{ij}/kT}. \tag{3.11}$$

Depending on the problem at hand and the accuracy required, a stochastic approach such as Kramer's theory or a more accurate rate theory formulation can be used [54]. It must be noted that in general at non-zero temperatures entropic effects become fundamental for a correct calculation

of rates, as well as equilibrium properties. For some examples of the estimation of the density of states see the extended harmonic superposition approach proposed by Strodel *et al.* [55], the Wang-Landau [56] and the basin-sampling [57] methods.

3.4 Exploring Energy Landscapes

As already mentioned, every simulation method performs a sampling of the energy landscape, and by analyzing simulation data it is in some case possible to map its most relevant regions, and the main paths connecting them. On the other hand the presence of high barriers may break the ergodicity of the system on the time scale practically reachable with the simulation method.

A way of studying high barriers systems, for which a coarse-grained method cannot be developed on the basis of MD data, is that of sampling the PES directly, looking for relevant critical points to be used for predicting the system's behaviour, or in order to build a suitable KMC model.

Direct study of the PES requires an efficient way of detecting critical points and their connections. Finding minima is an easy task, starting from any point, taken from a simulation trajectory or even randomly picked in the configuration space, and following a way down-hill i.e. minimizing energy via any of the many minimization methods available (for example the Newton-Raphson or the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method). Stillinger and Weber [58] proposed a way for detecting the so called *inherent-structure* of the PES, by quenching points in a long MD trajectory, so that mainly minima in the most relevant part of it are sampled. A more difficult task is that of locating the global minimum, as all other local minima are potential traps where the system can get stuck. Many different techniques have been proposed to avoid trapping, like simulated annealing and basin-hopping [59].

Once a minimum is located, a *single-ended search* technique [52] can be used to locate a saddle point leading out of the starting basin. Locating a first order saddle point, that is a TS, is more problematic than finding a minimum, as, while following any direction down-hill via the steepest descent path will most likely directly lead to a minimum, following up-hill steepest path will in general lead to a maximum. Eigenvector following is a way to reach a TS starting from a minimum by following the eigenvector corresponding to the lowest eigenvalue of the Hessian matrix evaluated at the minimum, that is the 'softest ascent' path up-hill [60]. Once the new stationary point is reached, the Hessian matrix is evaluated again, and if it has a single negative eigenvalue the point is a transition state. By stepping in the opposite verses on the direction of the eigenvector corresponding to the negative eigenvalue, and minimizing the energy, the two minima it connects

are located. At least one of these should be a new stationary point, from which the search for a TS can start again, and so on, resulting in an iterative procedure for locating new minima and the connecting transition states.

Another way of exploring the PES, once two or more minima are known, is by iterating *double-ended searches* on it via a Discrete Path Sampling (DPS) procedure [61, 62]. Instead of a starting point from which the algorithm generates a succession of points until an unknown final one, this approach requires two points (A and B), and aims to find a minimum energy path connecting them, in the form of a series of transition states and minima. The algorithm starts with a minimum energy path search via the doubly nudged elasting band method (DNEB) (see Trygubenko et al. [63] for more details on this variation of the well known NEB method [64]). The path found will in general show a maximum energy point, from which a single ended TS search is started. The two minima connected by this transition state (located in the same way as in the single-ended search), may be the two starting minima or not. In the first case a minimum-saddle-minimum connection has been found, and the algorithm stops. Otherwise the new points are added to the set of known minima, and new DNEB paths are calculated connecting A and B via the minima in the set. As the procedure goes on, the datasets of minima and saddles grows and a Dijkstra algorithm [65] is used to attempt connections, using as costs the heights of transition states connecting minima, and their distance in configuration space when the TS is missing. In this way a minimum energy discrete path is built connecting A and B, and at the same time a subset of the minima and transition sets of the PES is unveiled. Successive applications of DPS to pairs of points in the dataset and a perturbation procedure applied to known minimum energy paths result in a progressive sampling of the energy landscape minima and their connections.

3.5 Energy Landscapes for small molecules adsorbed in zeolites

We studied the energy landscapes of Ar, CH_4 and Xe adsporbed in the α-cage of zeolite ITQ-29 using the OPTIM and PATHSAMPLE free softwares, developed by the group of prof. David J. Wales at the University of Cambridge [66]. This suites offer many tools for the direct exploration of energy landscapes, employing the concepts and techniques described in the previous sections. For these calculations we adopted a drastically simplified model of the zeolite framework, consisting only of the 48 oxygen atoms of which the surface of a single α-cage consists. Periodic boundary conditions are applied, with a simulation box length of 12.3 Å and the atoms positions are taken from an energy minimized structure and held fixed. We remark that in all other calculations reported in this thesis work, only interactions between

	ε/k_B	σ
Ar	95.6	3.17
CH_4	115.0	3.47
Xe	136.2	3.45

Table 3.1: *Interaction parameters between adsorbed atoms and oxygen atoms belonging to the zeolite framework. Units are K^{-1} for ε, and $\AA$ for σ.*

the sorbate and the oxygen atoms are taken into account, as it is customary. The large electronic cloud of these atoms shields interactions with Si atoms, which are thus negligible, at least for small non polar molecules like those considered here. OPTIM offers a wide choice of force fields for specific systems and a general Lennard-Jones option for binary systems called LP, which we used. The force field parameters are reported in tab. 3.1.

More complex molecular systems can be studied by implementing in OPTIM the AMBER (or CHARMM) package. Unfortunately we had not the possibility of employing these option, and that is the reason why we only report here results for two atomic species and a spherically approximated model of methane. It would be very interesting to apply the method to structured molecules and to other zeolitic frameworks containing ions, as these systems may present high barriers, slowing down the dynamics to the point that in some cases standard MD simulations are not able to sample the relevant transitions for diffusion. It is in a situation like this that the energy landscape approach becomes really useful. However for the systems presented here it gives an idea of the the preferential sites of adsorption, and can add to the explanation of what is observed in MD simulations. The microcanonical disconnectivity graphs for Ar, Xe and methane are shown in fig. 3.3, with a spacing between levels of 115 J/mol. For all three systems, there are six deepest minima (shown in yellow), simmetrically located in front of the eight-membered ring windows connecting the various cages. Eight other deep minima (shown in blue) are located in front of the six-membered ring windows, in the inner corners of the cage, and a number of other minima (shown in violet), which are not exhaustively mapped, are also present. These are 24 in total and are located in pairs in front of the four-membered windows for Ar, and in groups of three, around the each of the eight previous intermediate minima, for CH_4 and Xe. In the following chapters, regions surrounding minima of the first kind will be called I-sites, and regions surrounding minima of the second kind (containing also related minima of the third kind) will be called II-sites.

As can be seen, the adsorption energies for argon is far less than that of methane and xenon, and barriers are less pronounced. The highest barriers for xenon and methane are on the plane of the inter-cage windows, and are of the order of roughly 1 kJ/mol, while the very low inter-cage barrier for argon

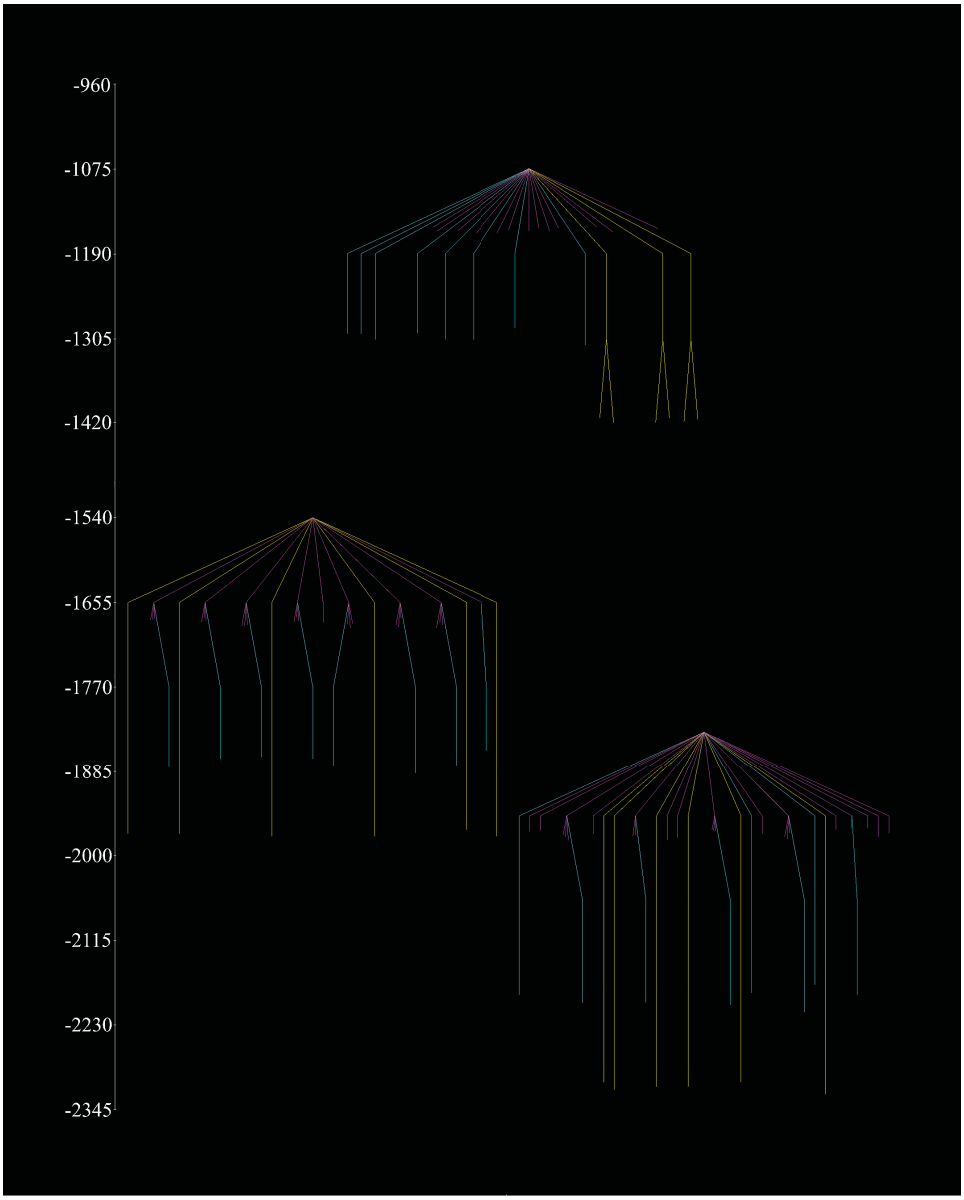


Figure 3.3: *Disconnectivity graphs for Ar (top), CH_4 (bottom-left) and Xe (bottom-right), in the α-cage of zeolite ITQ-29. The energy is given in units of J/mol and refers only to the zeolite-adsorbate interactions, as interactions among the zeolite oxygen atoms are set to zero. The level spacing for the grouping procedure is of 115 J/mol.*

is an order of magnitude smaller. This is the reason why the deepest minima for this system appear to be grouped in pairs: given the periodic boundary conditions each minimum in front of an eight-membered window seems to be connected by a low barrier to the corresponding minimum on the opposite side of the cage. This seems to suggest that inter-cage motion is faster than intra-cage motion. Argon shows indeed an higher frequency of inter-cage migration (see fig. 4.12) compared to Xe and CH_4, however entropic effects should be considered, and these account for the fact that inter-cage jumps are rare if compared to intra-cage jumps. The sketch of the energy landscape given here gives a qualitative idea of the sorption sites and of the general dynamics, however it is rigorously valid only at 0K. At any finite temperature entropic effects become important and the shape of attraction basins should be taken into account, as it can change drastically the thermodynamics and kinetic; the free energy landscape must thus be considered. Moreover, for all three systems considered here intra-cage barriers are well below kT at room temperature, and nonetheless they are not completely negligible for Xe and CH_4, so one should expect a behaviour which cannot be easily described in terms of jumps between attraction basins of the cage minima. On the other hand one could consider the whole cage as a single super-basin, and concentrate on the inter-cage jumps. This turns out to be a good approximation when few molecules are present, but for high sorbate density, equilibrium inside the α-cage after a jump in cannot be assumed before a new jump occurs. To understand how the energy landscape evolves as the number of molecules increases we tried to map it as before, but with more than one molecule present.In these calculations a permutation option was employed, allowing the permutation of coordinates of selected atoms. In this way the huge number of degenerate configurations which differs only for the identity of the atoms can be merged. We allowed the permutation of all atoms apart from one, aiming to obtain a map of the energy landscape experienced by one molecule in a zeolite cage with a varying number of other guests. The energy landscape of Lennard-Jones binary clusters with up to more than one hundred atoms have been studied and are reported in literature [67,68], however we were not able to obtain a satisfactory mapping of the PES for two or more molecules. The number of critical points increases very fast with the number of molecules, with the appearance of a huge amount of minima which are very close in energy and configuration geometry, and are connected by low barriers. This renders very challenging a satisfactory sampling of the relevant parts of the PES, as the algorithm gets lost in every small region of configuration space, and every search for a transition state between two minima results in the discovery of new uninteresting minima, so that the dataset of minima increases too fast if compared to the transition states dataset, remaining largely disconnected. The very intricate fine structure of the PES may be explained by the fixed zeolite framework inducing a first general shape on it, which is finely an densely perturbed by

the presence of other molecules. PATHSAMPLE offers a number of way of regrouping minima, in order to highlight significant feature of the EL, and to render the disconnectivity graph easier to analyze. However before such a regrouping can be performed, a sensible sampling of the PES should be performed, and this turned out to be impossible in our calculations, because of the aforementioned proliferation of minima and transition states. It is possible that a better expertise in running the softwares employed would allow an efficient EL exploration, overriding uninteresting shallow minima. This could be the matter of further investigations.

Chapter 4

First and second order markovian models of dynamics

4.1 PCCA and Markov State Models

Any trajectory obtained from a simulation method, such as Molecular Dynamics on which we will focus, constitutes a sample of the underlying energy landscape, so that the basin structure and kinetic hierarchies can be reconstructed from its analysis and used to build a coarse-grained model of the system. In general the aim of such an effort will be that of finding metastable states, corresponding to super-basins enclosed by high barriers, and the transition rates among these states.

One way of doing this is that proposed by Noè *et al.* [69] for building Markov State Models (MSM) [70,71]. The method starts with a fine-grained partition of the configuration space of the system into *microstates*, which is used to discretize the continuous MD trajectories via a *clustering* procedure. There are different ways of partitioning the space, which are in general Voronoi partitions based on different choices of the microstates centers number and distribution. One can chose a regularly spaced distribution by fixing a center-center distance, adding a new center whenever a point in the trajectory is further apart than this distance from any other previously determined center; otherwise new centers can be created every n time steps of the trajectory, with the coordinates of the system at that time. Another system-based approach, which tends to concentrate microstates centers in the more statistically relevant portions of the configuration space is the k-medoid method. The continuous trajectory is then mapped to the microstates, by assigning each point to the microstate whose center is nearest, and, given a certain *lag time* τ, the probabilities for transitions among microstates, occurring within τ, are calculated from the discrete trajectory. Thus the dynamics can be

reformulated using a discrete time transition matrix $\mathbf{T}(\tau)$, whose entries T_{ij} represent the probability of finding the system in state j at time $t+\tau$ given that it was in state i at time t.

$$\mathbf{p}((k+1)\tau) = \mathbf{p}(k\tau)T(\tau) \tag{4.1}$$

Equations 3.2 and 4.1 give equivalent results in the case of discrete time and are related by

$$\mathbf{T}(\tau) = e^{\tau\mathbf{K}} \tag{4.2}$$

where $\mathbf{K}$ is the transpose of $\mathbf{W}$. The analysis of the eigenspectrum and eigenvectors of matrix $\mathbf{T}(\tau)$ in the light of Perron-Frobenius (P-F) theorem and its corollaries, is at the basis of the Perron-Cluster Cluster Analysis method (PCCA) for detection of metastable aggregates of microstates [72].

Let us consider a matrix $\mathbf{T}(\tau)$, which is a row stochastic matrix i.e. such that for each of its rows $\sum_j T_{ij} = 1$. If the matrix is such that there exists a single m for for which $[T^m]_{ij} > 0$, it is a *primitive* matrix, and the Markov chain it produces is a *regular* Markov chain. All irreducible, non-negative, aperiodic matrices are primitive and obey the Perron-Frobenius theorem like positive matrices do. The theorem states that there is a single dominant real eigenvalue in the eigenspectrum of these matrices, called Perron root or Perron-Frobenius eigenvalue, λ_0, and all other eigenvalues, which can in general be complex, satisfy $|\lambda_i| < |\lambda_0|$. For stochastic matrices $\lambda_0 = 1$ and the corresponding left eigenvector is the equilibrium probability distribution $\mathbf{u}_0 = \mathbf{p}^{eq}$, while the right eigenvector is of the form $\mathbf{v}_0 = (1, ..., 1)^T$. Another consequence of the theorem is that, while $\mathbf{v}_0$ and $\mathbf{u}_0$ are are positive, all other eigenvectors have at least one negative element. Each pair of these left and right eigenvectors corresponds to a fundamental mode of the dynamics of the system, and their study reveals the basins and super-basins hierarchical organization of the underlying energy landscape. An example may be useful for understanding how this works. Let us consider the two transition matrices schematically represented in fig. 4.1, in which the value of entries is represented by the colour gradation. The first matrix is an uncoupled, reducible one, in which *invariant aggregates* (i.e. aggregates of states that do not communicate with the rest of space) are clearly visible as blocks on the diagonal. There are no transitions among this blocks, so the matrix can be decomposed into four distinct matrices each having its own eigenspectrum, eigenvectors and stationary distribution. So the eigenspectrum of matrix $\mathbf{A}$, which does not satisfy the conditions of the P-F theorem, will show four different $\lambda_i = 1$ eigenvalues, and the corresponding eigenvectors are the stationary distributions for each of the invariant aggregates. In the end this is because we are considering four distinct and entirely independent systems as a single one, and the stationary distribution will depend on the starting conditions.

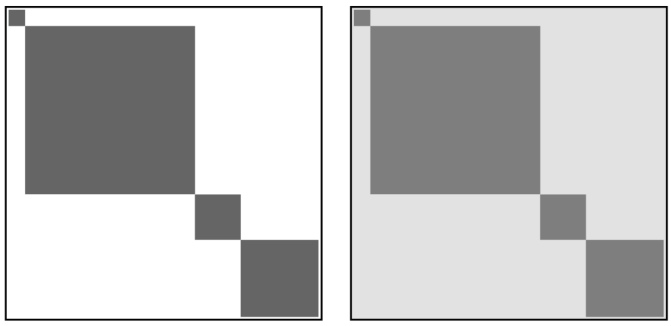


Figure 4.1: *Two hypothetical matrices are sketched. The color intensity roughly represent the average value of entries in the various blocks and out of them*

Let us consider now the second stochastic matrix **B**, which is in turn irreducible, thus ergodic, but shows a block structure as well. Intuitively one would expect in this case a slow transition dynamics among the blocks, and a fast dynamics inside each of them. The blocks are *almost invariant aggregates* [72], that is metastable aggregates of microstates, which we will call *macrostates*. One can think of these as super-basins, and of microstates as basins in the energy landscape picture given in section 3.2. However it is important to notice that, while properly defined macrostates should correspond to canonical super-basins at the MD trajectory temperature, microstates defined in the clustering procedure do not in general correspond to basins as they were defined in section 3.2.

Matrix **A** can be seen as a limit case of matrix **B**, in which the weak communications among the blocks has ceased completely, leading to four separated systems. So the eigenspectrum of **B** resembles that of **A**, but this time there is a single dominant eigenvalue, left and right eigenvector, because the P-F theorem conditions are satisfied, and the system has a single stationary distribution.

The second, third and fourth eigenvalues are now smaller but they are still in the vicinity of $\lambda_0 = 1$, and the four constitute what we call the *Perron-cluster*, which is separated from the rest of the eigenspectrum by a clear gap, proportional to the metastability of the aggregates. Each eigenvector corresponds to a fundamental mode of dynamics i.e. its structure is related to how probability flows between different regions of the configuration space, according to their different sign in the eigenvector entries. Namely, given a certain initial probability distribution $\mathbf{p}(0)$, and its component $C_i = \langle \mathbf{u}_i | \mathbf{p}(0) \rangle$ on the i-th eigenvector, microstates whose corresponding $\mathbf{u}_i$ entries have negative sign, gain probability at the expenses of those with positive sign if C_i is positive, and vice versa. This happens with a

characteristic implied time scale

$$t_i = -\tau / ln\lambda_i \tag{4.3}$$

As it is clear from Eq. 4.3 large eigenvalues correspond to slow relaxation processes towards equilibrium i.e. those related to the flow of probability among blocks, and the anatomy of this processes is embedded in the sign structure of their eigenvectors, which reflects how the configuration space splits into metastable aggregates. PCCA exploits this fact for detecting kinetically separated super-basins as lumps of microstates.

Once these have been determined they can be tested for use as states on which a coarse-grained Markov model can be built. Even if a gap in the eigenspectrum exists and the system exhibits a certain splitting in the characteristic time scales of its dynamics, markovity for the given lag time τ is not granted. It could be necessary to increment the lag time for the MSM to ensure that short time memory effects are avoided. A first hint to the threshold lag time beyond which markovity should be ensured, is given by the slope of a plot of t_i vs τ. This should reach a plateau, and the proper value of τ can be chosen in the region where t_i is constant, depending on the time scale of the process that one is interested to study. Even in the plateau region however markovity is not ensured, and the only way to ascertain it is to compare the MSM results with those obtained from the original simulation data (employing the Chapman-Kolmogorov equation or running a MSM simulation for example).

Using the concepts and techniques described in this section, we tried to develope a MSM for studying the diffusion of adsorbates in microporous materials. We have found that a simple MSM is not suitable for this purpose if one is interested in the intra-cage jumps dynamics alongside with inter-cage motion, because barriers inside cages are not high enough to induce a clearly separated dynamic hierarchy. Even if a MSM model based only on inter-cage jumps could be useful for studying long scale dynamics at low sorbate density, strong backscattering correlations makes it not suitable in case of high sorbate density. Moreover we wanted our coarse-graining to act as a filter for highlighting the relevant events determining diffusion trends, and for this purpose intra-cage motion is obviously fundamental.

4.2 A coarse-grained method based on the analysis of short MD trajectories for the simulation of non-markovian dynamics of sorbates in microporous materials

We developed a coarse-grained model suitable for the study of adsorbed molecules in microporous materials. A partition of the inner pores space was carried out, which allows to formulate the dynamics in terms of jumps between discrete regions. The probabilities of observing given pairs of successive jumps were calculated from Molecular Dinamics simulations, performed on small systems, and used to drive the motion of molecules in a lattice-gas model. Dynamics is thus reformulated in terms of event-space dynamics and this allows to treat the system despite its inherent non markovity. Despite some strong assumptions enforced in the algorithm, results show that it can be applied to various spherical molecules adsorbed in the all-silica zeolite ITQ-29, estabilishing a suitable direct bridge between MD simulation results and coarse-grained models.

The first step in our method is the clustering of MD trajectories frames to obtain a discretized version of the trajectory of adsorbed molecules within the zeolite. Each pore in the aluminosilicate is partitioned in a number of regions (or sites) and each point in the original trajectory is mapped to the proper region according to a distance criterion. The regions correspond roughly to the main basins in the free energy surface (FES) [73, 74]. This discrete trajectories in space are then mapped to 'event-space' trajectories: given a certain lag time we replace the two discrete positions at the start and end of each interval with the corresponding discrete displacements. We then perform a statistical analysis of the trajectories and obtain the transition probabilities from each discrete event to each other as a transition matrix. This matrix is at the basis of the evolution rule of the model. Considering the transitions between events allows one to circumvent the inherent non-markovity of the space partition, at least to a reasonable degree of approximation. Simulating the motion of guest molecules as a simple random walk on the lattice of regions obtained in the first step of coarse-graining would be inaccurate due to the low barriers in the FES, specially when few molecules are present. By considering the transitions between events, short time memory effects are naturally taken into account and this allows one to treat the dynamics of interest as a simple random walk in the event-space, driven by an algorithm based on the event-event transition matrix.The trajectory on this space then naturally implies a trajectory in the discretized physical space. As shown below our method gives self-diffusivities, events frequencies and occupation probabilities of regions and pores in reasonable agreement with the MD results for spherical, or roughly spherical, guest molecules, such as Ar, Kr, Xe and CH_4 in zeolite ITQ-29 [75].

4.2.1 Space partition and the transition matrix

Each α-cage [76] (see fig. 4.2) of the aluminosilicate is represented by 15 sites in our lattice-gas model, that can host one or more molecule each. These correspond to different regions of the inner space of the cage: six sites of kind I, corresponding to regions in front of the 8-membered ring windows connecting the cages, eight sites of kind II, corresponding to regions in front of the sodalite 6-membered ring windows, and one site of kind III, corresponding to the center of the cage. There are many possible ways to carry out the partition, the simplest being a Voronoi tessellation based on some choice of the regions centers. In order to have a more 'natural', system-based partitioning scheme we applied a version of the Perron-Cluster Cluster Analysis (PCCA+) [77] first proposed by Deuflhard *et al.* [72, 78], that gives more finely defined regions boundaries, better reflecting the underlying energy landscape. It should be noted that even this approach is to some degree arbitrary, given the nature of the systems studied, as explained below. The Perron-Cluster Cluster Analysis is based on some corollaries of the Perron-Frobenius theorem. Given a certain fine-grained partition of space into *microstates*, MD trajectories are discretized by mapping each point to the microstate containing it, and the transition probabilities between pair of microstates are calculated. If the transition matrix, which is stochastic, is irreducible and non-negative, it has a single eigenvalue $\lambda_{pf} = 1$, which is the Perron-Frobenius eigenvalue. If there are n metastable states in the dynamics, the eigenspectrum of the spectrum will show a 'cluster' of $n - 1$ other eigenvalues close to λ_{pf}, and, depending on the difference in time scales, a gap between this cluster and the rest of the eigenvalues. By analysing the sign structure of the eigenvectors corresponding to these eigenvalues it is possible to lump microstates into metastable *macrostates*. A more detailed explanation of the principles underlying PCCA+ can be found in Kube *et al.* [77]. The procedure does not ensure that the dynamics can be formulated in terms of a Markov chain on these macrostates, and the Markov State Model (MSM) [69–71] so obtained should be tested by comparing its results to those of the original MD trajectories. Indeed we have found that for the systems we studied (i.e. small non-polar molecules in an all-silica LTA zeolite at 300 K), a simple spatial MSM cannot be obtained, at least at the level of jumps within α-cages, regardless of the lag time used. This is due to the weak interactions between guest molecules and the cation-free zeolite framework, which imply low barriers to diffusion inside the α-cage, so that the time scales of jumps within macrostates and jumps in and out of them are not clearly separated. It must be noted that at low sorbate density a MSM can be built by considering the whole α-cage as a single macrostate, thus taking into account only jumps between cages. Even in this case markovity is lost at high sorbate density, when backscattering effects arise.

In the present study we are interested to the dynamics at the level of

intra- as well as inter-cage jumps [28] and thus will turn to a second-order Markov chain approach, as explained below.

We started from a MD trajectory of CH_4 in ITQ-29 at loading (average number of molecules per α-cage) $\langle n \rangle = 12$. As all MD trajectories in this study this was generated using the LAMMPS package [79, 80]. The zeolite framework is flexible. The force field parameters used for it were recently proposed by our group [7], while the parameters for guest CH_4 molecules were taken from Krishna *et al.* [81]. Lorentz-Berthelot mixing rules were used for cross interactions. After 10^6 steps of thermalization at 300 K, the simulation was run for $5 \cdot 10^7$ steps of 1 fs in the *NVE* ensemble, temperature being held at 300 10 K by the zeolite flexible framework acting as an effective thermostat. Snapshots were collected each 100 fs. The simulated system corresponds to a single zeolite crystallographic cell with 96 CH_4 molecules and periodic conditions (PBC) applied at the boundaries, with a simulation box length of 2.555 nm.

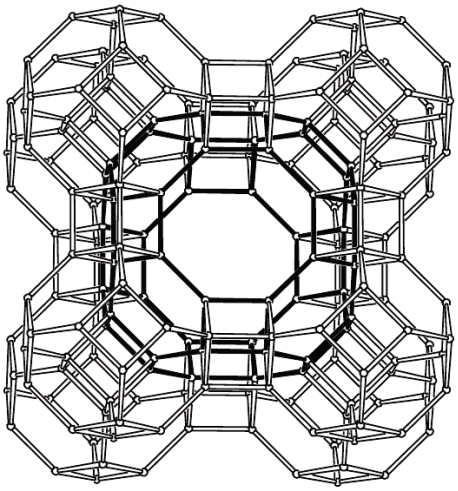


Figure 4.2: *The structure of zeolite ITQ-29 with the $\alpha-$cage highlighted.*

A relatively fine-grained clustering procedure was then applied to a collection of single-molecule trajectories, obtained from the original one, in order to generate discretized trajectories [82]. In this step the inner space of a single α-cage was partitioned in 480 microstates with a k-medoid algorithm [82]. The microstates-transition matrix was calculated from these discrete trajectories and it was analyzed by means of PCCA+ [77], to give 15 metastable macrostates as groups of the microstates defined in the previous step. All these calculations have been carried out using the free software EMMA [82].

In this way we obtained a space-partition where each macrostate is a

K	Start	Target
1	I	I
2	II	II
3	III	III
4	II	III
5	III	II
6	II	II*
7	I	II
8	II	I
9	I	III
10	III	I
11	II	II†
12	I	I‡

Table 4.1: *Event classes observed within a lag time of 100 fs in the MD trajectories. Event classes are defined by the site kind of the start-site and target-site.*
* *the site is adjacent to the start site*
† *the site is not adjacent to the start site*
‡ *the site is in the adjacent cage*

rough estimate of the energy landscape basins for single molecules, given a certain loading. Coadsorbed molecules are not taken directly into account in the analysis but obviously affect the single molecule behaviour. We have found that at higher loadings macrostates are more clearly defined and this is the reason why we used simulation data at $\langle n \rangle = 12$ for the definition of the sites.

We already discussed the difficulties of a simple MSM approach to the simulation of intra-cage dynamics. Our solution to the this problem was that of considering events instead of positions: the same space partition was used to obtain discretized trajectories for all the other systems studied, and these were then mapped to event-space trajectories. This was done by substituting each discrete position with the discrete displacement relative to the previous step. From these trajectories a transition matrix relative to each loading was obtained. These can be interpreted as second-order Markov matrices relative to the motion in the discretized space. From now on events will be called equivalently 'states', and being in state i, corresponding to a jump from site a to site b, means that the particle is actually in site b and was in site a during the previous step.

Given a lag time of 100 fs a finite number of events is observed and these can be classified in 12 different classes, as summarized in tab. 4.1 and shown schematically in fig. 4.3. The first three correspond to null-events, i.e. the particle stays in the same site. All the other classes correspond to

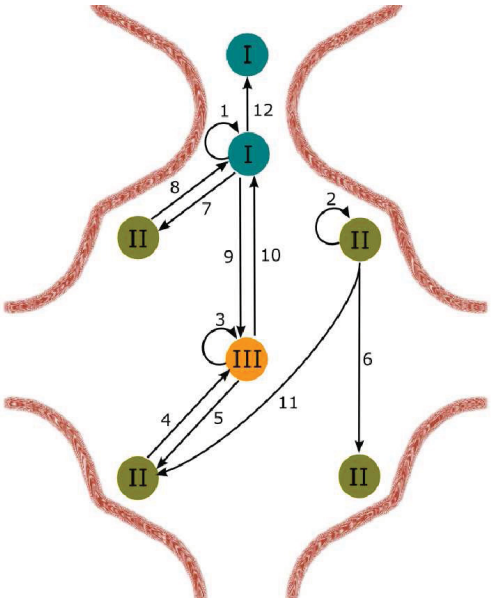


Figure 4.3: *A schematic representation of the* 12 *classes of events. A single* $\alpha-$*cage is shown plus one I-site belonging to an adjacent cage (top of the figure). Sites not strictly necessary for the definition of the events were omitted.*

jumps within a single cage (intra-cage jumps), apart from class 12 which groups jumps taking place between two adjacent 8-membered ring window sites belonging to different cages (inter-cage jump) [28]. No other inter-cage class of event was observed within our lag time.

For each class there are many different states, each corresponding to the particular pair of start- and target-site involved in the jump. Moreover we defined some additional null-event states in order to keep as much memory effects in the dynamics as possible. Let us consider a segment of discrete space trajectory $a \to b \to c \to d \to d \to ...$, corresponding to a sequence $ab \to bc \to cd \to dd \to d...$ in the event-space trajectory: the first states in this last sequence do not imply thermalization in any site, they keep memory of the previous site visited and allow one to calculate transition probabilities satisfactorily, reflecting the correlation among jumps; on the other hand the last step reset in some way the particle's story, as after the

null-event any of the possible events starting in d would be equivalent, which is true only if the particle effectively thermalizes in site d. Supposing that the cell occupancy is high, we know that, in this particular example, a jump back to c is instead more likely than any other, and this backscattering correlation memory effect turns out to last more than our lag time of 100 fs. In order to avoid this sudden 'amnesia', for each null-event we define a different state on the basis of the last site visited before the null-event, so that in our example dd becomes dd_c. This enhances the number of states significantly but drastically improves the reliability of the method. On the other hand using a longer lag time, in order to lower the correlation effects, would introduce even more events, and eventually lead to the occurrence of jumps between non-adjacent cages, thus destroying the locality of the model.

4.2.2 Evolution rule.

In the algorithm described below, particles move inside and between cells without direct exclusion interaction, and they only affect each other by means of the cell occupancy, acting as a sort of mean field. One of the interesting feature of the method is that, despite this, correlation effects are taken into account, at least to some degree. This is because of the way the event-event transition matrices are built, which allows them to retain a certain amount of the particle's story memory. We assume that in a cage containing n particles transitions involving only intra-cage jumps occur according to the transition matrix calculated from MD trajectories at loading $\langle n \rangle$. A critical point is that of inter-cell jumps, which poses some non-trivial challenges. In fact if we accept that transitions involving jumps between two cells with the same occupancy n occur with the probabilities encoded in the n-th transition matrix (i.e. calculated at loading $\langle n \rangle$), what should we do when jumps between two cells with occupancies n and m respectively, with $n \neq m$, are involved? One could try to keep on with the simple random walk in event-space as for intra-cage jumps, and let, for instance, the occupancy of the starting cell to decide which is the relevant transition matrix. But this gives very broad gaussian distributions of the occupancy probabilities, showing no resemblance with the MD data.

Our solution is based on a simple assumption, namely that we can express the transition probability $P_{ij}(n, m)$ for the transition from state i to a state j, where j corresponds to an inter-cage jump from a cell with occupancy n to a cell with occupancy m, factorizing it as

$$P_{ij}(n, m) = P^*_{ij}(n) P^{\text{in}}(m) \tag{4.4}$$

where $P^*_{ij}(n)$ may be interpreted as the probability of an attempt to perform a transition $i \rightarrow j$, which will be successful with probability $P^{\text{in}}(m)$,

which is independent of j and is simply the probability that a cell with occupancy m accept an incoming particle.

This is of course a strong assumptions as P_{ij} could be not factorizable, and in any case the choice of P^*_{ij} and P^{in} is non-trivial and to some degree arbitrary as we'll see. The previous relation is interesting as it indicates a way to separate the respective contributions of the starting and target cells, so that one can think of a procedure for inter-cage jumps consisting of a first trial step depending entirely on the starting cell occupancy and a second step depending only on the target cell occupancy, in which the trial jump is accepted or rejected. From Eq. (4.4) follows

$$P^*_{ij}(n) = P_{ij}(n, m)/P^{\text{in}}(m) \tag{4.5}$$

which allows one to calculate $P^*_{ij}(n)$ given $P^{\text{in}}(m)$. Our choice was that of considering a $P^{\text{in}}(m)$ as a linearly decreasing function of m

$$P^{\text{in}}(m) = (n_{max} - m)/n_{max} \tag{4.6}$$

where n_{max} is a tunable parameter and should corresponds roughly to the maximum number of molecules that a cell can accomodate. This makes sense intuitively as one expects that the higher is the number of particles in a given cell the higher is the probability that an incoming particle is rejected: in fact the target site, in a high occupancy cell, is likely to be occupied itself and even if the incoming particle has enough energy to displace the other one, this will have in turn to find room inside the cell, which is harder and harder as occupancy grows. The functional form in Eq. (4.6) is just one among many possible ones, and it was chosen as it is simple and gave better results than all other candidates, nonetheless a further effort to find a better function could surely improve the method.

Given the previous discussion, the n-th transition matrix needs to be changed slightly, by substituting entries $P_{ik}(n)$, where k is an inter-cell jump, with the corresponding $P^*_{ik}(n)$, obtained via Eqs. (4.5) and (4.6) by considering the occupation of the target cell to be n. In order to keep outgoing probabilities normalization, for each state i connected with an inter-cell jump state, we keep all P_{ij} - where j is an intra-cage non-null jump state - unchanged, and set

$$P_{il} = 1 - \sum_{j \neq l} P_{ij} \tag{4.7}$$

where l is the null-event on the target site of i.

The right probability of observing l is preserved by the algorithm as explained below.

At the begin of each run, particles are distributed randomly on the lattice sites, regardless of each other's position. Each particle has then a starting random state assigned, consistent with the occupancy of its cell.

At each timestep all particles are considered in random order for update of their state and position. The proper transition matrix is considered for each particle, depending on the actual occupancy of the host cell. A random real number r in $0 < r < 1$ is compared to the progressive partial sums $s(i)$ of transition probabilities from the actual state j to all other reachable states, until $r < s(i)$, with

$$s(i) = \sum_{k=1}^{i} P_{jk} \tag{4.8}$$

then, if the target-position of event i is within the same cell, the new state is set to i and the new position is set to the target position of i, regardless of its occupancy. Otherwise, if i is an inter-cage event, another random number is extracted and compared with $P^{\text{in}}(m)$, where m is the actual occupancy of the target cell. If this step is successful the particle jumps to the new cell and its new state is set to that corresponding to i at the new cell's occupancy. Otherwise the transition is rejected and the particle state is set to the null-event of the starting site - i.e. the target site of j. The probability of directly choosing a null-event was previously lowered in order to enforce normalization (Eq. (4.7)), but with the present rule the overall probability is actually restored (see fig. 4.4). As the occupancy of cells evolve the procedure causes the transition matrix relevant to a given particle, to change very often, and a conversion rule matching corresponding states in different matrices is needed. For example consider a cell with occupancy n at time t, and a particle in state i, corresponding to the jump from site a to site b at loading $\langle n \rangle$. If at time $t+1$ the cell occupancy has become m then, before attempting any transition, the particle state must be changed to that corresponding to the jump from a to b at loading m. In doing this one must take into account the possibility that a particle find itself in a state with no analogous in the transition matrix corresponding to the new occupancy (as an example events of class 11 tend to dwindle in number at high loadings). As null-events are by far the most likely to occur, in this case we choose a 'default' state for each possible event corresponding to the most probable null-event for its target site, so that this is surely present in all transition matrices.

The described model can be thought of as stochastic algorithm leading to a random walk in the event-space, with a fixed time step (i.e. the lag time used to define the transitions). The usefulness of this approach is in its ability to deal with systems showing weak metastability, so that their dynamics is slow but non truly markovian. All particles in the lattice are considered, and eventually allowed to move, in a random order at each time step, but at least in its present version the algorithm is not parallelizable to give a Cellular Automaton (CA) model, as there is not a suitable way of partitioning the lattice so that each block can be updated independently

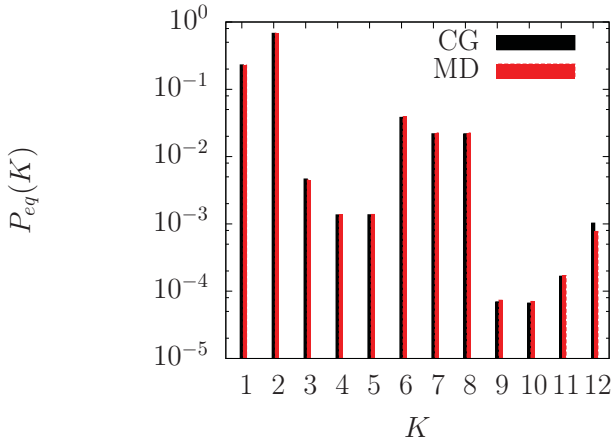


Figure 4.4: *The probabilities of observing each class of events at equilibrium for Kr at loading $\langle n \rangle = 6$ are shown. Results from the coarse-grained model are compared with those obtained from MD runs.*

of others [9, 19, 20]. This is because of the evolution rule for inter-cage events, which depends on the total occupancy of the target cell. A different evolution rule, depending only on the target site of the inter-cage event, could allow the development of a CA model, and a further study in this direction is envisioned.

Strict detailed balance cannot be satisfied in the model, as obviously most of the event-event transitions are not reversible in a single step. Let us consider as an example the transition $i \to j$, where i is again a jump from site a to site b and j is a jump from b to c. The reverse transition $j \to i$ cannot occur as the target site of j and the start site of i do not coincide and going back to a implies some other transition. The shortest route is by a state $c \to a$ if this exist, otherwise a longer detour is needed. What is important is that a way back exists and this is ensured by the nature of the transition matrix.

In our case all matrices are irreducible, non-negative (as the elements represent probabilities), and with at least one positive element on the main diagonal (as any null-event i allows the transition $i \to i$, because of the metastability of sites), and this implies that they are primitive (i.e. there exists a single m such that $[A^m]_{ij} > 0$, where A is the transition matrix). We remark that all systems for which our approach, as well as a generic MSM, is meaningful must be at least ergodic, thus irreducible, and must allow self-loop transitions for null-events, as a meaningful definition of sites implies at least some degree of metastability.

Primitivity of the transition matrices means that the sampling which they lead to is regular, a condition which ensure attainment of the station-

ary distribution. This turns out to be the same equilibrium distribution generated by MD, for all intra-cage events, as can be seen in fig. 4.4, which implies that the balance condition is satisfied at the thermodynamic equilibrium. According to Deem *et al.* [29] satisfying the balance and regularity conditions together is enough to ensure the reliability of a model, even if strict detailed balance is not satisfied.

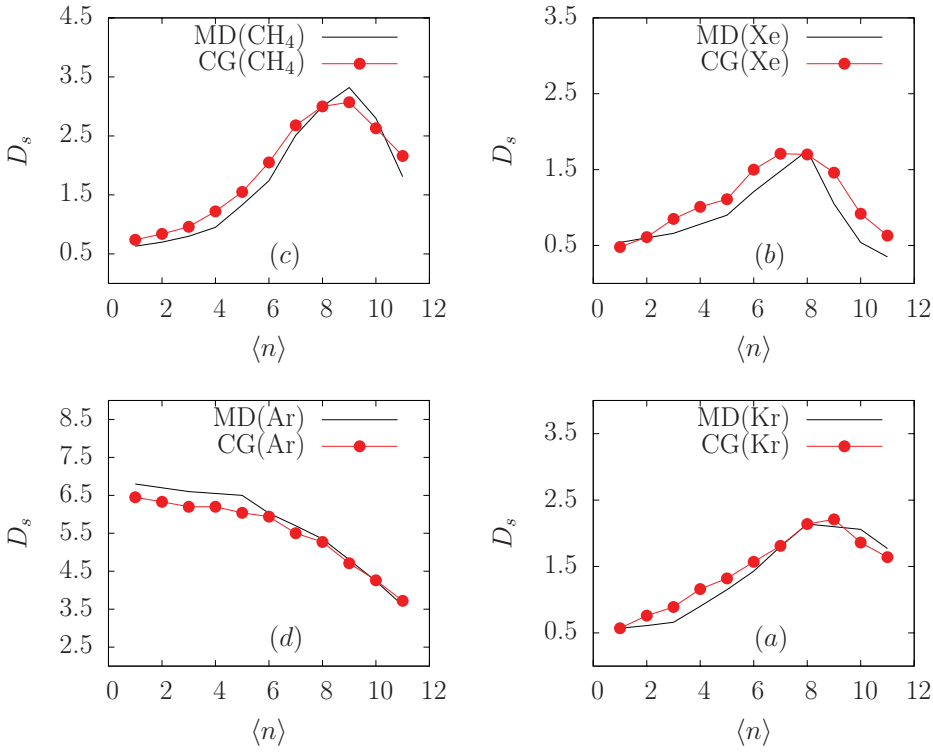


Figure 4.5: *Self-diffusivity ($D_s/10^{-8}m^2s^{-1}$) for the four systems studied as a function of the average number of molecules per cage ($\langle n \rangle$). Results from the coarse-grained model are compared with those obtained by means of MD.*

The only event class that does not satisfy balance is class 12 which has a different treatment from all others as explained above. The development of a proper rule for inter-cage events, satisfying balance and leading to a good occupancy probability distribution, is currently under study. It should be noted that, while finding such a rule is obviously desirable in order to put the model on a really solid basis and improve its performances, the results shown below are already in reasonable agreement with MD results.

4.2.3 Results

We applied the method to CH_4, Ar, Kr, and Xe adsorbed in zeolite ITQ-29 with loadings ranging in $1 \leqslant \langle n \rangle \leqslant 12$. MD simulations were performed with the same overall parameters described in section I. Force field parameters for CH_4, Ar and Kr are taken from Krishna *et al.* [81], while the parameters for Xe are taken from Tunca *et al.* [83]. All the calculations with our method were carried out on a cubic lattice consisting of 125 α-cages for 10^6 time steps, corresponding to 100 ns, with $n_{max} = 16$ for all system. High loading calculations required only few minutes on a single processor. This means that the coarse-grained method is more than 10^4 times faster than MD performed with the previously described settings on a single processor. The equilibrium probabilities of observing each class of event at a given loading are shown in fig. 3, where the results of the simulations are compared with those obtained from MD runs at the same loading. Only an example is shown (Kr at $\langle n \rangle = 6$), which is representative of all the systems studied. All self-diffusivities were calculated by means of the Einstein formula

$$D_s = \frac{1}{6} \lim_{t \to \infty} \frac{d}{dt} \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle \tag{4.9}$$

and are shown in fig. 4.5. In general a reasonable agreement with MD results is observed, and the overall trends are all reproduced, despite the less satisfactory results in the case of Xe. In fig. 4.6 occupancy probability ($F_{eq}(n)$) distributions are compared with those obtained from MD showing reasonable agreement expecially at intermediate loadings.

Deviations from MD results may arise from many different factors, the first being the inherent approximations underlying the transition matrices definition and evaluation, and the algorithm rules. The state reassignation step, applied each time the number of molecules in a cell changes, introduces a certain loss of memory as, in case the original particle state is not one of those available at the new loading, the particle story is somehow reset. Given that most of the correlation effects arise from backscattering, at all loadings apart from very low ones, this will in general causes D_s to be higher than expected. Another problematic point is the choice of $P^{\text{in}}(n)$. Apart from the balance issue discussed above, considering a linear dependence with the same value of n_{max} for all species, as we did, is clearly an oversimplification, and finding a more realistic function of occupancy, taking into account the molecules size and interactions, would surely improve significantly the results. It should be finally pointed out that, apart from the case of CH_4, for which MD calculations were carried out at all loadings between 1 and 12, available MD data were only for $\langle n \rangle = 1, 3, 5, 6, 8, 10, 12$, so that only certain occupancies had a corresponding transition matrix. For the other ones we randomly choose the occupancy to be considered as $n + 1$ or $n - 1$, thus adding an interpolation error. Moreover for occupancies $n > 12$ we

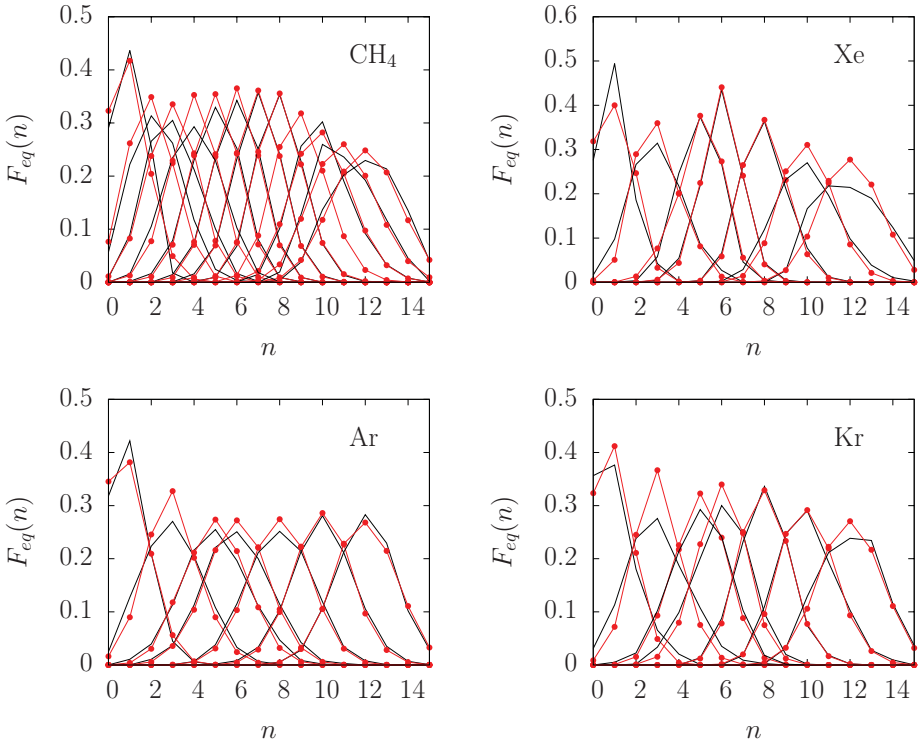


Figure 4.6: *Occupancy distributions at various loadings. Results from the coarse-grained model (lines with points) are compared with those obtained by means of MD (simple lines).*

consider the transition probabilities obtained at $\langle n \rangle = 12$. The occurrence probability of cells with such occupancies is very low for $n \leqslant 10$, but start to affect significantly the results for $n \geqslant 11$.

4.2.4 Conclusions

We proposed a coarse-grained method which stems in a straightforward way from short MD simulations performed on a single zeolite crystallographic cell, and suitable for the extension of the in-silico study of microporous systems with adsorbed molecules to greater spatial and time scales. The application of a simple Markov State Model to these systems is not feasible, at least at room temperature and for weakly interacting adsorbed molecules, but by switching our attention from site-site dynamics to event-event dynamics a relatively simple algorithm was built which is able to reproduce intra- and inter-cage molecular motion, accounting for memory effects despite the lack of direct interaction among molecules. The host-guest system can be seen as a lattice of cells each characterized by its occupancy which

implies a different resistance to the motion of guests, and a certain tendency to release/accept traveling molecules. This simple scheme can be applied to the study of mass transport in microporous materials and could be a candidate for simulating out of equilibrium systems such as membranes under working conditions [84].

4.3 Equilibrium probability distibution of events

We present here the equilibrium probabilities for each class of events, as obtained from the first left eigenvector of the event transition matrix. The analysis of these results gives a qualitative interpretation of the overall behaviour of the various systems studied

Let us consider figs. 4.7 and 4.8, showing how the equilibrium probability of each class of events develops with varying $\langle n \rangle$ in the case of methane. The first thing which comes to the eye is the large gap, throughout the whole range of $\langle n \rangle$, between the probabilities of events of class 1-2 and all the other ones. This will be seen for all other cases reported here, and means that null-events on adsorption sites are by far dominant in the dynamics. The 3-events on the other hand are not so prominent, because the site III is not an adsorption site, if not at very high loadings, and represents only a region of transit in most conditions.

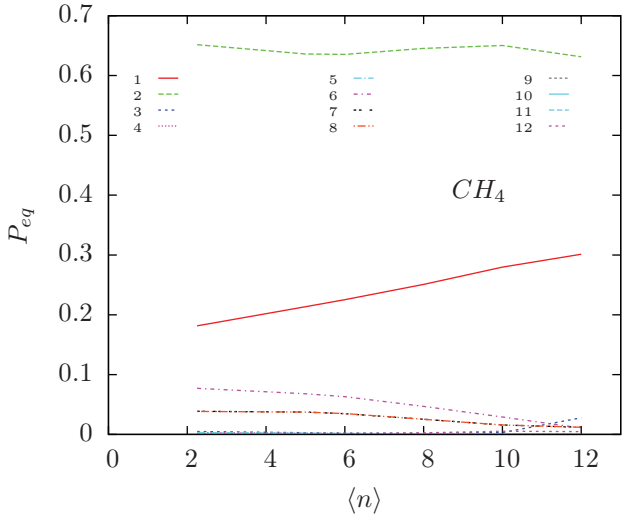


Figure 4.7: *The equilibrium probability of all event classes for methane at 300K.*

The large probability of null-events is to be expected if the coarse-graining of space is a meaningful one, as one usually wishes to have metastable

sites, and this implies a higher probability of residence in one of them, rather than of jumping from one to the other. In particular we notice that 2-events dominates over 1-events, and this seems to be at odds with what we saw in section 3.5: namely that sites of kind I correspond to the six deepest minima in the PES. However we already remarked that at room temperature the shape and connectivity of attraction basins should be taken into account.

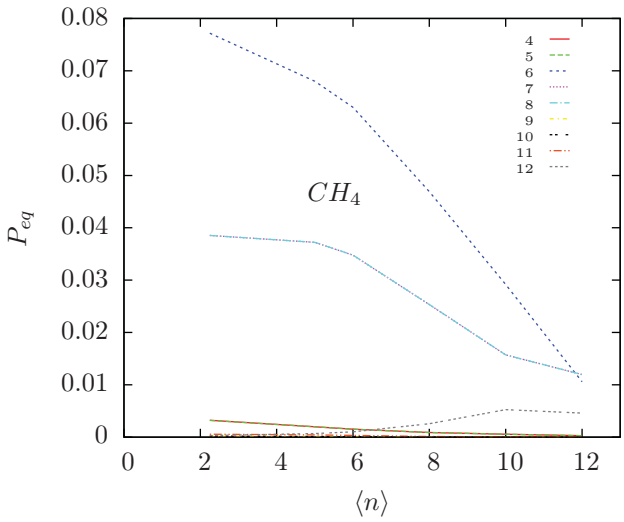


Figure 4.8: *The equilibrium probability of all non null-event classes for methane at 300K.*

In fact the shallow minima found in the energy landscape analysis of sec. 3.5 fall inside the super-basins represented by sites of kind II, and even if these are not as deep as minima of kind I, they result thus to be larger. The connection of the shallow minima to minima of kind II is evident in fig. 3.3, where some of them appear in groups of three as small clumps connected to the II-minima. So molecules tend to spend most of their time in sites of kind II. As the loading increases however this sites are progressively saturated, and molecules populates more and more sites of kind I, and finally sites of kind III, as is evident from the increasing probability of events 1 and 3 in figs. 4.7.

The lower part of fig. 4.7 is enlarged in fig. 4.8, where the curves for non null-events can be seen clearly. The dominant non null-events are event 6,7 and 8, corresponding to the motion on the α-cage inner surface. The fact that the probability of events 7 and 8 coincide at all loadings, confirms that the balance condition is satisfied at equilibrium, and all other pairs of events in which one is the opposite of the other show the same behaviour. Events in and out of site III are very rare.

As can be seen the overall trend for intra-cage non null-events is of

monotonic decrease, as it can be expected when the number of molecules increases, crowding the inner space of the α-cage, and rendering motion more and more difficult. On the other hand we see that the only non-null event whose probability grows higher with loading is event 12, that is the inter-cage event.

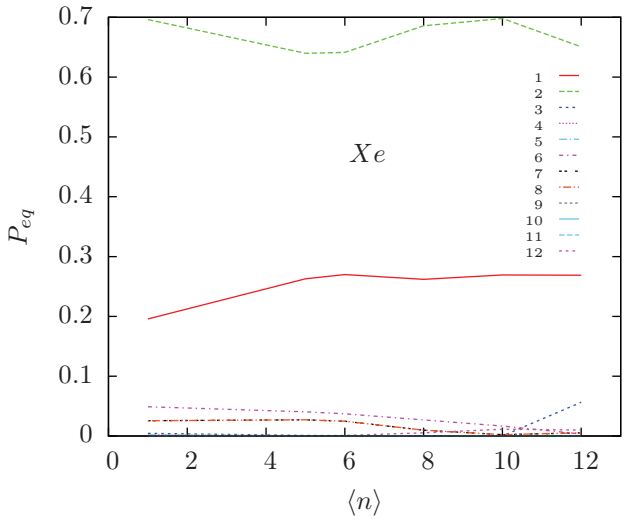


Figure 4.9: *The equilibrium probability of all event classes for xenon at 300K.*

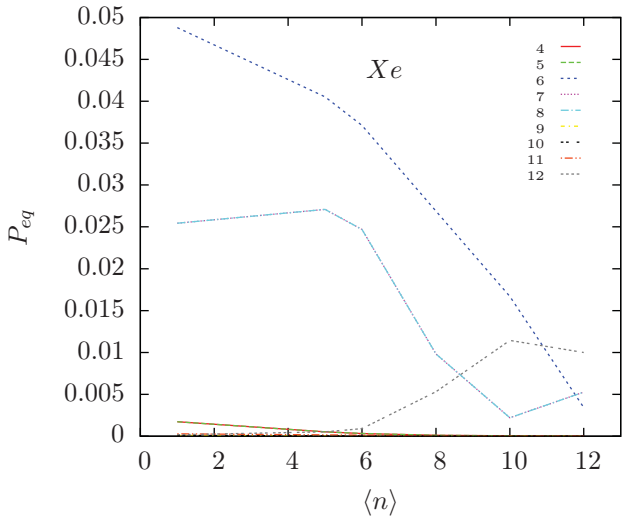


Figure 4.10: *The equilibrium probability of all non null-event classes for xenon at 300K.*

This is very interesting as it seems to point to the possible explanation of the D_{self} trend shown by methane as loading increases. We can reckon that the maximum in the the plot of D_{self} vs. $\langle n \rangle$ is due to the balance between the increasing probability of inter-cage jumps, which are rate determining at low loading, and the decreasing probability of all inter-cage kinds of motion. A similar behaviour for D_{self} is in fact seen in the case of xenon and kripton and from fig. 4.10 it is clear that events probabilities for Xe behave according to the same overall trend. In the latter case curves are steeper: the increase of probability for event 12 is still more pronounced and some of the intra-cage events probabilities fall to zero in the end.

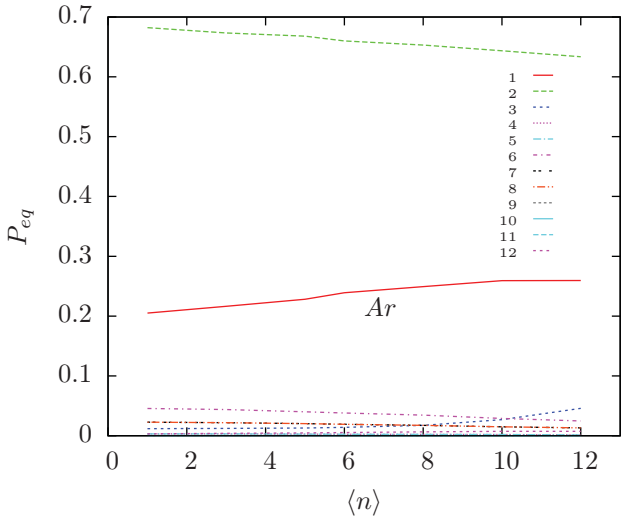


Figure 4.11: *The equilibrium probability of all event classes for argon at 300K.*

This seems to be a good hint to the reason for the somehow counter-intuitive behaviour of CH_4, Kr and Xe. But what happens with a species showing monotonic decrease of self-diffusivity with increasing loading, such as Ar? In fig. 4.12 we can see an overall similar trend for Ar. However curves are less steep in this case, and it seems that the increase of loading has a weaker effect compared to the cases of methane and xenon. This can be explained with the smaller size of argon, and the fact that xenon, which is larger than methane, shows the strongest dependence on $\langle n \rangle$, corroborates this idea. It is intuitive to think that adding a small Ar atom has a small influence, particularly at low loading, on the behaviour of the other guest molecules, while adding a massive Xe atom has a stronger impact. However if the probability of intra-cage jumps decreases and that of inter-cage ones grows as for the other systems examined, why does the plot of argon self-diffusivity looks completely different? This could be explained considering

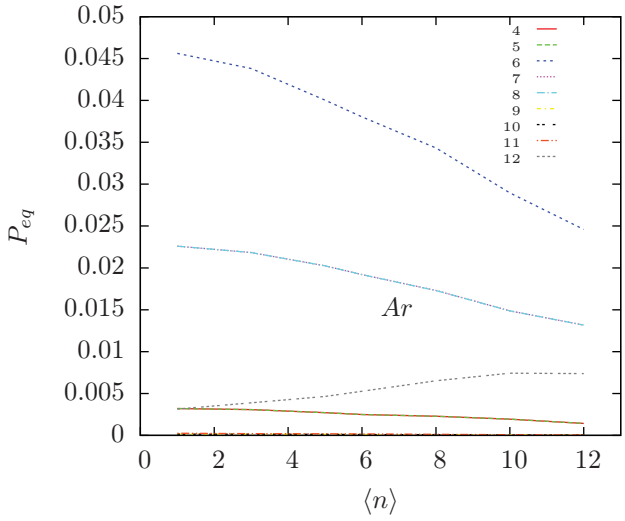


Figure 4.12: *The equilibrium probability of all non null-event classes for argon at 300K.*

the relative contribution to diffusion of the various events. As we saw in sec. 3.5, the energy barrier for inter-cage diffusion of argon is one order of magnitude smaller than those of methane and xenon, and in fact smaller than intra-cage barriers. Even considering entropic effects, we can thus expect that inter-cage jumps do not play as strong a bottleneck role in diffusion dynamics as for the other systems. In fact the equilibrium probability of event 12 for Ar, is lower than those of events 6, 7 and 8, but is by far higher than in the case of CH_4 and Xe. So its increase, beside being less pronounced than in other systems, has a minor impact on the diffusion dynamics and the decreasing intra-cage mobility plays the crucial role in determining the self-diffusivity trend.

A very different case is that of CO_2. As we have already seen, and is evident in fig. 4.13 sites of kind I plays a very different role for this sorbate species. At low loading molecules spend most of their time in the channels connecting cages and are progressively forced to populate the inner part of the cage as $\langle n \rangle$ grows. In fact the definition of sites used for the other systems is not best suited for this case, as we know that the 'channel-site' comprises part of two sites of type I, belonging to two adjacent cages. So event 12 is not the jump between two metastable states, but it connects two parts of what we know to be a single metastable state, whose total residence probability is the sum of the probabilities of events 1 and 12. A different partition of space should be used for this system, however we present it here as an example of how far one can go with an arbitrary choice of macrostates.

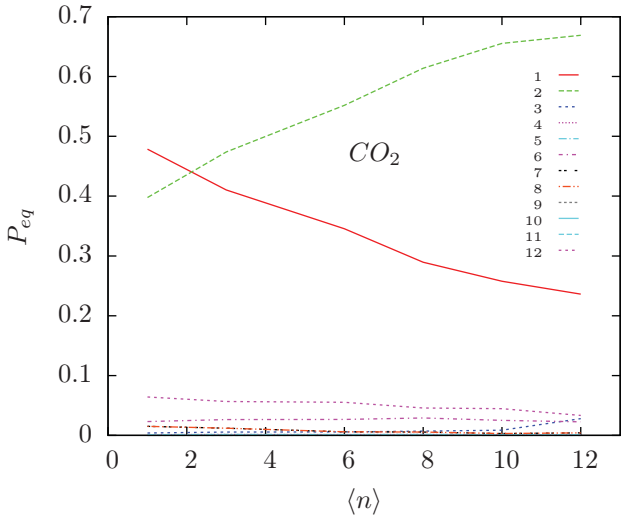


Figure 4.13: *The equilibrium probability of all event classes for carbon dioxide at 300K.*

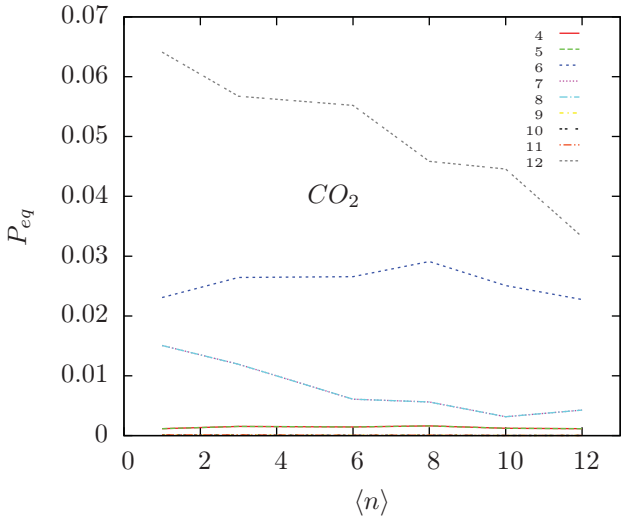


Figure 4.14: *The equilibrium probability of all non null-event classes for carbon dioxide at 300K.*

The non null-events in fig. 4.14 show a less definite trend if compared to what we saw before. Most of the events do not show a clear increase or decrease of their probability, apart from events 7,8 and 12. This can be related to a general crowding of space, but while event 12 is not in any way

a significant event for diffusion, as we have seen that it take place inside a single effective macrostate, events 7 and 8 probably determine the monotonic drop of D_s.

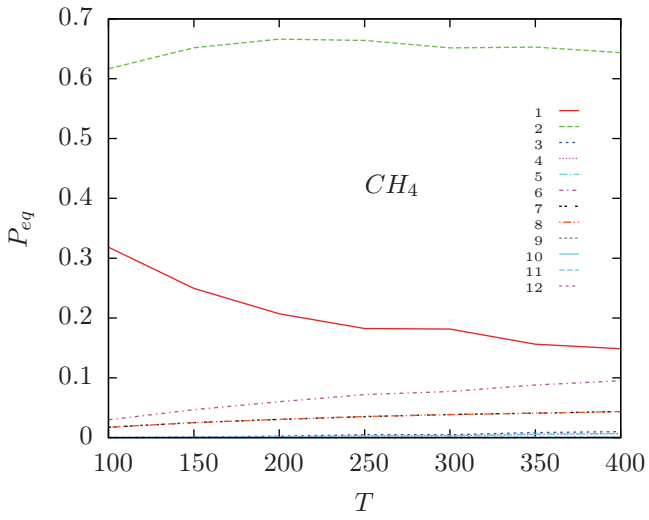


Figure 4.15: *The equilibrium probability of all event classes for methane at* $\langle n \rangle = 1$*, as a function of temperature.*

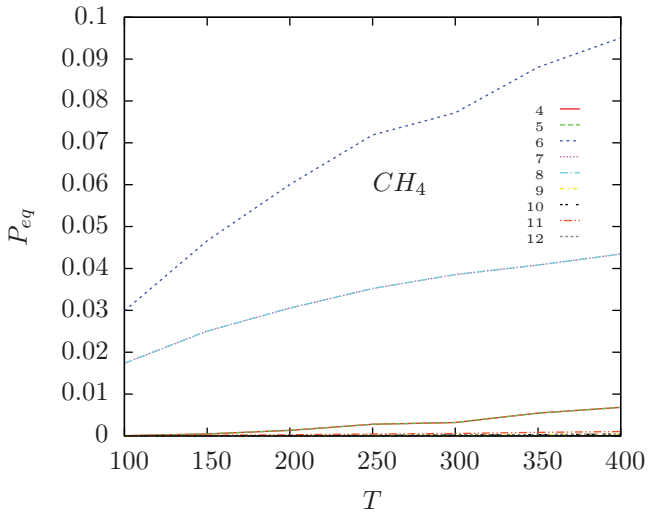


Figure 4.16: *The equilibrium probability of all non null-event classes for methane at* $\langle n \rangle = 1$*, as a function of temperature.*

The last two plots refer to the evolution of the event probabilities with varying temperature in the case of CH_4. In fig. 4.15 we notice how the probability of residing in site I decreases as the temperature is raised. This happens as on one hand, for the entropic effect already discussed, molecules tend to spend their time in sites of type II more and more with increasing T, and on the other hand because of the steady increase of mobility. As event 1 is a static event, it loses probability to non null-events. The same must be true of event 2, which is static as well, but the effect is counterbalanced by the increased probability of occupying II-sites.

All non null-events probabilities, as shown in fig. 4.16, increase monotonically with temperature, and this is reasonable as all kinds of motion are enhanced by the increasing kinetic energy of molecules. So in this last case there is a simple, direct relationship between the behaviour of D_{self} and the equilibrium population of events.

Before concluding this section it is worth spending two words on site III and correlated events. As can be seen for all the species considered, event 3 is practically negligible at low loading, but its probability grows with $\langle n \rangle$, and in particular for $\langle n \rangle > 10$. This is because, as already remarked, site III does not correspond to a region of preferential adsorption. As we have seen in sec. 3.5, there is no energy minimum at the center of the cage, nonetheless, as the number of molecules in the α-cage grows, the adsorption regions on its surface become progressively saturated, and in the end some molecule is forced to occupy the center of the cage. The most marked increase of the probability of event 3 is observed for Xe. It seems that up to ten molecules can be accomodated on the surface of the α-cage, distributed among sites of type I and II, but as this threshold is exceeded, this arrangement loses stability: a molecule can leave a site of kind II, and occupy site III, as it is evident from the opposite change of the probability of event 2 and 3 in fig. 4.9. At the same time, events 7 and 8 experience an inversion of their decreasing probability trend. This could be explained with the vacancy left by the molecule which now occupies the center of the cage, allowing again an exchange between sites of kind II and I, which was practically blocked, because of their saturation, at $\langle n \rangle = 10$. A similar line of reasoning can be applied to the other cases. These show similar, but progressively weaker phenomena, in the order $CH_4 > CO_2 > Ar$.

Chapter 5

The event-event transition matrix as a network

5.1 Network theory

In recent years an impressive amount of work has been devoted to the analysis of the most diverse problems in terms of graph theory [85]. A graph is an ordered pair (V, E), where V is a set of vertices, or *nodes*, and E is a set of pairs of element of V, called edges or *links*. These can be unordered or ordered pairs, and in the last case the graph is a directed graph, or *digraph*, in which the link connecting node v_i to node v_j and that connecting v_j to v_i do not coincide. Virtually any system can be represented as a graph, by mapping the objects of which it consists to the set of nodes of the graph, and the relations among these objects to the set of edges [86]. When a graph, which is in itself a perfectly abstract object, represents a real system, or in general every time we interpret its elements in terms of something else, we speak of a *network*. Network theory concerns any attempt to understand systems and phenomena (and usually complex ones), using graph theory concepts and machinery. Often the proper description of a real world system requires directed graphs and the introduction of *weights* characterizing each link. The weight of a link may represent the distance between the nodes it connects, the strength of their relation or the flow of something between them. Here we give a brief overview of some of the very basic 'metrics' of a network [85, 87]:

-*degree*: the number of links attached to a node v_i is the node's degree k_i. This simple definition is the only possible for undirected networks, while in the case of a digraph an in-degree k_i^{in} and an out-degree k_i^{out} can be defined corresponding to the number of links entering the node and the number of links leaving it, respectively. An overall degree can be defined as the sum of the two. In the case of an undirected weighted network it is useful to

define the *strength* of a node as the sum of the weights of all edges attached to the node $s_i = \sum_{j=1}^{k} w_{ij} + w_{ji}$. If the network is a directed one then we have again s_i^{in}, s_i^{out} and a global s_i, defined in a way analogous to k_i^{in}, k_i^{out} and k_i for unweighted digraphs. One of the most characterizing features of a network is the distribution $F(k)$ of the degree of its nodes. Regular networks have $k_i = k\ \forall i$, as their structure is perfectly homogeneous, while random graphs (see below) usually show a gaussian or a Poisson distribution, with a well defined average $\langle k \rangle$. However the most interesting networks are those showing a *scale-free* degree distribution, obeying a power law relation of the form $F(k) = e^{-\gamma k}$, where γ is a small positive real number.

-*average path length*: the *distance* between two nodes v_i and v_j in a graph is the length of the shortest path connecting them. In an unweighted graph this is simply the number of links on the path, while for a weighted graph also weights should be taken properly into account, depending on their interpretation. The *average path length* is the average distance between any two nodes in the graph. A peculiar feature of most real networks is the very short average distance between nodes, scaling roughly as lnN, where N is the number of nodes in the graph. A network with this characteristic is a *small world* network.

-*clustering coefficient*: the ratio C of the number of closed triples, that is triples of nodes each connected to each other, to the number of all possible triples in the graph (both closed and open, i.e. with only two out of three possible links), is called *clustering coefficient*. The ratio of the number of closed triples to the number of all possible closed triples is called transitivity ratio. C is a global measure of the degree to which nodes in the graph tend to form clusters. It is useful to define also a *local clustering coefficient* C_i for each node v_i. Watts and Strogatz [88] define it as:

$$C_i = \frac{|e_{jk} : v_j, v_k \in N_i, e_{jk} \in E|}{\varepsilon_i} \tag{5.1}$$

where the numerator is the effective number of edges linking nodes in the neighborhood N_i of v_i, and ε_i is the maximum possible number of links among nodes in N_i, and we have $\varepsilon_i = k_i(k_i - 1)$ for directed graphs, and $\varepsilon_i = k_i(k_i - 1)/2$ for undirected ones. The *average clustering coefficient* for the whole network will be then the average of C_i over all the nodes of the graph (and is in general different from the clustering coefficient defined above).

-*centrality measures*: there are many different ways of defining the centrality of a node v_i in a graph. The degree and the strength are two examples. It is intuitive to think that a node will be more central, or influential, or significant, if it is connected to many other nodes. In the case of a weighted

graph it is seems intuitively more appropriate to consider the strength. However it should be noted that strength alone is not a very sensible measure, as we will see below. Two useful definitions are those of *betweenness centrality* and *closeness centrality*, which answers to the questions: 'how much is the node v_i involved in the shortest routes connecting the other nodes?' and 'how close is the given node to all others?' respectively. Betweenness is formally defined as

$$C_i^B = \sum_{j \neq i \neq l} \frac{\sigma_{jil}}{\sigma_{jl}} \tag{5.2}$$

where σ_{jil} and σ_{jl} are the number of shortest paths connecting v_j to v_l via v_i, and the total number of shortest paths from v_j to v_l , respectively. Closeness is defined as

$$C_i^C = \frac{1}{\sum_{j \neq i} d_{ij}} \tag{5.3}$$

where d_{ij} is the distance between v_i and v_j, and $\sum_{j \neq i} d_{ij}$ is the *farness* of node v_i.

-*modularity*: a network may present a modular structure, with groups of nodes forming tightly bound clusters. Many and/or strong links exist among members of a cluster while connections to members of other clusters are sparse. Many different *modularity* definitions have been given in literature, aiming to quantify the degree to which a network is split in such clusters, also called *modularity classes*.

Many real networks show peculiar features which makes them different from both regular networks (such as lattices) and random ones, and can be classified as *complex networks*. Namely they tend to differ for three main reasons: they usually show the small world property, tend to have high clustering coefficients and the degree of nodes follows a power law distribution, i.e. they are scale-free. Small world random graphs [89] can be generated via the Erdős-Rényi approach, but they do not show neither high clustering nor power law distribution of k. Generating graphs with all these three features is not easy, the Watts-Strogatz and the Albert-Barabasi methods, for instance, give more realistic graphs if compared to those generated by the Erdős-Rényi algorithm, but they are able to reproduce only two out of the three aforementioned features: apart from a short average path length, the first generates graphs with high clustering coefficient but not scale-free k distribution, while the second gives scale-free graphs, but with low clustering.

5.2 A network of events

The analysis of the transition matrix described in the previous chapter is not easy with the algebraic approach described in section 3.3. This is because it is in general a non-symmetric matrix and it does not satisfy the detailed balance condition, so that there seem to be no obvious way of transforming it into a symmetric matrix. Being non-symmetric implies that the existence of a complete basis of real eigenvectors for it cannot be ensured. The kind of eigenspectrum analysis described in section 3.3 and 4.1 cannot in general be applied, at least directly, to it, because part of the eigenvalues are complex, thus defying the requirements of PCCA, and cannot be related in a meaningful way to the characteristic time scales of the processes taking place in the system.

The existence of a symmetrizing transformation cannot be excluded, however at the moment we are not aware of any such operation.

A different, intriguing approach to the event-space transition matrix is that offered by network theory. In fact any matrix can be translated to a graph, and vice versa.

In the last years different attempts to apply networks theory tools and concepts to chemistry have appeared in literature. A great part of them was devoted to represent and unravel the complex interactions of molecules involved in genetic expression, metabolic reactions and other biochemical processes. Some really interesting studies also focused on the same kind of problems treated in this thesis. Doye *et al.* [90] analyzed the topology of the energy landscape of a 14 Lennard-Jones atoms cluster, by representing it as an undirected, unweighted network, where nodes represent minima, and their connectivity mirrors the connectivity of minima in the PES. They also proposed an explanation for the power-law distribution of the degree in energy landscape networks, based on an Apollonian packing model of the PES [91].

The network theory approach was used for studying the conformation space of dipeptides, proteins, water clusters and other systems [92–95], highlighting the degree, weight and clustering coefficient distibutions for these systems, and their correlation with the statistical weight of the configurations represented by nodes. Coarse-graining techniques based on the configuration space network connectivity were also proposed, in which the metastable aggregates of states can be defined as modularity classes [96], or according to a random walk based procedure preserving the main spectral properties of the original network [97].

It should be noted that most of these networks are undirected or unweighted. They all represent the connectivity of configurations, or PES minima, by putting an edge linking any two of them for which a transition was observed in a MD simulation, or a connecting TS is known to exist, respectively.On the contrary we are going to define a weighted digraph, where

each node represents an event, as defined in the previous chapter, and each edge represents a transition between two events.

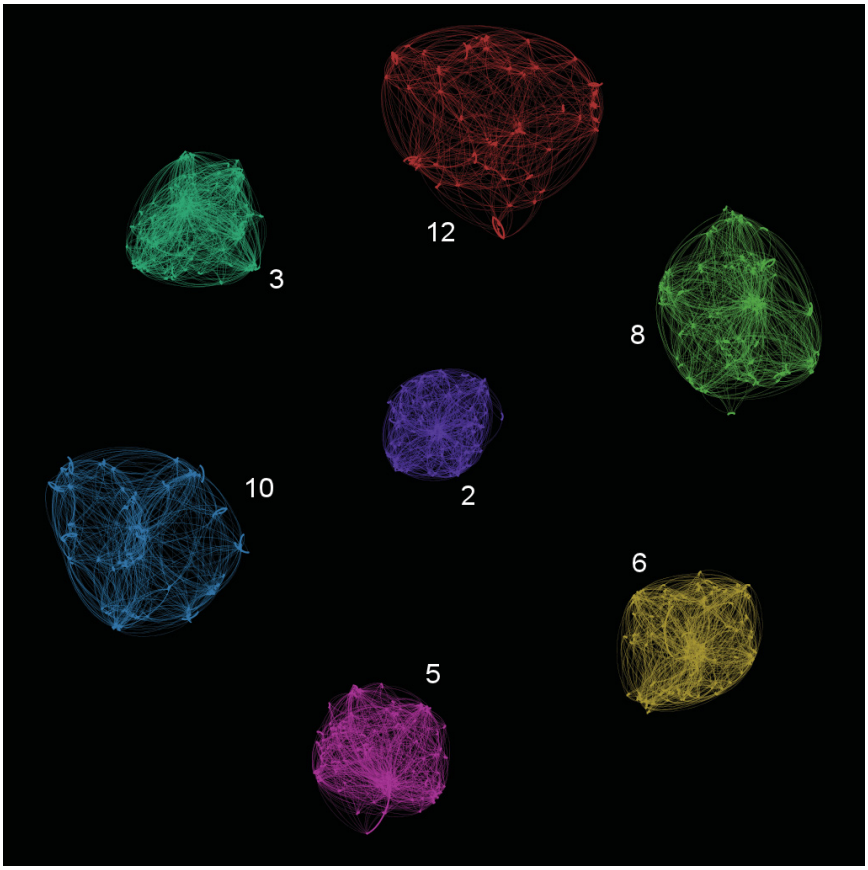


Figure 5.1: *The networks obtained from MD simulations of methane at various loadings are shown. As all other networks graphic representation, this figure was generated using Gephi [98]. The algorithm used for displaying the networks is force atlas.*

Given our $n \times n$ event matrix A we can build a n nodes graph representation of it by connecting each pair of nodes i and j if the entry A_{ij} of the matrix is non null, and attaching to the edge e_{ij} a weight $w_{ij} = A_{ij}$.

In this way connections between nodes are non-symmetric and they are characterized by a weight, representing the probability of transition between the events represented by the nodes, within the given lag time. This network is a representation of the dynamics of the system, but what information can we extract from it and how? It must be made clear here that a definite answer to this questions won't be found in these lines. However we will try to give an overview of the problems faced in the analysis of a weighted

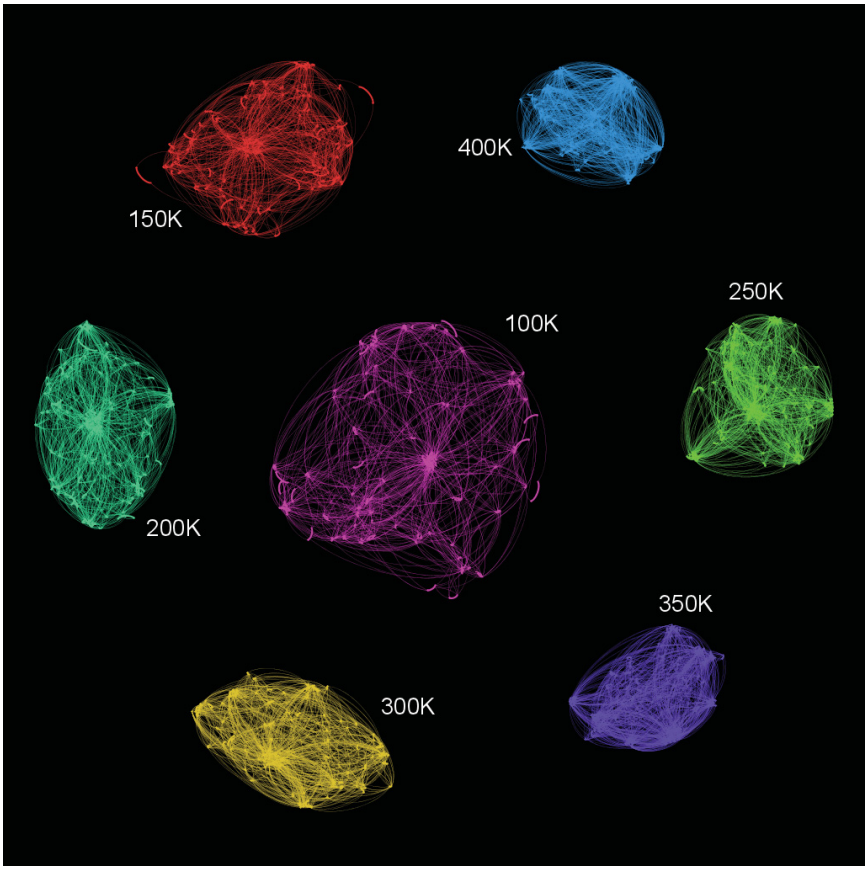


Figure 5.2: *The networks obtained from MD simulations of methane at various temperatures are shown. The algorithm used for displaying the networks is again force atlas.*

digraph representing events in a molecular system, and of the perspectives for further investigations.

We remark that this is a preliminary study, aiming to gain confidence with the matter. As explorative tools we employed the free software GEPHI [98], mostly intended for graph visualization and simple analysis, and the TNET [99] package for weighted network analysis using R [100]. We restricted ourselves to the network metric calculations available with these softwares, as long as they were suitable for our purposes, while more sophisticate calculations are left for further studies.

As a first question, one could ask if there is any immediate overall characteristic of the network which is directly connected to the diffusion trends observed for the various systems. A number of event networks for methane are shown in figs. 5.1 and 5.2, corresponding to various loadings and tem-

peratures respectively. The visualization algorithm employed is called *force-atlas*, and arranges nodes so that their distance is proportional to the number and strength of their links. This can immediately highlights the presence of highly connected clusters of nodes and how they are related to other clusters. What can be immediately realized in figs. 5.1 and 5.2 is the progressive shrinking of the graphs with decreasing loading and with increasing temperature, respectively. This means that nodes, and thus events, tend to be more tightly bound at high temperature and low loading. Should we expect then a corresponding increase of diffusivity in this conditions? If all events are closer together as the graphical representation vividly suggests, it seems reasonable to think that the molecular motion is more free and should give a higher diffusivity. But while this is true for the case of varying temperature at fixed loading (in this case $\langle n \rangle = 1$), for the case of varying loading we know that the trend is very different from what fig. 5.1 suggests: self-diffusivity increases up to $\langle n \rangle \sim 9$ and then it decreases.

This means that no such simple general relationship as we suggested before can always be expected, and while it is roughly valid in the case of varying temperature, in the case of varying loading other characteristic of the network structure must be considered, if one hopes to relate the observed behaviour to the network architecture.

K	Start	Target
1	I	I
2	II	II
3	III	III
4	II	III
5	III	II
6	II	II*
7	I	II
8	II	I
9	I	III
10	III	I
11	II	II†
12	I	I‡

Table 5.1: *Event classes observed within a lag time of 100 fs in MD trajectories of various sorbates inside zeolite ITQ-29. Event classes are defined by the site kind of the start-site and target-site.*
* *the site is adjacent to the start site*
† *the site is not adjacent to the start site*
‡ *the site is in the adjacent cage*

A number of network metrics show the same trend graphically seen in figs. 5.1 and 5.2. The in-degree and out-degree for the two cases of methane

are shown in figs. 5.3, 5.4, 5.5 and 5.6.

Both k_{in} and k_{out} increase with temperature and decrease with loading, in line with what we saw before. This is reasonable, as, when the temperature is raised, progressively more and more events become possible within the given lag time, and their connections increase in number. At low temperature almost all jumps from a site to another are followed by a long stasis in the target site, meaning that in the network the direct links between non-null events are very rare. As the temperature is raised the probability of consecutive jumps becomes non negligible and new links appear. On the contrary, adding molecules in the cage has the effect of lowering both the number of events and their connectivity, as new guests tend to hinder mobility.

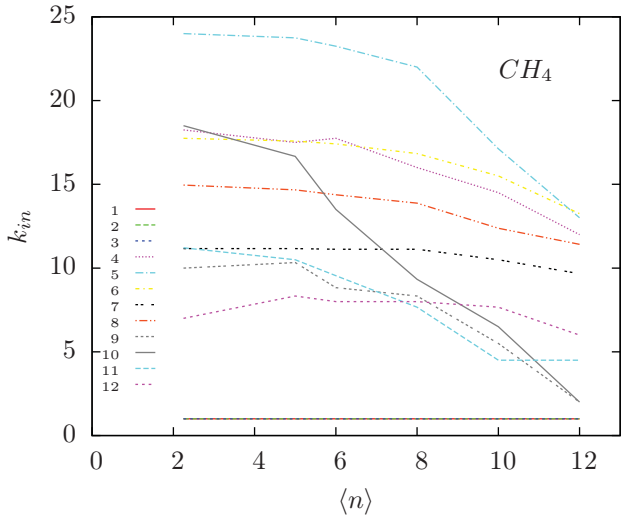


Figure 5.3: *Average in-degree for all event classes of methane, plotted against loading.*

An important exception to this trend is that of class 12 events (as we will often refer to it, the table with the definition of all event classes is reproduced in this section (tab. 5.1).) : for this class k_{in} increases at first, then it remains more or less constant and then decreases only for $\langle n \rangle \geqslant 10$, while k_{out} is constant up to $\langle n \rangle = 10$ before decreasing. This is worth of attention, as inter-cage jumps have a prominent role in diffusion of methane. As we have seen in sec. 4.3, it seems that the non monotonic trend of diffusion for methane, xenon and kripton can be explained in terms of two contrasting phenomena connected with increasing loading: an increase of the probability of performing an inter-cage move, and the dwindling possibilities for intra-cage motion. It would be interesting to understand how the nodes of kind

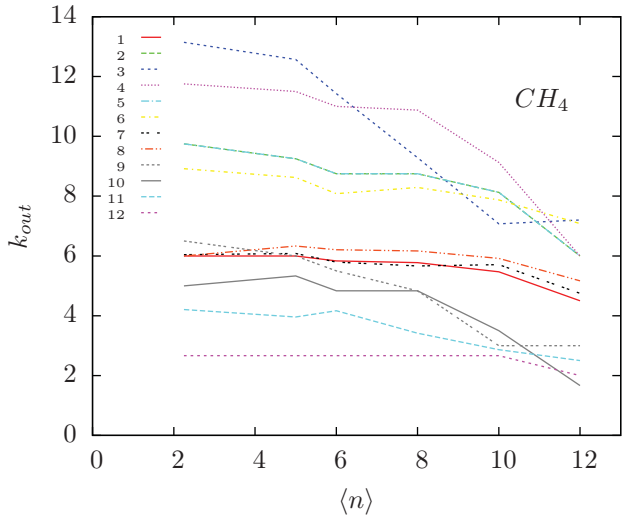


Figure 5.4: *Average out-degree for all event classes of methane, plotted against loading.*

12 gain probability, and this is a first observation in this direction: as $\langle n \rangle$ increases, it connects to more and more other events. Until now we did not take into account the weight of links. However the weight plays a crucial role, as the number of edges entering or leaving a node can in principle remain unchanged in a network like ours, and weights can make all the difference as conditions are varied. So we consider the strength (see figs. 5.7 and 5.8), and notice that in general the overall trend is not so clear as for degrees. In the case of varying temperature s_{in} of class 2 and 3 slightly decreases, while for class 1 it remains almost constant. The first fact can be explained with the overall increase of mobility implying a smaller in-flow for null-events. Class 2 events suffer the same effect, but they experience at the same time an increment due to the progressively less attractive nature of sites of type I (as was explained in discussing the equilibrium probabilities of the various classes): as the probability of moving from site II to site I decreases, the flow of probability into event 2 grows, while for events that leave site II it decreases. The other non-null event classes showing decrease with growing temperature are indeed those pointing to sites of type I, namely class 8 and 10.

One could be in fact tempted of interpreting degree and strength in terms of the probability of observing a given event. However it should be stressed that these are strictly local properties, accounting only for the number of nodes directly bound to the one under examination and the flow of probability of its first neighbours toward and out of it, respectively. While these

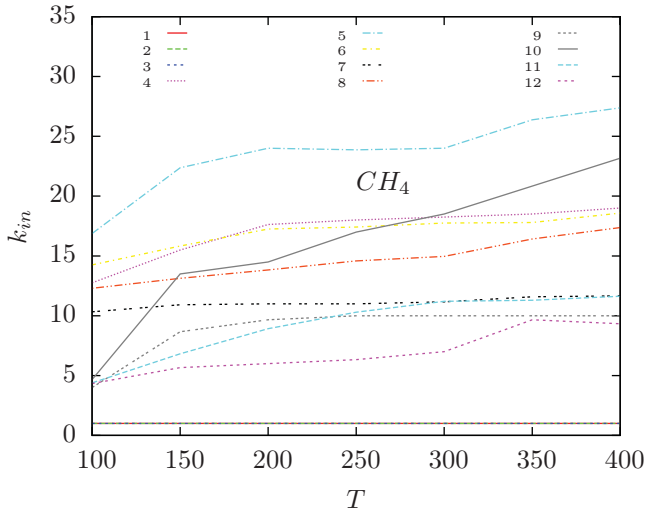


Figure 5.5: *Average in-degree for all event classes of methane, plotted against temperature (K).*

properties are of course related to the overall probability of observing that event, they are only a part of the story. A node can have high in-strength for instance, and nonetheless the probability of observing the corresponding event can be low, if its neighbours are weakly bound to the rest of the network. This is particularly clear in the case of class 5 events. As can be seen in figs. 5.7 and 5.8, this class show the highest value of s_{in}, but as it can be seen in fig. 4.8, it is globally a class of rare events compared to others with lower s_{in}, and this is due to the 'instability' of site III at low loading. As $\langle n \rangle$ increases the events of class 5 show a drastic drop of s_{in}, because as other sites become unavailable, site III is 'stabilized' in the sense that leaving it become more and more difficult (at the same time the in-strength of events of class 3 increases).

For the rest we see a somehow more complicated situation in fig. 5.7, where each class of events seems to have an independent behaviour. However, at a closer look, certain interesting features emerge. We see again that the non-null event class showing the strongest general increase, up to $\langle n \rangle \geqslant 10$, is class 12. Some pairs of event classes seem to have a specular plot, in particular class 2 with 1, and 6 with 8. This points to the fact that the events from which the incoming edges for class 2 and 6 originate, tend to distribute more of their s_{out} to events targeting sites of type I rather than II as the loading increases, because of the latter's saturation. In fact 2- and 6-nodes correspond to II→II events, and lose s_{in} to 1- and 8-nodes, corresponding to I→I and II→I events, respectively. We remark here that,

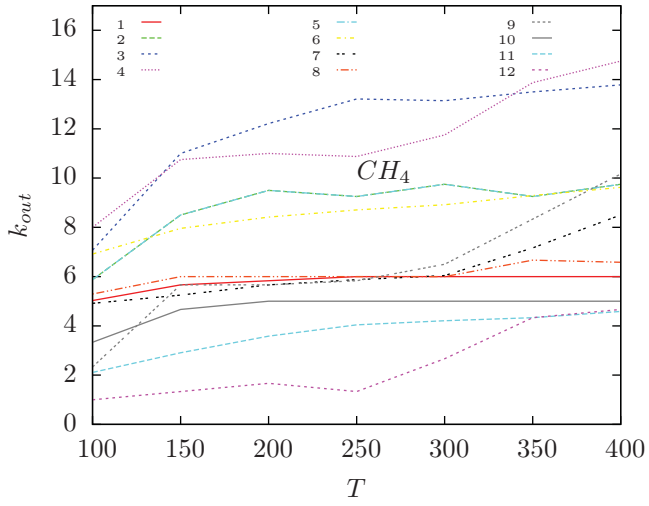


Figure 5.6: *Average out-degree for all event classes of methane, plotted against temperature (K).*

given the stochastic nature of the event transition matrix, $s_{out} = 1$ for all nodes, and this is why we do not discuss it here. However this fact should be remembered in order to understand certain facts concerning s_{in}, and the last observations are based on it: for example, as s_{in} of II→I events increases, that of II→II (and II→III) events should be expected to decrease, as the nodes from which their incoming edges come from are the same, namely those corresponding to events targeting II-sites, whose s_{out} is constant.

Apart from these and other similar observations on particular cases, we are not able, at present, to say much more on the interpretation of degree and strength plots, and more reflection on the data presented may be needed in order to highlight the key for a meaningful and comprehensive explanation of the role played by these network metrics in the diffusion and adsorption properties of the systems examined. If such a key exists.

However it can be that, given the aforementioned inherent limitations in considering degree and strength for themselves, more sophisticate related metrics are needed. Opsahl *et al.* [101] proposed a different way of describing the connectivity of a node, which can account for both the number of connections and for their weight, at one time. They define this generalized degree centrality as:

$$C_D^{w\alpha}(i) = k_i^{(1-\alpha)} s_i^{\alpha} \tag{5.4}$$

where α is a positive tunable parameter. When $\alpha = 0$ we have $C_D^{w\alpha}(i) = k_i$, while for for $\alpha = 1$ obtain $C_D^{w\alpha}(i) = s_i$. By tuning this parameter we

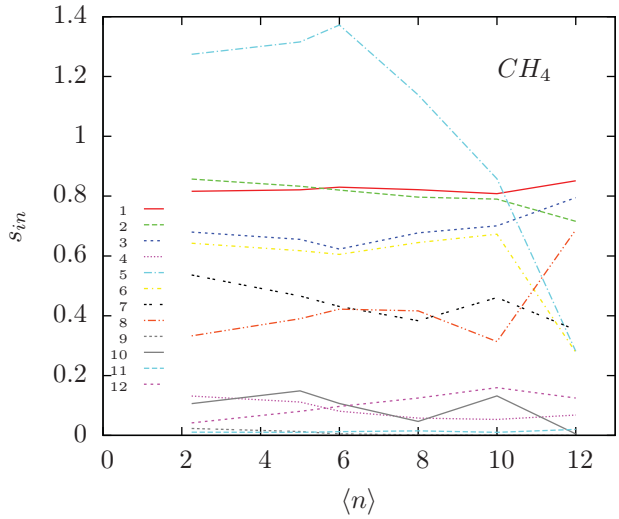


Figure 5.7: *Average in-strength for all event classes of methane, plotted against loading.*

can make $C_D^{w\alpha}(i)$ act in different ways. For example, given a certain fixed strength, if $0 < \alpha < 1$ the measure increases with the number of links, while if $\alpha > 1$ it favours nodes with less connections. In other words it can be higher for nodes which distribute strength on many links or vice versa. Different values of α can thus highlight different subtle aspects of a node's connectivity, and in particular how strength and degree are interrelated. This can shed light on interesting behaviours and correlation, and it would be interesting to investigate a whole range of values of α, to see if more illuminating tendencies emerge, linking the strictly local connectivity to the behaviour of the system.

Even by considering the degree centrality in this way it remains the simplest of centrality measures, and it is largely insensible to the global structure of the network. So what happens when we turn our attention to other 'long range' centrality measures, such as betweenness and closeness? We remark here that this measures are primarily defined for unweighted networks. Opsahl *et al.* [101] defined two measures suitable for weighted graphs, based on a different notion of distance which is intended as the sum of weights connecting two nodes.

Moreover a parameter α can be again introduced, playing a similar role to that seen for the degree centrality. This metrics seem to be better suited for our network, however it must be remarked that, while considering the weights seems to be in general a better choice if one wants to deal with weighted graphs, the real meaning of these weights must also be taken into

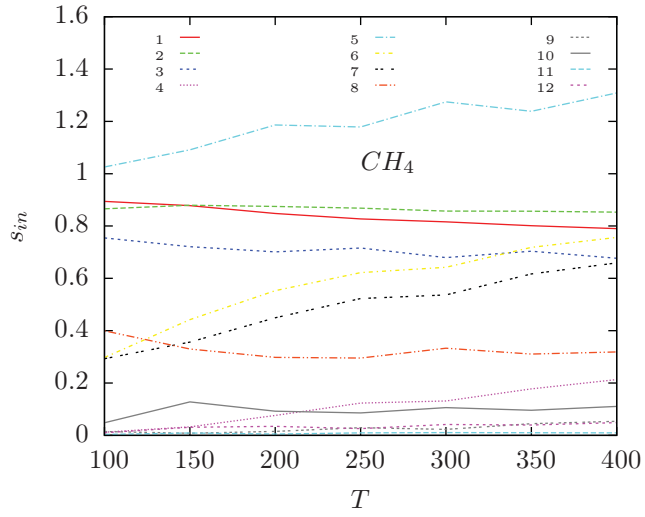


Figure 5.8: *Average in-strength for all event classes of methane, plotted against temperature (K).*

account. Given that ours represent probabilities of moving from one event to the other, we should define the distance between two nodes as the product of weights along a path, rather than as their sum. So, for the moment, we rely on the simple unweighted definitions of closeness and betweenness (eqs. 5.2 and 5.3) to see if they show any interesting behaviour.

In figs. 5.9 and 5.10 the networks for methane, respectively at various loadings and temperature, are shown in different colors, and the vertical position of nodes corresponds to their closeness value. The size of nodes is proportional to their total degree. This is only one of the many possible ways of displaying the networks, each highlighting the correlations existing among the various measures, where the latter can be represented by colors, size or position, depending on what one is interested to see.

What we notice at first sight is an overall similar trend to that already seen for k_{in}, k_{out} and the global networks appearance in figs. 5.1 and 5.2: closeness increases on average with T, and decreases with increasing $\langle n \rangle$. Moreover the distribution of C_i^C is narrower at high loading and temperature (apart from the lowest one), and there seem to be lumps of nodes with similar closeness, separated by more or less clear gaps. What we discover by analyzing the identity of nodes, is that those with highest closeness are all related to site III (i.e. having it as target or starting point), which occupies the center of the α-cage. At intermediate values we find events related to sites of kind II, while events related to sites of kind I show the lowest values. This makes sense as we are considering as distance only the number of edges

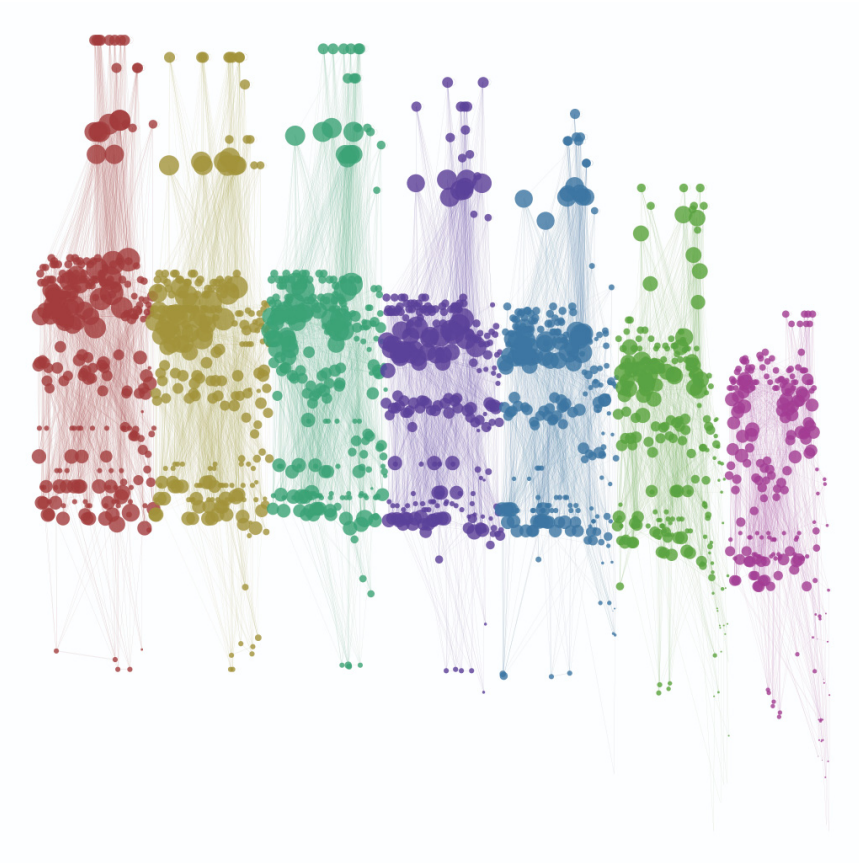


Figure 5.9: *The networks of methane corresponding to loadings 2, 3, 5, 6, 8, 10 and 12 (each shown in a different colour, and arranged in ascending order along the horizontal axis). The vertical displacement of nodes is proportional to their closeness centrality value, while their size is proportional to the total degree.*

separating nodes, and events related to the central III-site are only one or two edges away from almost all others. However the situation could change completely by considering the weights.

In figs. 5.9 and 5.10 no evident correlation between the closeness and the degree of nodes seems to appear. On the contrary, fig. 5.11, showing the networks corresponding to various $\langle n \rangle$, with the size of nodes proportional to their degree and the vertical coordinate corresponding to their betweenness, makes evident that the degree of a node is strongly correlated with its betweenness centrality. In particular, as one would intuitively expect, high degree nodes tend to have the highest betweenness values. This is because nodes with many connections are more likely to be found on the shortest

Figure 5.10: *The networks of methane corresponding to T=100K, 150K, 200K, 250K, 300K and 350K at* $\langle n \rangle = 1$ *(each shown in a different colour, and arranged in ascending order along the horizontal axis). The vertical displacement of nodes is proportional to their closeness centrality value, while their size is proportional to the total degree.*

paths connecting other nodes. As a final example we present the networks obtained from simulations of CO_2 in fig. 5.12, displayed in a way analogous to figs. 5.9 and 5.10. In this case no overall trend can be recognized as we did for methane, apart from the progressively lower closeness value as we move away from events related to site III. This behaviour leave an open question, as we are not able, at present, to tell if the fact that nothing here seems to have any connection with the diffusivity trend is due to some true characteristic of this system or to the fact that, as already remarked, the partition of space used is not very well suited for the case of CO_2. So it seems that very little can be said at present on the relationship of betwenness and closeness centrality of the nodes in the networks with the

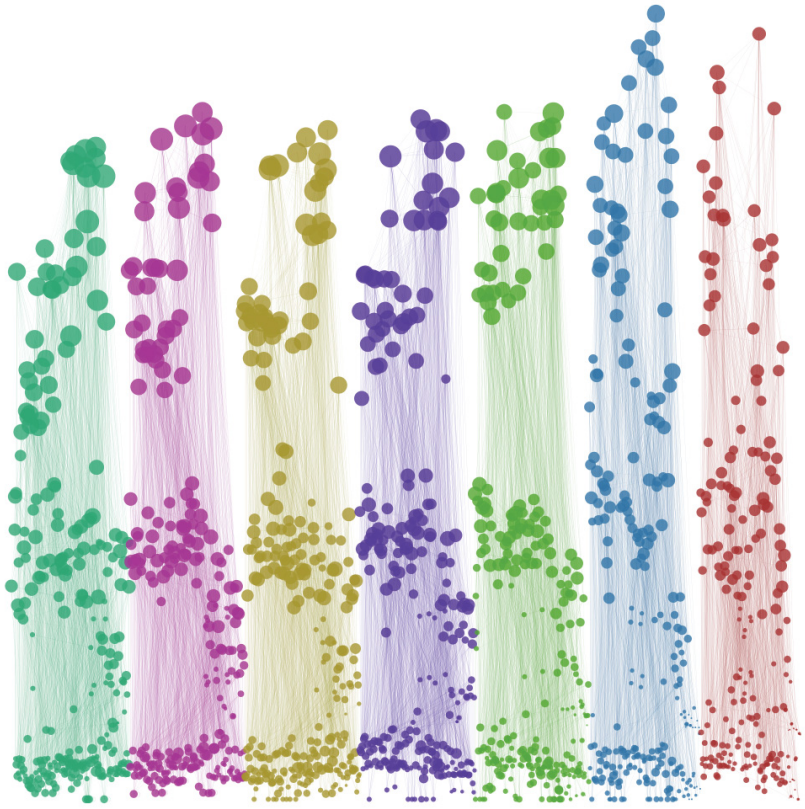


Figure 5.11: *The networks of methane corresponding to loadings 2, 3, 5, 6, 8, 10 and 12 (each shown in a different colour, and arranged in ascending order along the horizontal axis). The vertical displacement of nodes is proportional to their betweenness centrality value, while their size is proportional to the total degree.*

behaviour of the systems they represent, if we neglect the role of weights. However some clues point to the existence of such a relationship, and we believe that by developing more appropriate metrics for studying weighted digraphs representing stochastic matrices, the kind of analysis attempted in this last lines could unveil interesting trends and correlations between the networks and their associated physical systems.

So far we considered only properties of single nodes. In fig. 5.13 we report the average Erdos number for the networks of methane, argon and carbon dioxide, referred to one of the events of class 12. The Erdos number is a popular measure often used in scientific collaboration networks. It appeared first in the community of mathematicians, as the number of in-

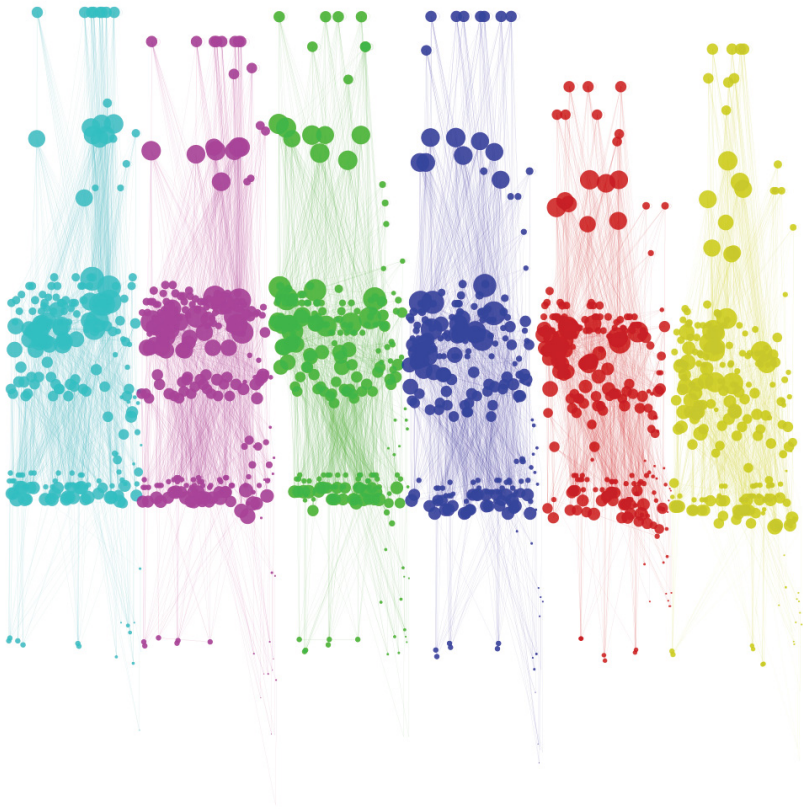


Figure 5.12: *The networks of carbon dioxide corresponding to loadings 2, 5, 6, 8, 10 and 12 (each shown in a different colour, and arranged in ascending order along the horizontal axis). The vertical displacement of nodes is proportional to their closeness centrality value, while their size is proportional to the total degree.*

termediate persons needed to link a scientist to the famous mathematician Paul Erdos, one of the pioneers of graph theory. He collaborated with an impressive number of colleagues, and it became customary to assign mathematicians their Erdos number, by looking at the coworkers of their published papers, and seeing how many steps it takes to reach Paul Erdos. Amazingly the average number for mathematicians is around 4.5. In our case it represents the average number of nodes between the inter-cage event and any other node. As can be seen the number slightly increases in all cases, but shows a minimum for CH_4 at $\langle n \rangle = 8$. This is a very rough measure but seems nonetheless to point to a particularly favourable condition for reaching inter-cage jumps at that loading, which is in line with the existence of a

maximum of self-diffusivity in its vicinity.

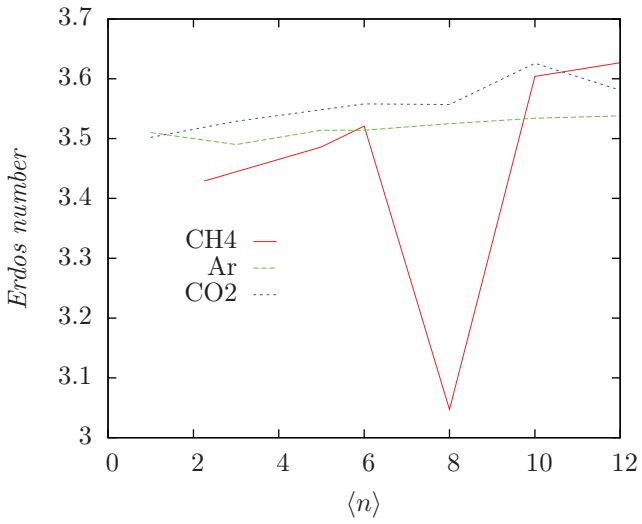


Figure 5.13: *Average Erdos number of nodes, relative to one of the nodes of class 12, for the networks of methane, argon and carbon dioxide, as a function of loading.*

In figs. 5.14, 5.15, 5.16 and 5.17 we report some global properties for networks of methane, argon and carbon dioxide. The total number of nodes (fig. 5.14) tends globally to decrease for all three systems, but while this tendency is weak and discontinuous for Ar and CO_2, in the case of CH_4 it is stronger, showing a plateau at low loading, followed by a drop for $\langle n \rangle \geqslant 8$. Similar observations apply to the network connection density (fig. 5.15), while the modularity (fig. 5.16) and the number of modularity classes (fig. 5.17) show strong increase for methane and again a weaker and discontinuous increase for the other two cases. This facts are all coherent with what we have already seen under many aspects: namely that as the density of molecules increases, the number of possible events dwindles and so does the probability of directly jumping from one to the other, causing the number of nodes and the density of connections in the network to decrease. As already remarked this effect is stronger for larger molecules, such as methane and xenon, and the effect should be related to the reduced available space for molecular motion. In the case of CO_2 we should also take into account the fact that this molecule experience a somehow larger α-cage space, because of its preferential adsorption in the channel regions connecting cages, which are largely unpopulated for other species. In this way part of the molecules are removed from the inner part of the cages where other species tend to crowd, and their volume interactions are thus reduced. The same lines

of reasoning applies to the increase of modularity and modularity classes, meaning that the dynamics is more and more divided in clusters of strongly connected events. This clusters emerge and strengthen as molecules get trapped in sites and jumps to other sites lead progressively more often to backscattering rather than to thermalization in the new destination.

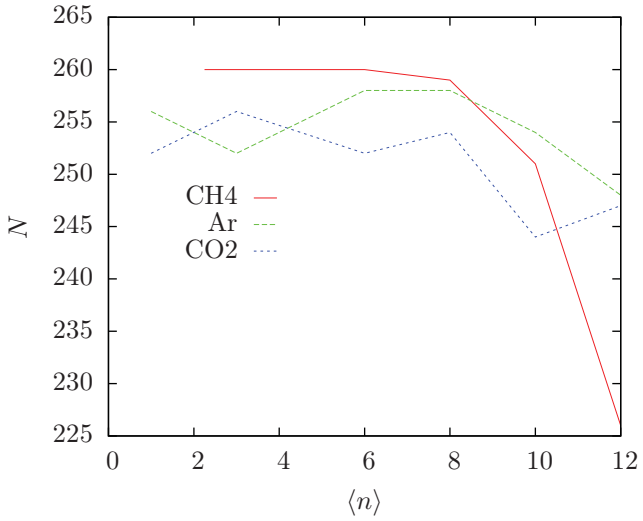


Figure 5.14: *The number of nodes in the networks of methane, argon and carbon dioxide, as a function of loading.*

At the end of this section we report the plots for in-degree, out-degree and in-strength for the networks of xenon, argon and carbon dioxide. We do not embark here on a detailed anlysis of these plot as we did for methane. The same lines of reasoning as before can in principle be applied, and we just remark a pair of general facts. The most evident aspect in the following figures is the close resemblance of those concerning Xe with what we have seen previously for CH_4. As already discussed in sec. 4.3 these two systems (together with kripton) show a similar general behaviour, with the only difference that all variations, associated with varying loading, for the Xe data are more pronounced and result in steeper curves than for CH_4. We explained this fact in terms of the larger size of Xe, and in the following we see that the same thing happens in all plots for Xe. On the other hand, both Ar and CO_2 show very different plots, with no pronounced trends and generally flatter curves. The analysis of these data would require a case by case study, and for the moment we leave it to the curiosity of the reader.

In the end of this provisional survey on the event network, we would like to suggest what the real interpretation of the various trends observed could be. At the beginning of this section we considered the possibility

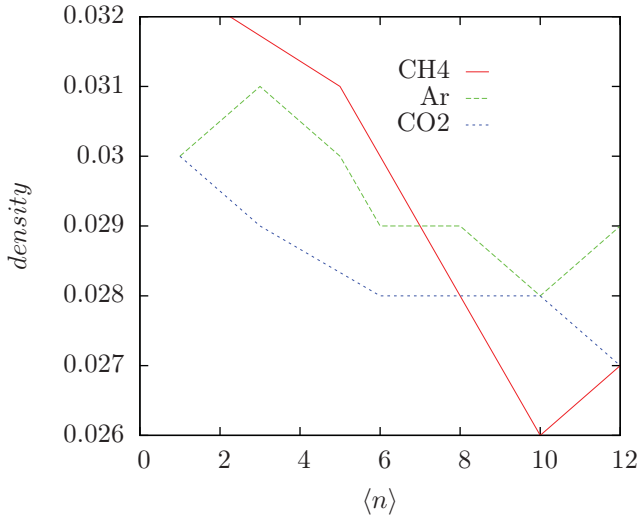


Figure 5.15: *The connection density of the networks of methane, argon and carbon dioxide, as a function of loading. Given the number of nodes, the connection density is defined as the ratio of the effective number of edges present, to the maximum possible number.*

that the evolution of the network architecture, under varying conditions, could reflect in some aspect the mechanisms determining the observed self-diffusivity trends. Then, given the available metrics and their limitations, we have seen how our network behaves with varying adsorbed species and conditions, and how some of its features can be related to the mechanisms underlying diffusion, and to the equilibrium probabilities of events. However no evident, global trend of the network seems to relate to the self-diffusivity trend in the case of varying loading. The only clear fact is that the network tends to become smaller and less connected as $\langle n \rangle$ increases. And while this makes sense, as we have seen, it seems to contrast with what we know to be the behaviour of self-diffusivity. The same is not true when we look at a given system under varying temperature, as in this case everything seems to be intuitively coherent.

The reason could be that, as the network structure represents the connectivity of events and their causal relationships, we should not expect after all to find in it a direct reflection of the diffusion overall behaviour. The problem may arise from the idea that as events connectivity decreases, so should sites connectivity. But while the two things are of course strongly related and can show a parallel behaviour, as in the case of varying temperature at low loading, they remain two separate things. So may be that the only fact that can be inferred from the connectivity of a network like

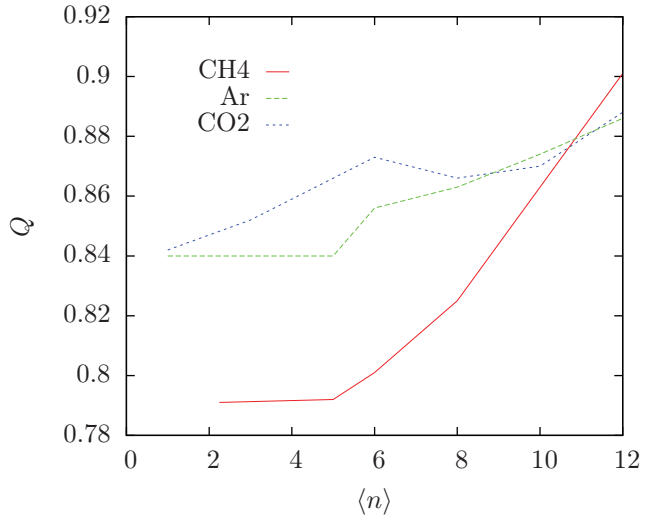


Figure 5.16: *The modularity value for the networks of methane, argon and carbon dioxide, as a function of loading.*

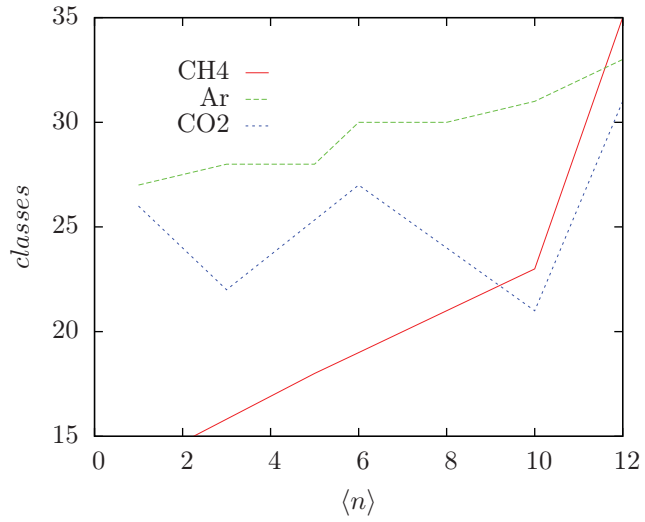


Figure 5.17: *The number of modularity classes in the networks of methane, argon and carbon dioxide, as a function of loading.*

ours is the rate of diffusion of a perturbation on it. If we consider a certain event realization at a given instant of time, the network structure tells us how long it will take for this perturbation to reach all of its part by random-walking on the graph. If the network is strongly interconnected and has low modularity, the information flow on it will be fast and uniform, leading to a rapid restoration of the equilibrium probability of all events. While if

modularity is high the time for losing memory of the initial event will be longer. This point of view seems to suggest that the structure of an event network like this is primarily related to the correlation properties of the underlying dynamics. And in this light it is evident that what we should see in the progressively sparser and less homogeneous tangles in fig. 5.1, is the increase of correlation effects in the dynamics of methane molecules as their density grows. And indeed this is what happens, tempering other effects, which would otherwise cause a still stronger increase of diffusivity.

On the other hand we remark that the network we described is in some sense hybrid. A pure event network can be imagined as one built on events defined only by displacement, while, on the contrary, our events bring also information concerning the sites involved. Let us consider a simple squared lattice of regularly spaced points. The events on this symmetric, homogeneous lattice can be defined by only considering displacements, and the connectivity of the resulting event network would be related only to the nature and strength of displacements correlation. The same can be said of a system with no fixed underlying framework in space, like a molecule in a bulk liquid. But given the nature of the zeolite framework, and the inhomogeneity of sites, we must make a difference between events corresponding to the same displacement, but connecting different pairs of sites. This makes things slightly more complicated and while it is a legitimate way of defining events, it leads to a network where different aspects of the dynamics are mixed together. However, once this is made clear, an analysis like that exposed in the previous paragraphs can proceed and could demonstrate to give interesting insights in the mechanisms underlying the diffusive process.

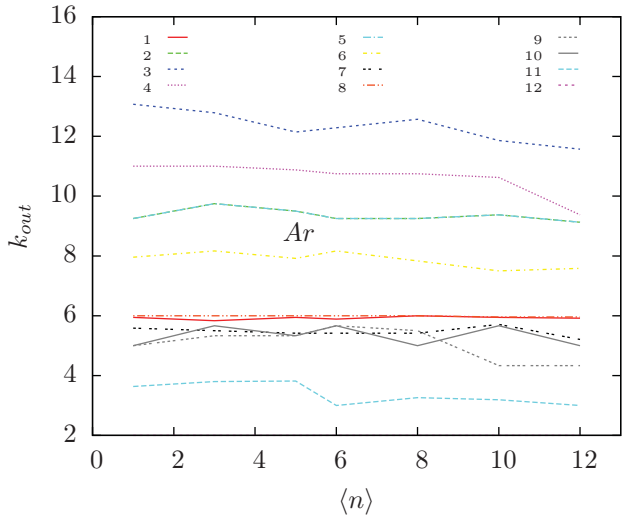


Figure 5.18: *Average out-degree for all event classes of argon, plotted against loading.*

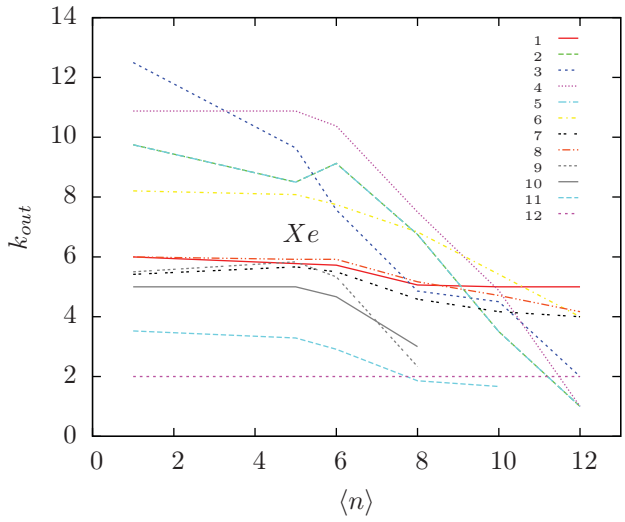


Figure 5.19: *Average out-degree for all event classes of xenon, plotted against loading.*

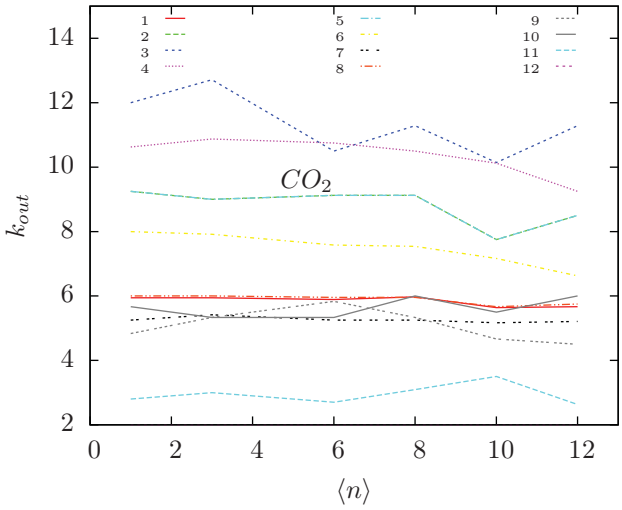


Figure 5.20: *Average out-degree for all event classes of carbon dioxide, plotted against loading.*

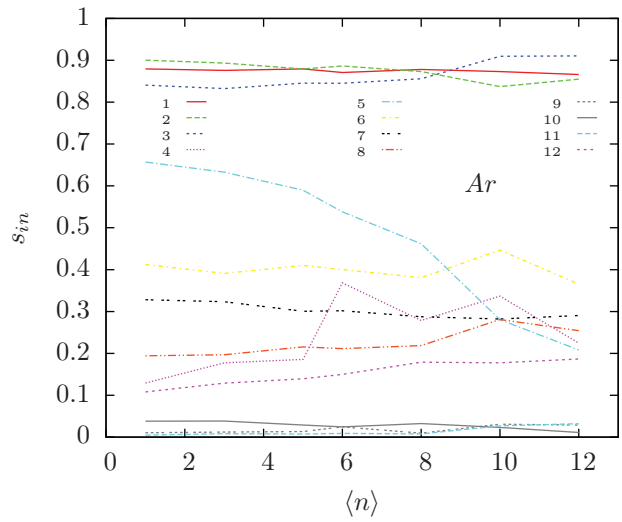


Figure 5.21: *Average in-strenth for all event classes of argon, plotted against loading.*

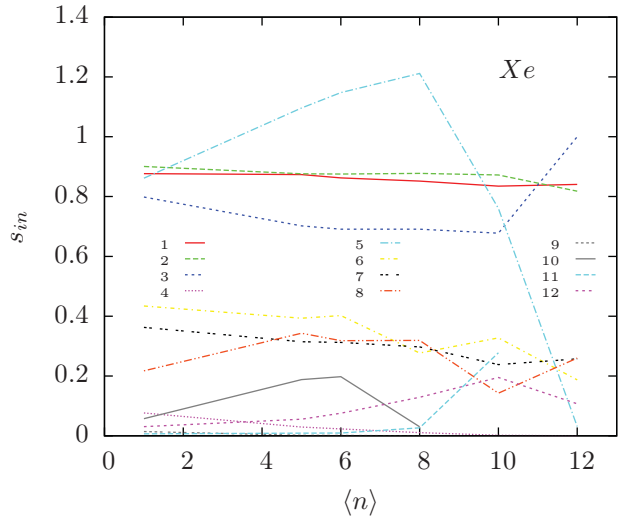


Figure 5.22: *Average in-strength for all event classes of xenon, plotted against loading.*

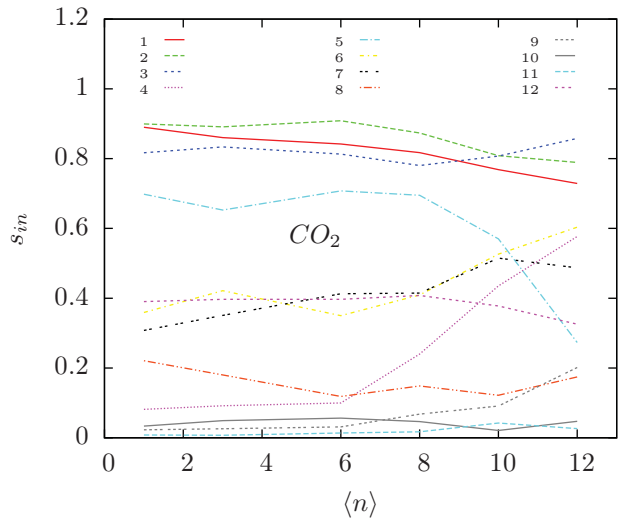


Figure 5.23: *Average in-strength for all event classes of carbon dioxide, plotted against loading.*

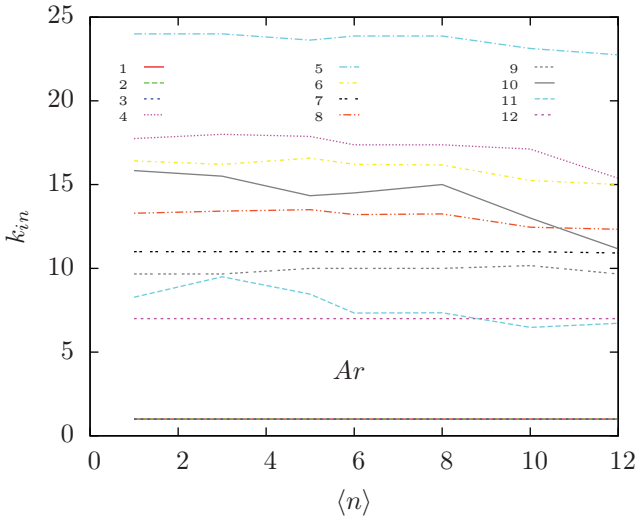


Figure 5.24: *Average in-degree for all event classes of argon, plotted against loading.*

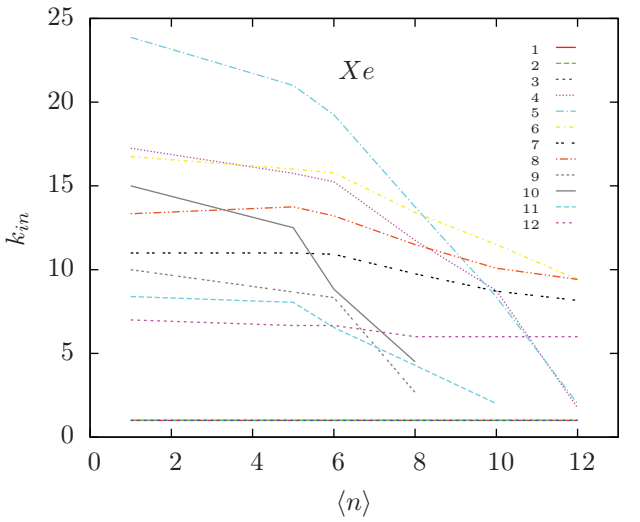


Figure 5.25: *Average in-degree for all event classes of xenon, plotted against loading.*

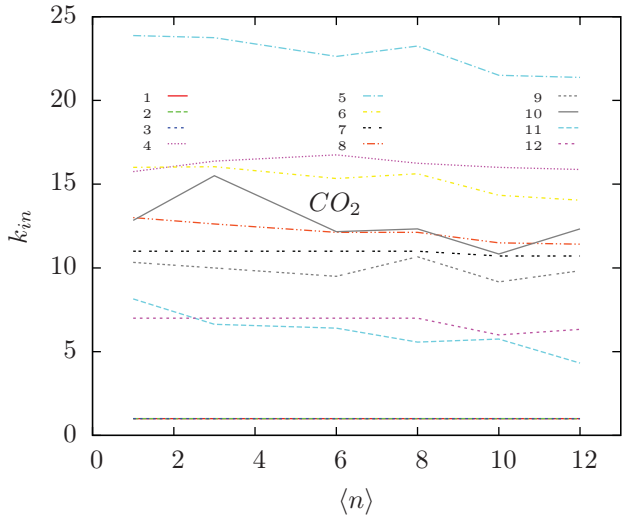


Figure 5.26: *Average in-degree for all event classes of carbon dioxide, plotted against loading.*

Chapter 6

Conclusions and Future Perspectives

Many different approaches to the study and simulation of adsorbed molecules in zeolites were described in these pages, on the general unifying background of energy landscape theory. Some conclusions on the models proposed have been already drawn at the end of the relative sections. What we present in the following is a brief outline of the main common themes, pointing at further development of these ideas.

Cellular Automata seem to be valuable candidates for simulating large systems, having their strength in their naturally parallel evolution rules. This allows a more efficient utilization of computational resources, and can act as a general framework, where different concepts of computational chemistry can find expression. As an example we implemented parts of various Monte Carlo algorithm in our probabilistic CA models. As already remarked, a fundamental problem in this field is that of linking the model's parameters to the physical reality it aims to represent. This is one of the reasons why we tried to analyze the MD dynamics, taken here as 'reality', with the tools described. Obviously a good knowledge of the energy landscape of the system could give useful information in this sense. However it seems that this approach is really effective for systems showing high barriers. In the rather flat energy landscape for methane, argon and xenon, uncovered by our study, little can be added to what Molecular Dynamics already offers. This is because small barriers make the kinetic difficult to be treated with Transition State Theory, and moreover, in this conditions, the shape and extension of basins play the dominant role. Unfortunately these information can only be known approximately in this theoretical framework, using for instance the eigenvalues of the Hessian matrix at a minimum, to estimate the extension of the corresponding basin. But while in the presence of high barriers, and large energy gaps between minima, so that the enthalpy is dominant in determining the free energy, this approximation of the entropic

effects is fine, in our case entropy plays a leading role and accounting for it properly could be more difficult. Nonetheless the exploration of the energy landscape, using tools like OPTIM and PATHSAMPLE, may be fruitful in different microporous systems, or in the case of strongly interacting adsorbed species. If we are interested to the motion of zeolite extra-framework cations for instance, or in the case of a molecule like benzene, which is known to move very slowly in faujasite type zeolites, enthalpy should dominate the dynamics, and given that MD is ineffective in such systems, a mapping of the energy landscape could in principle be a very useful approach for the study of kinetics, and the formulation of coarse-grained models, including a sensible definition of parameters for our Cellular Automaton model.

The effort to mould the CA from a realistic casting led to the other two interrelated approaches reported.

Markov State Models are valuable tools for describing and simulating processes which are hardly handled by means of simple MD. However, they appear to be again really promising for systems with higher kinetic barriers compared to those addressed here. This is because they require a clear separation in the characteristic time scales of dynamics in order to be really effective, i.e. markovian.

If the systems studied were not best suited for this way of modeling, this was the opportunity for an interesting detour, as it led to the reformulation of the problem in terms of an event-event transition matrix. In the way it is formulated here, this is little more than a second-order transition matrix, but by switching to consider it as a first order transition matrix acting in the event space, we gained an interesting and new point of view, which was then translated in the language of network theory.

Beside being the basis for the coarse-grained model described in sec. 4.2, it allowed us to qualitatively account for the diffusivity trend observed for argon, methane, xenon and kripton, by looking at the equilibrium probability distribution of events.

With the word *equilibrium*, referred to the event space, we mean the situation in which the system is at thermodynamic equilibrium, and we have no knowledge of its previous history. In other words, the equilibrium distribution of events gives us the probability of observing an event at any randomly picked time, at thermodynamic equilibrium. But is it enough to know this distribution in order to predict the behaviour of the system, or to explain it? If we know that, at a particular time, a molecule has gone through a given event, then for a certain time we know that the probability of performing any other move is related to the first one. The way by which the molecule loses memory of its starting point in the event space is related to the displacement autocorrelation function, and this information is embedded in the structure of the event network defined in sec. 5.2. Diffusion can be seen as a random walk on the network of events, a byproduct of which is the random walk in configuration space. And it is intuitively clear that

the way events are connected should be considered beside their equilibrium distribution. In fact we implicitly took this into account in sec. 4.2.

Looking at the event network architecture is one way by which we should be able to infer what the various contributions to the displacement autocorrelation function are. This could also be done by standard algebraic analysis of the event transition matrix. However, if we look at its eigenspectrum, we find that in general not all eigenvalues are real, and this renders this way to proceed less intuitive and appealing.

Nonetheless, whichever way we look at the problem, it is useful to think in terms of the spectral analysis of this matrix, as, similarly to what we have seen for the case of configuration space in sections 3.3 and 4.1, its eigenvectors and eigenvalues must be related to the structure of the processes by which equilibrium is reached, and their characteristic time scales, respectively. But, given that we are dealing with the event space, the processes and time scales are those by which autocorrelation effects die out after a given event.

Apart from this considerations, knowing the probability distribution of events, and and their connectivity, could suggest an original way for determining the parameters for a CA model, like the one described in section sec. 2.2. One could think of a fitting procedure based on the information embedded in the network, which optimizes the set of parameters of the model until the latter generates a dynamics with the same fundamental pattern of connections. This optimization problem can be cast in various way, and we attempted a Neural Network (NN) approach to it. This objects are capable of fitting very complex functions, and are particularly useful when even a tentative formulation of the analytical functional form of a problem is impossible. Let us consider the set of unknown parameters of a model, already properly designed for the problem at hand, and the set of weights of the event transition network (suitably simplified by means of symmetry considerations): there should be a way of adjusting the parameters so that they cause the model to produce a dynamics whose event transition network fits that obtained from MD.

A suitably trained NN should be able, in principle, to take as input a set of network weights, and output the set of parameters which, implemented in the CA, generate an event transition network with the same set of weights. The training procedure consists of feeding a large number of input and output set pairs to the NN, so that it 'learns' how to output the proper set given an input one. We used sets of weights of the event network, obtained from CA runs, as input, and tried to train the NN to output the correct set of parameters used for the simulations. After a successful training, one can expect that, giving the set of weights of a network, this time obtained from a Molecular Dynamics run, the Neural Network output the set of right parameters, causing the CA to properly emulate the system. Unfortunately we were not able to succesfully train the NN, and so this attempt was left

hanging, and is briefly reported here just as a suggestion for further studies.

Apart from applying the energy landscape analysis, and techniques based on the clustering of MD trajectories and PCCA, to other microporous structures and adsorbate species, which seems a straightforward task, many aspects of the idea exposed require further investigation.

Concerning the exploration of energy landscapes, we have seen how, by iteratively searching minima and transition states, a representative part of the PES can be uncovered. Beside the Discrete Path Sampling method succesfully proposed and implemented by Wales and coworkers, other approaches are reported in literature [103, 104]. In [104], as an example, a transition network is built on the energy landscape. This is done by homogeneously distributing points in the relevant part of configuration space, minimizing their energy before adding more points by interpolation, and, when a certain number of minima is reached, connecting them by searching minimum energy paths. This gives a weighted transition network, whose weights are determined by means of Transition State Theory. This and other techniques rely on algorithms for minimizing energy along paths connecting minima. However it is well known that at finite temperature the minimum energy path between two points in configuration space may lose relevance. This is because other paths starts to be available as temperature grows. The statistical weight difference among these paths tends to decrease as the kinetic energy of the system increases and, depending on the system, a huge number of practically degenerate different routes can render the knowledge a single minimum energy one irrelevant. Moreover, as temperature increases a relatively high energy path can take over the leading role, as the entropy contribution can render its free energy particularly low. Imagine as an example a low energy path on the bottom of a narrow valley of the energy landscape; at 0K it will be dominant, but at finite temperature, another higher energy path can overwhelm it, if the latter lies on the bottom of a wider valley, with a larger catchment basin. Discrete Path Sampling and similar techniques should not suffer much from the first problem, as they effectively sample a number of paths, properly accounting for degeneracy. The second problem is more subtle and was independently addressed by Elber *et al* [105, 106] and Faccioli *et al* [107–110], by proposing the minimization of an *action* functional of the path, in the spirit of Feynman's Path Integral formulation of quantum mechanics.

Various formulations of this stochastic action (also called Onsager-Machlup action) exist. One of them contains a Hessian term, that is an energy second derivative term, which account for the curvature of the PES along a path, and thus allows an estimation of the aforementioned entropic contribution.

To date, techniques based on stochastic actions have been applied to the study of the single dominant reaction path (DRP [111]) and to the sampling of paths in its immediate neighborhood [112]. It would be interesting to see the impact of applying an action minimization algorithm, taking into

account entropic effects, to a global energy landscape exploration, in the spirit of the other algorithms seen, which are all based on minimum energy paths networks. However much of the applicability of this idea depends on the processor time required by action minimization, which can be expected to be more computationally expensive than energy minimization.

Finally, concerning the event space approach to the dynamics of diffusing particles, we can only conclude with a series of open questions. First of all, is it a useful way of formulating the problem? Can it really add to our understanding, or is it little more than a complicated reformulation of it? Is the difference a matter of concepts or just of words? And for what attains the most suitable tools for addressing it, is there a way of simmetrizing the matrix in order to analyze it in terms of real eigenvalues and eigenvectors? Is there another suitable analytical way? What is the advantage of looking at the corresponding graph? What can be the physical interpretation of the metrics characterizing the network? If the modularity is inversely related to the forgetfulness of dynamics, is there a way of quantifying this? Is there a way of reformulating centrality measures, so that they are meaningful in the context of a stochastic matrix network? How could the problem of degenerate paths be addressed in the definition of distance and related metrics? What is the most useful definition of the clustering coefficient in this context? How is it related to dynamics? Is there any critical phenomenon that can be highlighted in the evolution of the network with varying conditions? What correlations among various metrics are meaningful, and in which sense? Is there a more elegant, useful or meaningful way of defining such a network? Could it act in any way as a conceptual bridge between the physical reality of diffusion and a generic coarse-grained model of it? And if it make sense, is it suitable only for the kind of systems described here? By a proper definition of the relevant events, could it be sensible to extend this idea to other processes, such as diffusion in bulk solutions, where no fixed framework exist? Could it shed any new light on processes that have already been studied from the viewpoint of configuration space, such as protein folding and other complex structural transitions?

Only time, reflection and hard work will tell.

Bibliography

[1] P. Demontis and G. B. Suffritti, Chem. Rev. **97**, 2845 (1997),

[2] J. Kärger and D. M. Ruthven, *Diffusion in Zeolites and Other Microporous Materials* (John Wiley and Sons: New York, 1992)

[3] B. Smit and T. L. M. Maesen, Chem. Rev. **108**, 4125 (2008)

[4] S. M. Auerbach, Int. Rev. Phys. Chem. **19**, 155 (2000)

[5] M. Schenk, S. L. Vidal, T. J. H. Vlugt, B. Smit and R. Krishna, Langmuir **17**, 1558 (2001)

[6] C. Saravanan and S. M. Auerbach, J. Chem. Phys. **107**, 8132 (1997)

[7] A. Gabrieli, F. G. Pazzona, P. Demontis, and G. B. Suffritti, Phys. Rev. E, **83**, 056705 (2011)

[8] C. Tunca and D. M. Ford, Chem. Eng Sci. **58**, 3373 (2003)

[9] F. G. Pazzona, *A cellular automata model for diffusion and adsorption in zeolites: Construction of a mesoscopic model*, LAP Lambert Academic Publishing, 66123 Saarbrücken, Germany, 2010.

[10] http://www.iza-structure.org

[11] S. Wolfram, *A new kind of science* (Wolfram Media, 2002)

[12] B. Chopard and M. Droz, *Cellular Automata Modeling of Physical System* (Cambridge University Press, 1998)

[13] A. M. Pintus, F. G. Pazzona, P. Demontis and G. B. Suffritti, J. Chem. Phys. **135**, 124110 (2011)

[14] P. Demontis, F. G. Pazzona, and G. B. Suffritti, *J. Chem. Phys.* **126**, 194709 (2007).

[15] P. Demontis, F. G. Pazzona, and G. B. Suffritti, *J. Chem. Phys.* **126**, 194710 (2007).

[16] F. G. Pazzona, P. Demontis, and G. B. Suffritti, *J. Chem. Phys.* **131**, 234703 (2009).

[17] P. Demontis, F. G. Pazzona, and G. B. Suffritti, *J. Phys. Chem. B* **112**, 12444 (2008).

[18] F. G. Pazzona, P. Demontis, and G. B. Suffritti, *J. Chem. Phys.* **131**, 234704 (2009).

[19] T. Toffoli and N. Margolus, Physica D, **45**, 229 (1990)

[20] T. Toffoli and N. Margolus, *Cellular Automata Machines: A New Environment for Modeling* (MIT Press, 1997)

[21] B. Smit and R. Krishna, Curr. Opin. Solid St. M. **5**, 455 (2001)

[22] D. Paschek and R. Krishna, Langmuir, **17**, 247 (2001)

[23] K. A. Fichthorn and W. H. Weinberg, J. Chem. Phys. **95**, 1090 (1991)

[24] B. D. Lubachevsky, J. Comput. Phys. **75**, 103 (1988)

[25] A. Chatterjee and D. G. Vlachos, J. Computer-Aided Mater. **14**, 253 (2007)

[26] S. Wolfram, Rev. Mod. Phys. **55**, 601 (1983)

[27] A. G. Hoekstra, J. Falcone, A. Caiazzo, and B. Chopard, Lect. Notes Comput. Sc. **5191**, 192 (2008)

[28] P. Santikary, S. Yashonath, and G. Ananthakrishna, J. Phys. Chem. **96**, 10469 (1992)

[29] V. I. Manousiouthakis and M. W. Deem, J. Chem. Phys. **110**, 2753 (1998)

[30] D. Frenkel and B. Smit, *Understanding Molecular Simulation. 2nd Edition* (Academic Press, 2002)

[31] R. Krishna and J. M. van Baten, Sep. Purif. Technol. **61**, 414 (2008)

[32] R. Krishna and J. M. van Baten, Chem. Engng. J. **133**, 121 (2007)

[33] R. Krishna and J. M. van Baten, Chem. Engng. Sci. **64**, 3159 (2009)

[34] R. Krishna and J. M. van Baten, J. Phys. Chem. B, **109**, 6386 (2005)

[35] D. Paschek and R. Krishna, Langmuir, **133**, 121 (2001)

[36] R. Krishna, Ind. Eng. Chem. Fundam. **16**, 228 (1977)

[37] S. Yashonath, J. Phys. Chem. **95**, 5877 (1991)

[38] P. R, Van Tassel, H. T. Davis, and A. V. McCormick, Langmuir, **10**, 1257 (1994)

[39] J. P. Boon, D. Dab, R. Kapral, and A. T. Lawniczak, Phys. Rep. **273**, 55 (1996)

[40] F. G. Pazzona, A. Gabrieli, A. M. Pintus, P. Demontis and G. B. Suffritti, J. Chem. Phys, **134**, 184109 (2011)

[41] J. Klafter and J. M. Drake eds., *Molecular Dynamics in Restricted Geometries*, John Wiley and Sons, New York, first edition, 1989. 5

[42] M.-O. Coppens, A. T. Bell, and A. K. Chakraborty, *Chem. Eng. Sci.* **53**, 2053 (1998).

[43] C. Saravanan, F. Jousse, and S. M. Auerbach, *Phys. Rev. Lett.* **80**, 5754 (1998).

[44] Z. Chvoj, H. Conrad, V. Cháb, M. Ondrejcek, and A. M. Bradshaw, *Surf. Sci.* **329**, 121 (1995).

[45] J. van den Bergh, S. Ban, T. J. H. Vlugt, and F. Kapteijn, *J. Phys. Chem. C* **113**, 17840 (2009).

[46] D. Dubbeldam, E. Beerdsen, T. J. H. Vlugt, and B. Smit, *J. Chem. Phys.* **122**, 224712 (2005).

[47] D. Frenkel and M. H. Ernst, *Phys. Rev. Lett.* **63**, 2165 (1989).

[48] D. F. M. A. van der Hoef, *Phys. Rev. A* **41**, 4277 (1990).

[49] D. F. M. A. van der Hoef, *Physica D* **47**, 191 (1991).

[50] R. Gomer, *Rep. Prog. Phys.* **53**, 917 (1990).

[51] P. Demontis, L. Fenu, and G. B. Suffritti, *J. Phys. Chem. B* **109**, 18081 (2005).

[52] D. J. Wales, *Energy Landscapes. With Application to Clusters, Biomolecules and Glasses* (Cambridge University Press, 2003)

[53] D. J. Wales, Curr. Opin. Struc. Biol. **20**, 3 (2010)

[54] P. Hänggi, P. Talkner and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990)

[55] B. Strodel and D. J. Wales, Chem. Phys. Lett. **466**, 105 (2008)

[56] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2000)

[57] T. V. Bogdan, D. J. Wales and F. Calvo, J. Chem. Phys. **124**, 44102 (2006)

[58] F. H. Stillinger and T. A. Weber, J. Phys. Chem. **87**, 2833 (1983)

[59] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A, **101**, 5111 (1997)

[60] C. J. Cerjan and W. H. Miller, J. Chem. Phys. **75**, 2800 (1981)

[61] D. J. Wales, Mol. Phys. **100**, 3285 (2002)

[62] D. J. Wales, Mol. Phys. **102**, 891 (2004)

[63] S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **120**, 2082 (2004)

[64] G. Henkelman, B. P. Uberuaga and H. Jonsson, J. Chem. Phys. **113**, 9901 (2000)

[65] E. W. Dijkstra, Num. Mat. **1**, 269 (1959)

[66] http://www-wales.ch.cam.ac.uk/

[67] J. P. K. Doye, M. A. Miller, and D. J. Wales, J. Chem. Phys. **111**, 8417 (1999)

[68] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A, **101**, 5111 (1997)

[69] J. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schutte and F. Noé, J. Chem. Phys. **126**, 155101 (2007)

[70] F. Noé and S. Fischer, Curr. Opin. Struc. Biol. **18**, 154 (2008)

[71] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill and W. C. Swope, J. Chem. Phys. **134**, 174105 (2011)

[72] P. Deuflhard, W. Huisinga, A. Fischer and C. Schutte, Linear Algebra Appl. **315**, 39 (2000)

[73] O. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997)

[74] D. J. Wales and T. V. Bogdan, J. Phys. Chem. B, **110**, 20765 (2006)

[75] A. Corma, F. Rey, J. Rius, M. J. Sabater and S. Valencia, Nature, *431*, 287 (2004)

[76] C. Baerlocher, L. B. McCusker and D. H. Olson, *Atlas of Zeolite Framework Types, 6-th edition* (Elsevier, 2007)

[77] S. Kube and M. Weber, J. Chem. Phys. **126**, 24103 (2007)

[78] P. Deuflhard and M. Weber, Linear Algebra Appl. **398**, 161 (2005)

[79] S. Plimpton, J. Comp. Phys. **117**, 1 (1995)

[80] http://lammps.sandia.gov

[81] R. Krishna and J. M. van Baten, Micropor. Mesopor. Mat. **109**, 91 (2008)

[82] M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, C. Schutte and F. Noé, J. Chem. Theory Comput. **8**, 2223 (2012)

[83] C. Tunca and D. M. Ford, J. Chem. Phys. **111**, 2751 (1999)

[84] J. Caro and M. Noack, Micropor. Mesopor. Mat. **115**, 215 (2008)

[85] A. L. Barabasi, *Linked. The new Science of Networks* (Perseus Books Group, 2003)

[86] G. Mack, Commun. Mat. Phys. **219**, 141 (2001)

[87] A. Barrat, M. Barthelemy, R. Pastot-Satorras and A. Vespignani, PNAS **101**, 3747 (2004)

[88] D. J. Watts and S. Strogatz, Nature **393**, 6684 (1998)

[89] B. Bollobás, *Random Graphs (2nd edition)* (Cambridge University Press: Cambridge, 2001)

[90] J. P. K. Doye and C. P. Massen, J. Chem. Phys. **122**, 84105 (2005)

[91] C. P. Massen, J. P. K. Doye and R. W. Nash, Physica A, **382**, 683 (2007)

[92] F. Rao, S. Garrett-Roe and P. Hamm, J. Phys. Chem. B, **114**, 15598 (2010)

[93] F. Rao and A. Caflisch, J. Mol. Biol. **342**, 299 (2004)

[94] D. Gfeller, D. Morton, P. De Los Rios, G. Caldarelli and F. Rao, Phys. Rev. E, **76**, 26113 (2007)

[95] D. Gfeller, P. De Los Rios, A. Caflisch and F. Rao, PNAS, **104**, 1817 (2007)

[96] D. Prada-Gracia, J. Gomez-Gardeñes, P. Echenique and F. Falo, Plos Comput. Biol. **5**, 1 (2009)

[97] D. Gfeller and P. De Los Rios, Phys. Rev. Lett. **99**, 38701 (2007)

[98] https://gephi.org

[99] T. Opsahl, *Structure and Evolution of Weighted Networks.* University of London (Queen Mary College), pp. 104-122 (2009)

[100] http://www.r-project.org/

[101] T. Opsahl, F. Agneessens and J. Skvoretz, Soc. Networks, **32**, 245 (2010)

[102] http://www.visualcomplexity.com/vc/

[103] F. Noe, M. Oswald, G. Reinelt, S. Fischer and J. C. Smith, Multiscale Model. Simul. **5**, 393 (2006)

[104] F. Noe, D. Krachtus, J. C. Smith and S. Fischer, J. Chem. Theory Comput. **2**, 840 (2006)

[105] R. Elber and D. Shalloway, J. Chem. Phys. **112**, 5539 (2000)

[106] R. Elber, A. Ghosh and A. Cardenas, Acc. Chem. Res. **35**, 396 (2002)

[107] P. Faccioli, M. Sega, F. Pederiva and H. Orland, Phys. Rev. Lett. **97**, 108101 (2006)

[108] M. Sega, P. Faccioli, F. Pederiva, G. Garberoglio and H. Orland, Phys. Rev. Lett. **99**, 118102 (2007)

[109] P. Faccioli, J. Phys. Chem. B, **112**, 13756 (2008)

[110] P. Faccioli, J. Chem. Phys. **133**, 164106 (2010)

[111] E. Autieri, P. Faccioli, M. Sega, F. Pederiva and H. Orland, J. Chem. Phys. **130**, 64106 (2009)

[112] G. Mazzola, S. Beccara, P. Faccioli and H. Orland, J. Chem. Phys. **134**, 164109 (2011)