Topology-based Visualization of Transformation Pathways in Complex Chemical Systems

K. Beketayev [1,3] G.H. Weber [1,3] M. Haranczyk [1] P.-T. Bremer [2] M. Hlawitschka [3] and B. Hamann [3]

[1]Computational Research Division, Lawrence Berkeley National Laboratory
[2]Center for Applied Scientific Computing, Lawrence Livermore National Laboratory
[3]Institute for Data Analysis and Visualization, Department of Computer Science, University of California, Davis
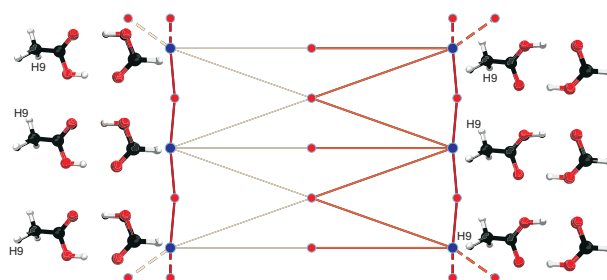
Figure 1: A 2D graph showing the topology of the 3D potential energy function of a complex of formic and acetic acids, which depends on the positions of constituting atoms. Blue and red dots represent minimum energy configurations and lowest barriers connecting neighboring minima, respectively. Edges represent the energy cost of a particular transformation, with darker and wider edges corresponding to transformations through lower barriers (more likely transformations). Two vertical branches corresponding to different positions of protons are visible on the left and right side of the graph. Energy barriers for transforming the right branch into left one are lower than for the reverse transformation.

Abstract

Studying transformation in a chemical system by considering its energy as a function of coordinates *of the system's components provides insight and changes our understanding of this process. Currently, a lack of effective visualization techniques for high-dimensional energy functions limits chemists to plot energy with respect to one or two coordinates at a time. In some complex systems, developing a comprehensive understanding requires new visualization techniques that show relationships between* all *coordinates at the same time. We propose a new visualization technique that combines concepts from topological analysis, multi-dimensional scaling, and graph layout to enable the analysis of energy functions for a wide range of molecular structures. We demonstrate our technique by studying the energy function of a dimer of formic and acetic acids and a LTA zeolite structure, in which we consider diffusion of methane.*

Categories and Subject Descriptors (according to ACM CCS): J.2.4 [Computer Applications]: Physical Sciences and Engineering—Chemistry I.3.8 [Computing Methodologies]: Computer Graphics—Applications

1. Introduction

In chemistry, transformation processes that involve changes of relative positions of atoms in chemical systems are of fundamental interest. Examples of such transformations include internal rotation of fragments of a molecule (e.g., conformation change), translation of atoms or molecules within a chemical system (e.g., diffusion), and shifting of atoms leading to breaking and/or re-arrangement of chemical bonds (e.g., chemical reactions). An important factor of any transformation is its cost. Chemists usually focus on transformations between configurations of chemical systems corresponding to *energy minima*, representing stable states of the

system. The term "configuration" is generalized here and can bear two meanings: It can refer to a set of strictly defined positions of all atoms of a studied system as it is done in the vast majority of quantum chemical calculations relying on the Born-Oppenheimer approximation. On the other hand, it can also refer to a representative or dominantly populated state within a class.

The cost of a transformation between stable states can be defined as the energy difference between the two minima involved in the transformation. In a typical case, a transformation involves a transition through a higher energy configuration, a *barrier* that determines the probability of the transformation (or time required for it to happen). The analysis of *transformation pathways* in a chemical system usually follows the same general scheme. After identifying important energy minima, connecting pathways and corresponding barriers are found. A chemical system has $3n$ degrees of freedom, where n is the number of atoms. Analyzing transformations in such a system requires identifying minima and transition states in a $3n$-dimensional space. Chemists can often reduce the dimensionality of this space by exploiting prior knowledge about the system. For example, when considering a diffusion process, it is often sufficient to consider only translation and rotation of a rigid molecule, and the analysis can be performed for a six-dimensional energy function. When considering only conformational changes of a molecule, it is sufficient to investigate $3n - 6$ internal degrees of freedom. The latter can be simplified further by assuming that some coordinates are constant due to very high energy cost associated with their change.

Once minima and transition states are identified in the original or simplified high-dimensional energy function, it becomes necessary to visualize and analyze the relation between minima and their connecting paths. Due to the system's complexity and a lack of effective means for visualizing relationships in high-dimensional space, chemists usually select only one or two most important coordinates in the transforming chemical system (e.g., a reaction coordinate) and show the corresponding two- or three-dimensional energy plots. For example, in case of the dimer of formic and acetic acid, shown in the Figure 2a, a chemist would select the rotation angle of one methyl group as target coordinate and measure system energy as a function of it, see Figure 2b. An analysis of the energy function in this case is simple, and such one- and two-dimensional plots are commonly seen in discussions and publications demonstrating their usefulness. Similar attempts to plot higher-dimensional energy data are not common, although complex transformations involving more than two coordinates arise in many studied systems. We aim to provide a convenient approach to visualize the energy cost and barriers of possible transformation pathways in chemical systems.

We have developed our method based on the observation that minima of the system as well as their relationships to
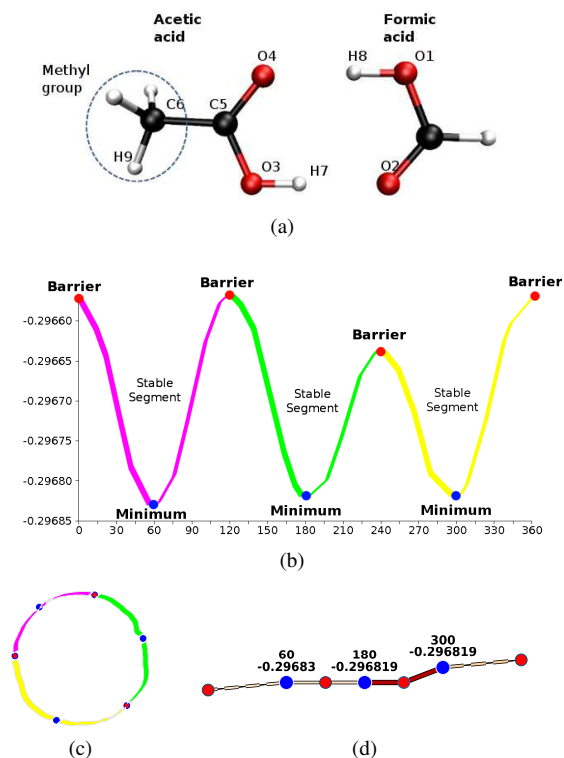

Figure 2: (a) Chemical system configuration of a dimer of formic and acetic acid. (b) Energy of the system in Hartrees (y-axis) as a function of the methyl group rotation angle between 0 to 360 degrees (x-axis). Blue and red dots mark energy minima and barriers, respectively. (c) Naive graph representation of transformation pathways. (d) Proposed graph representation, where the top row of a energy minima label displays its coordinate and the bottom row its energy value.

each other are of primary interest. A graph-like structure describing this relationship is the Morse complex (cf. Section 3.1), and descending regions in the Morse complex correspond to classes of configurations associated with a stable state. The saddle between two adjoining regions, i.e., the lowest function value along their boundaries, corresponds to the barrier between the states, and transformation paths are given by the edges connecting two minima through a saddle. Figure 2b shows the segmentation of our example energy function for the dimer of formic and acetic acid system. Three segments correspond to each of the three minima (blue dots) of the system and three barriers (red dots) separate each pair of neighboring minima. Figure 2c shows one way of drawing the resulting graph representing transformational pathways, where each edge corresponds to a segment that is colored differently. Based on this relationship between stable states and barriers, we construct a graph that represents the system. To preserve relational context between stable states for higher-dimensional cases, we utilize multidimensional scaling to project the Morse complex graph onto the plane and incorporate the resulting location information into the

graph layout. We also embed chemical information, such as the likelihood of a transformation.

Figure 2d shows the graph for our 1D energy example of the dimer of formic and acetic acid system (Figure 2a). We note that the barrier (red dot) between the violet and yellow segment is split into two barriers, each connected to respective minima by dashed lines, signifying the periodic nature of the barrier. Even though we use a 1D example to illustrate the concepts behind our idea, the true power of using a combination of Morse complex and multidimensional scaling to project this information onto a plane lies in the fact that we can apply it to higher-dimensional systems. In Section 4 we provide several 2D, 3D, and 4D examples that illustrate how our graphs can be used to gain insight into complex chemical systems. Our main contributions are:

- Combining the structural information contained in the Morse complex with geometric information obtained by multi-dimensional scaling in a novel way to facilitate analysis by chemists.
- Incorporating geometric information and properties of the chemical system, such as the likelihood of a transformation, into a graph layout with constrained node placement.
- Handling of periodic conditions in chemical systems and highlighting corresponding transitions in the resulting graph.

2. Background and Related Work

2.1. Energy Representation of a Chemical System

We assume that a numerical grid representation of a (multidimensional) energy function is available. The considered dimensions correspond to changes of atomic coordinates leading to a desired sampling of the configuration space of the studied system. The exact definitions of energy and configuration of a chemical system may vary from case to case; the considered energy may be the potential energy (e.g., including electrostatic interactions between particles in the system) or free energy, which also includes thermal and entropic contributions. The configuration of the system may be defined by precise positions of nuclei (or atoms and larger fragments) or average positions (average over time, states in ensemble or fragment of space) [KGK*96].

2.2. Scalar Field Topology

There is a tight link between pathways in chemical systems and the topology of the scalar field representing energy. Scalar field topology characterizes data by topology changes of level sets. Given a Morse function, i.e., a smooth, real-valued function without degenerate critical points, level set topology changes only at isolated critical points [Mil63]. For energy data, these critical points are energy minima (stable states), saddle points (lowest energy barrier from one stable state to another), and energy maxima. The *contour tree* [BR63, CSA03, TFO09] (or the more

general *Reeb graph* [Ree46]) and the *Morse* and *Morse-Smale* complex are structures that relate these critical points to each other and support ranking topological features by importance as well as simplifying the global topological structure [CSvdP04, GNP*05].

A contour tree tracks changes of the number of connected components of the level set of a function defined on a simply connected domain, resulting in a tree-like structure representing the topological structure of the function. Omitting local maxima in the contour tree results in the so-called *barrier tree* (cf. Section 2.3), which provides a meaningful representation of scalar energy functions by indicating the energy required to transition from one stable state to another [FHSW02]. Whereas these structures show all minima of interest, they show only a subset of barriers. More importantly, only paths across lowest barriers are indicated in the graph layout, which makes them unusable for our application, thus rendering visualization paradigms based on contour trees, such as the topological landscape metaphor [WBP07], are unusable.

In contrast, the Morse complex and the Morse-Smale complex [EHZ03, EHNP03] capture *all* barriers. Assuming the function on the manifold is differentiable and the gradient is defined at each location, it is possible to start at any location in the domain and follow a gradient line either to its origin (its minimum or lower saddle) or its destination (its maximum or upper saddle). Following gradient lines to their origin or their destination yields the Morse complex comprised of either descending or ascending manifolds, respectively. A superimposition of these ascending and descending manifolds results in the Morse-Smale complex. Gerber et al. [GBPW10] proposed a novel visual analysis technique based on Morse-Smale approximation. As we are only interested in energy minima (stable states) and their transition, we have adopted a simpler Morse complex approximation that only considers descending stable segments corresponding to minima.

2.3. Topological Analysis in Chemistry

Flamm et al. [FHSW02, FHSS07] proposed the barrier tree concept for the analysis of degenerate energy landscapes, and Heine et al. [HSF*06] described a visualization method for multiple barrier trees. As discussed in the previous section, barrier trees are related to contour trees and omit barriers that are of interest to our application. Similarly, James et al. [JWR07] introduced the concept of the *disconnectivity graph* to visualize energy landscapes of water clusters in a uniform electric field. This graph is similar to a barrier tree but uses a different visual representation.

Recently, Okushima et al. [ONIS09] proposed to generalize the disconnectivity graph concept to a *saddle connectivity graph*. To construct this saddle connectivity graph, their method locates minima and saddle points and combine them

into a schematic line representation. While the saddle connectivity graph shows all relevant barriers, our method is computationally more efficient and our layout method facilitates easier tracking of separate transition pathways.

The *stochastic roadmap simulation* [ABG*02] represents another approach to the analysis of energy landscapes. However, this method focuses on obtaining a high-level overview, e.g., finding the global minimum in a protein folding problem. Consequently, it can miss minima and saddles that may be important. This method can analyze system with larger number of degrees of freedom at the price of being stochastic and not exploring the entire domain of the configuration space. On the contrary, our method is oriented to provide information about the entire energy landscape within the domain and let the user choose the desired level of detail.

2.4. Dimension Reduction

Whereas the graph itself is an abstract structure, we have to find a proper layout to visualize it. While direct visualization of the Morse complex graph (MCG) using the original minima positions is possible for up to three-dimensional systems (see Figure 9a), this representation quickly becomes cluttered and too complicated for visual analysis. Furthermore, higher-dimensional systems require an alternate layout of the MCG in a plane. Although standard graph-layout algorithms may produce visually pleasing visualizations of the MCG, this representation does not consider relational information of the original system. This information is important for chemical analysis due to impact of distances on the probability of certain transformations.

To preserve this relational context, it becomes necessary to project the high-dimensional point locations of the graph onto the plane. Depending on the type of input data and the required results, there exist a variety of methods to reduce the dimensionality of data. (We refer to the survey by Fodor [Fod02] for an overview). Multidimensional scaling is a well-accepted method [GTS04] to reduce the dimensionality of data while maintaining spatial distances of points from the original data in their projection. Considering the fact that our points are provided in an Euclidean space and we are using a linear scaling function, the method simplifies to what is commonly known as principal component analysis (PCA). Using the principal components of the spread of points the new coordinate system ensures a good preservation of distances which is important for understanding the data.

3. Algorithm

Our technique uses a combination of the Morse complex capturing the relationship between energy minima (Section 3.1), multi-dimensional scaling for projecting minima positions to the two-dimensional plane, and, finally, a graph layout incorporating all necessary chemical information (Section 3.4). We further consider simplification schemes

for noise reduction and for focusing attention on the most relevant features of the chemical system (Section 3.2). We also introduce a strategy for handling additional complexity added by a possible periodic nature of the input data (Section 3.3). Resulting graphs show relevant transitions between energy minima and provide a succinct summary view of relevant transformations of the considered chemical system.

3.1. Morse Complex

To obtain structural transition information, we approximate the Morse complex for the energy function f_e by producing a segmentation of the input data into stable/descending manifolds. We implemented this segmentation similar to the approach of Gerber et al. [GBPW10] by finding a neighbor with the steepest descending gradient for each point, and subsequently applying a union-find algorithm to determine its corresponding minimum.

Segmented descending regions in the resulting representation correspond to classes of configurations associated with a minimum (i.e., stable state). Each descending region can adjoin several other regions. The saddle between two regions, i.e., the lowest function value along their boundaries, corresponds to the barrier between the states. Since minima and saddles in the Morse complex correspond to energy minima and barriers, chemical transformation paths are given by the edges connecting two minima through a saddle. Assigning minima and saddles to nodes and connecting pathways as edges creates the Morse complex graph (MCG).

The major steps of our algorithm can be summarized as follows. First, we perform a sweep over all input points and identify and store the steepest descending neighbor for each point. Subsequently, we use a union-find data structure to segment all input points into regions whose gradient flow ends at the same minimum, thus approximating stable manifolds. For each segment, we consider its neighboring segments and identify the lowest value along the boundary. We construct the MCG by adding or updating three nodes—two minima and saddle, and the two connecting edges. Finally, we check whether any of the minimum-saddle edges crosses a periodic domain boundary, and we also determine the direction when this happens. We preserve this information for future use for the graph layout stage (see Section 3.4).

3.2. Simplification

Noise in a dataset along with increased data complexity, e.g., due to a larger number of dimensions can result in convoluted final graphs. To ensure that a generated graph does not overwhelm users, we need effective means of simplifying the MCG before presenting and analyzing it. To simplify the MCG, we have implemented a scheme that eliminates low-persistence minima. In our approach, we define the persistence for each minimum-saddle pair as the absolute difference of energy function values. Since one min-

imum might belong to several minimum-saddle pairs, and thus have multiple associated persistence values, we select the lowest among them to decide whether a minimum is eliminated. This approach gives us a good measure to identify the noise, which usually corresponds to minima that appeared due to numerical error. Furthermore, we found that minima with a very low persistence correspond to a latent chemical configurations, hence their elimination does not affect the analysis. Finally, we specify a persistence threshold, such that minima with persistence below the threshold are merged with a neighboring minima that have the lowest saddle value.

We perform sequential elimination of minima, ordered by their persistence, as follows. First, we identify the minimum with the lowest persistence. Subsequently, we identify a minimum that is connected through the saddle to the minimum we want to reduce. We then merge the two minima into the new minimum, which inherits all the neighbors of its parents. Finally, we recalculate all the saddles(barriers) along the newly created border of the new minimum and pass on any information associated with both minima (e.g., periodicity information).

We consider two approaches for finding a suitable persistence threshold. The conventional approach for finding this threshold is based on a persistence diagram, which is obtained by gradually increasing the simplification threshold value within the possible value range and determining the current number of minima for each value. (We refer to Section 4 for an example of a persistence diagram.) Big drops in the number of minima indicate a large number of related minima and thus candidates for appropriate simplification threshold values. For example, the persistence of the minima introduced by numerical errors normally varies within a certain range corresponding to the noise threshold. While this approach can be very useful due to its generality, we also utilize an alternative calculation based on chemical domain knowledge about the considered system. Combining both approaches provides a user usually with sufficient information to select an appropriate simplification persistence.

We believe that the combination of this simplification scheme and our graph layout algorithm (see Section 3.4) sufficiently simplifies the resulting visualizations. (We introduce further means for further simplifying the resulting visualization in Section 5).

3.3. Periodicity

Proper handling of periodicity for different chemical structures and presenting them in a simple way for analysis is key to providing readable graphs. Chemistry datasets can also have some of their dimensions being periodic, and handling and presenting such datasets leads to additional visualization challenges. Our approach handles such datasets elegantly by marking edges of the MCG that cross the domain boundary

of a periodic dimension and presents them in a graph using dashed edges. Furthermore, we indicate directions, which could be crucial in the analysis of porous materials.

Our interpretation of periodicity in the data is straightforward: The maximum boundary and the minimum boundary along an axis corresponding to a periodic dimension will be virtual neighbors when wrapping around the data set. Thus, a maximum and minimum cell along that axis share a face, and the triangulation must subdivide that face consistently. To ensure this consistency and avoid possible cracks and hanging nodes in a triangulation that might occur on the domain's boundary faces, we use Freudenthal's subdivision scheme [Fre42], which subdivides opposite cell faces consistently. While Freudenthal subdivision generalizes to arbitrary dimensions, the number of edges in a mesh grows exponentially, making this approach infeasible for dimensions larger than three. Thus, we utilize a *k*-nearest-neighbor algorithm [TFO09] to create neighborhood graphs for high-dimensional datasets, and guarantee a correct triangulation by removing edges in the neighborhood graph that are not bi-directional.

3.4. Graph Layout and Drawing

As a first step in our graph layout algorithm, we project the location of minima to two dimensions using classical metric multidimensional scaling (MDS). While it seems intuitive to apply MDS both to minimum and saddle positions, we found that this approach often leads to a cluttered graph layout. Projecting only minima and placing saddles along lines between minima results in a much cleaner graph layout and supports encoding additional information based on the location of barriers.

While the projection at this point is simple to navigate, overlaps of edges and nodes might conceal important information. Thus, we implemented an iterative optimization algorithm that minimizes node/edge overlap in the final layout. We define the node's center N_{center} and the node's projection to the edge's mid-axis as $P_{N_{center}}$. Then, depending on a distance between them, we can derive if the node and the edge overlap or not. We formally define the overlap function as $D_{overlap} = (N_{radius} + E_{width}/2) - \text{distance}(N_{center}, P_{N_{center}})$, where N_{radius} is the radius of the node, and E_{width} is the width of the edge. If the $D_{overlap}$ is positive, we have an overlap, otherwise not. Thus, the goal is to iteratively move the node away from the edge along the projection line, until $D_{overlap}$ becomes zero or negative. To ensure the preservation of the overall structure, we add a condition that the center of the node does not move further than the predefined ε radius. Although we allow users to set parameters for faster or clearer view of the final graph layout, we found the convergence rate of the algorithm satisfactory.

Once we have obtained final positions for all nodes of the graph, we add chemical information from the model system

to the graph. Since the most important information is the *energy difference* between a minimum and a barrier, we color edges according to the energy difference value. The Boltzmann distribution states that the probability of accessing a state decreases exponentially with its energy. The lower the energy difference is, the higher is the probability of a transformation. Hence, we display paths with smaller energy differences more prominently (darker colors and wider edges), see Figure 1. We note that a higher energy difference leads to very narrow and lightly colored edges, making them almost invisible. As a consequence, these edges can be neglected when drawing the final layout. Thus, we introduce the notion of an upper threshold that signifies the highest possible and/or interesting energy difference threshold, so that a user can choose to discard certain edges and thus further simplify the layout. Periodic edges are dashed and the common saddle of the two connected minima is broken into two nodes (see Figure 2c). This helps chemist to identify periodic edges right-away. Finally, we add labels to all minima that specify coordinates (first line) and energy value (second line). This information helps chemists to relate the graph to the state of the model system and it guides analysis.

4. Results

We demonstrate our visualization technique using two distinct examples of chemical systems: (i) a dimer of formic acid and acetic acid (DFA), in which we study energetics of conformational changes and proton transfer reactions; and (ii) methane molecule in a crystalline porous material, in which we investigate energies related with diffusion.

4.1. Dimer of Formic Acid and Acetic Acid

Small "model" systems are often selected by chemists to study intra- and inter-molecular interactions. The small size of such complexes does not only facilitate gas-phase experiments (small systems usually have high vapor pressure) but also enables performing accurate *ab initio* calculations to investigate properties and simulate transformations in these complexes. In this context, a dimer of carboxylic acids can serve not only as model of hydrogen-bonded system but also can be used to study proton transfer reactions [BHD*05]. In the current study we selected a dimer of formic acid and acetic acid as a model system to present generation of our graph representations of transformation pathways in this chemical system.

The initial geometry of DFA presented in Figure 2a was optimized at the semi-empirical PM3 level of theory [Ste89a, Ste89b] to find a minimum energy configuration. Here, we consider up to four degrees of freedom in this system. Therefore, we used the latter optimized geometry to generate a four-dimensional grid with energies of DFA in the considered configurations. The geometries of each configuration were generated in the following manner. All atomic

positions (in internal coordinates), except those involved in the considered degree of freedoms, were fixed at their optimal positions. The remaining coordinates of DFA were systematically modified along the four considered directions. Specifically, dimensions of the energy grid are defined as follows:

- First dimension: length of H7–O3 bond is scanned between 0.85Å and 1.85Å with a step size of 0.05Å($1Å=10^{-10}$m).
- Second dimension: length of H8–O1 bond is scanned from 0.85Å to 1.85Å with a step size of 0.05Å.
- Third dimension: H7–O3–C5–O4 dihedral angle, which is scanned with a step size of $15°$(24 samples, periodic). This transformation rotates the H7–O3 group around the C5–O3 axis and has been selected to investigate the energetic effect, associated with disrupting the O3–H7…O2 hydrogen bond.
- Fourth dimension: rotation of C6 methyl group along the C6–C5 axis. It is defined by the H9–C6–C5–O3 dihedral angle, sampled at a step size of $15°$. During the rotation of the methyl group, all internal coordinates of atoms of the methyl group are fixed.

For each of the generated geometries, we used the Gaussian03 program [FTS*] to obtain the corresponding energy at the semi-empirical PM3 level. The resulting 4D energy grid was used to generate our graph representations. The energy unit used throughout this example is Hartree ($1Ha \approx 27.211eV$).

To simplify the referencing, we define F_X as an energy function of the system, where $X = \{1, 2, 3, 4, 12, 13, 14, 23, 24, 34, 123, 124, 134, 234, 1234\}$ is a set of dimension combinations. For example F_{124} corresponds to the energy function, defined in space of the first, second and fourth dimensions. Absent dimensions in our case are assumed to have a default (optimal) fixed coordinate. Each graph representation is referred to correspondingly as a M_X. Note that we use scaled coordinates between zero and one in our calculations and figures presenting F_X.

In Figure 3 we show the energy function for each dimension of the system. We can see that F_1 and F_2 have only one minimum each, thus no transformations between minima exist in the system. However F_3 has three minima, thus three transformations are possible. M_3 describes all possible transformations and barriers, and handles the periodic nature of the dimension by showing the barrier at 0 (and 359) as the same node on the left and right sides. Therefore, we have two ways of getting from one minimum to any other. (It naturally corresponds to left and right rotations of the methyl group of acetic acid.) Now, one would expect that since F_4 is also periodic, M_4 would have two different paths to get from one minimum to the other. However, we are interested only in the lowest barrier between two minima, thus M_4 shows only the lower of two barriers. Indeed, if the user has to choose
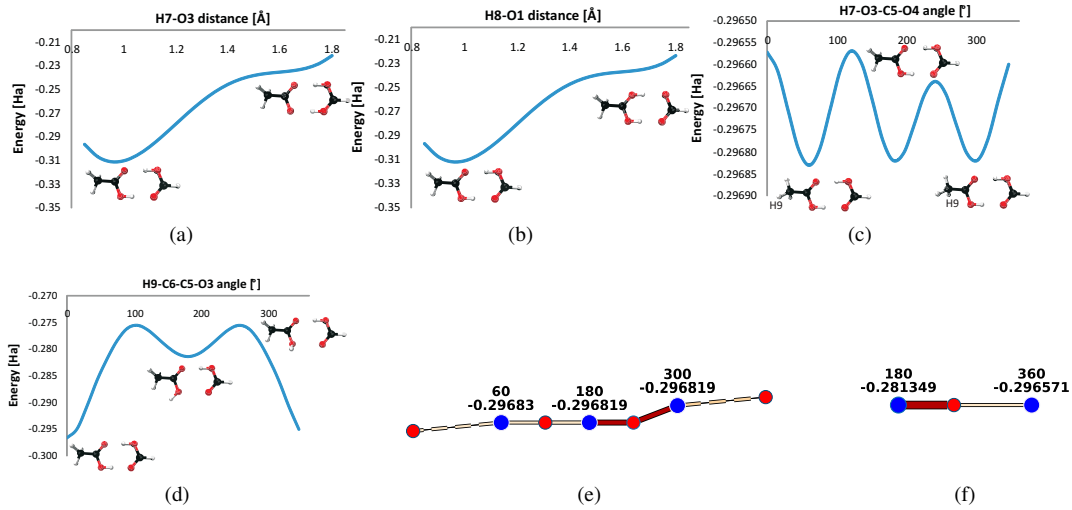
Figure 3: Energy functions of DFA along the (a) first dimension, (b) second dimension, (c) third dimension, and (d) forth dimension, and respective graphs for (e) the third dimension, showing the periodic nature of the dimension and (f) the fourth dimension, where it correctly handles the special case of periodicity. All labels in (e) and (f) use the convention from Figure 2.
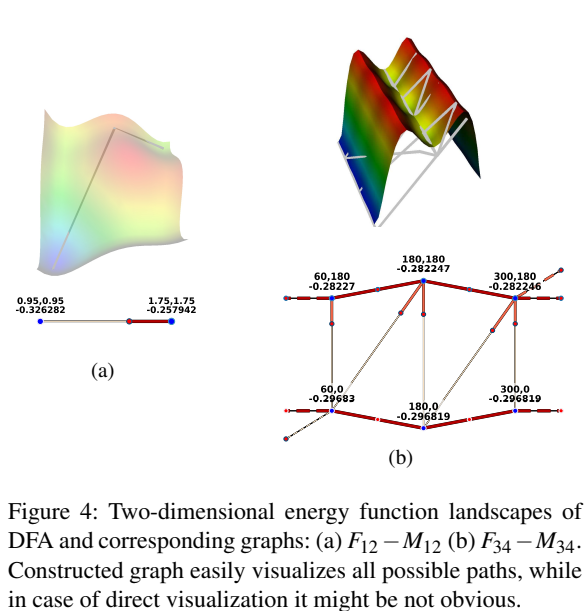

Figure 4: Two-dimensional energy function landscapes of DFA and corresponding graphs: (a) $F_{12} - M_{12}$ (b) $F_{34} - M_{34}$. Constructed graph easily visualizes all possible paths, while in case of direct visualization it might be not obvious.

Figure 5: Three-dimensional energy function volume renderings of DFA and corresponding graphs: (a) $F_{123} - M_{123}$ (b) $F_{124} - M_{124}$. As the number of dimensions grows, direct visualization becomes complicated.

the rotation direction, (s)he would choose the one that goes through the lowest barrier.

Figure 4 shows the energy function landscapes and corresponding graph representations for each of two selected dimension pairs. One-dimensional analysis is simple, but two-dimensional functions introduce more complexity. F_{12} in Figure 4a is still relatively simple. We find two minima and the lowest point connecting them. $F_{13}, F_{14}, F_{23}, F_{24}, F_{34}$ are not so intuitive due to one or two periodic dimensions. While it is obvious from the energy function landscape, we might have several transformation paths between each two
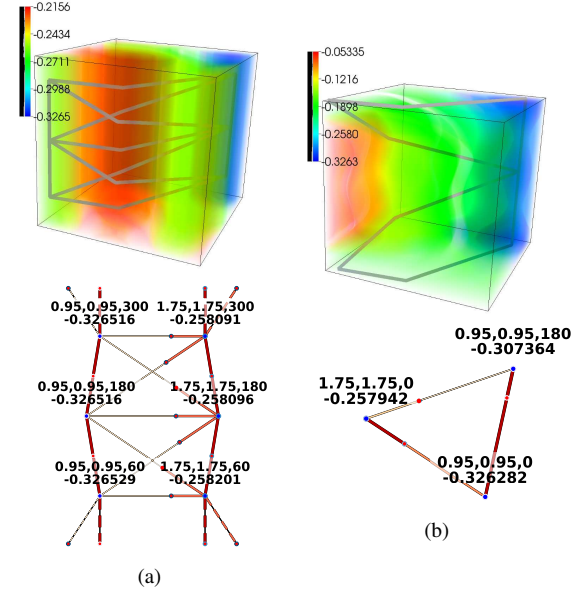
minima. For example in Figure 4b we see multiple possible paths between the minimum at (60,180) and the minimum at (300,180).

Figure 5 shows two examples of the energy functions for three dimensions. Since the energy functions becomes complicated in three dimensions, we have to rely on our graph representations to observe next important characteristics.
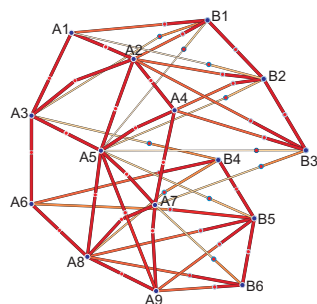
Figure 6: Graph of the 4D energy function of DFA. Nodes are labeled A1 to A9 and B1 to B6 to highlight two subgraphs. A1-A9 correspond to minima with coordinates of 0.95Å along the first and second dimension. The remaining third and fourth coordinates are (45,345), (180,345), (60,180), (285,345), (180,180), (60,0), (300,180), (180,0), and (300,0) for A1 to A9, respectively. Similarly, the first and second coordinate for B1-B6 are 1.75Å. The remaining coordinates are (60,345), (180,345), (300,345), (60,0), (180,0), (300,0) for B1 to B6, respectively.

Thorough analysis of M_{124} in Figure 5b shows that in a given system, a much better transformation between the minimum at (0.95,0.95,180) and the minimum at (1.75,1.75,0) can be found by visiting the third minimum at (0.95,0.95,0), although direct transformation is also available.

The final example is the graph representation of the 4D energy function of DFA depicted in Figure 6, for which plotting the original energy functions in all four dimensions was not feasible. The resulting graph is fairly easy to analyze. It clearly shows energy separation of the graph into two subgraphs: one corresponding to rotamers of DFA (with minima denoted A1-A9) and another with double proton transfer (minima denoted B1-B6). Transitions between minima within each subgraph are at low cost as shown by heavy edges.

4.2. Free Energy of a Guest Molecule in a Porous Material

The second example of application of our approach involves analysis of free energy of a guest molecule inside a porous material. Porous materials contain complex networks of void channels and cages that are exploited in many different industrial applications. Zeolites, probably the most recognized class of crystalline porous materials, have found wide use in industry since the late 1950s. They are commonly used as chemical catalysts, membranes and adsorbents for separations and water softeners [ACD04, SM08b, SM08a, KvB07].

One of the key processes that determine performance of membranes is diffusion of guest species. Diffusion of gases inside a porous material is controlled by free energy barriers. Analysis of possible diffusion pathways and the associated barriers is therefore critical to understand and design optimal
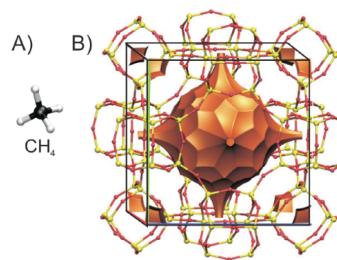

Figure 7: Picture of (a) methane (b) LTA zeolite structure. The orange isosurface highlights the closest distance to which the center of the guest molecule center can approach. The large cage is located in the center of the unit cell. The small cage is shared among eight cells and is visible in the corners of the unit cell.
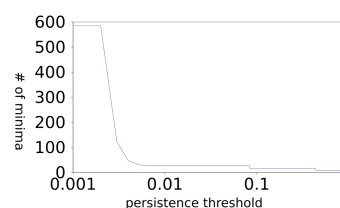

Figure 8: Persistence diagram of the energy function of a CH_4 guest molecule in porous material.

separation devices. Our representation may facilitate such understanding.

In the following we demonstrate our graph representation of free energy of CH_4 molecule inside LTA zeolite. The corresponding 3D free energy grids were prepared by the following procedure. The 3D space describing a periodic box of LTA zeolite was divided into 239x239x239 volumetric bins. We performed a Monte Carlo simulation to predict an average free energy of a guest molecule inside each bin. The details of this procedure can be found in [KGK*96]. We used a force-field approach developed by García-Pérez et al. [GPPA*07]. The energies used in this example are expressed in k_BT.

The free energy function of CH_4 in LTA zeolite is substantially different from the energy functions of DFA. The noise of the dataset is high, thus it required application of the described simplification scheme to reduce the final graph, based on the persistence diagram in Figure 8. Figure 9 shows free energy and the corresponding graph representation. This graph represents only the large cage of LTA zeolite, which is the only fragment of void space in LTA accessible to CH_4 (see [HS10] for discussion of accessibility). The graph representation of the free energy reveals important information about the material. There are 14 important energy minima per periodic unit cell of LTA corresponding to favorable locations of the adsorbing CH_4 molecule (adsorption sites).
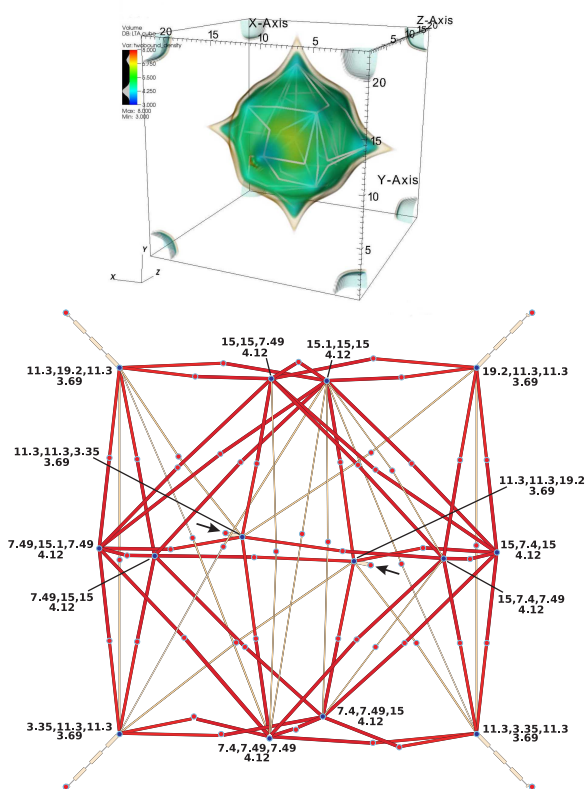
Figure 9: (a) Volume rendering of the energy function of a CH_4 molecule in an LTA zeolite and lowest energy paths connecting neighboring minima. (b) Corresponding graph showing lowest energy paths (an edge going through a face of the periodic box marked with arrows).

Six of them correspond to lower energy (ca. 3.7 $k_B T$), and are localized near windows connecting two periodic cells (near faces of the unit cell). The remaining eight minima are localized further away from the windows, on the surface of the large cage of LTA. The further analysis of connections between nodes/minima in our representations reveals that all 14 minima localized within the big cage are separated by low barriers, and therefore hops of CH_4 between the adsorption sites are feasible. However, connections between large cages in the extended material lead through high barriers. These high barriers along diffusion paths in every direction are reflected in slower diffusion rates.

5. Conclusions and Possible Extensions

The described visual representation provides new capabilities to visualize complex multidimensional energy functions. Our approach highlights the most relevant information for chemists: number and location of energy minima and heights of connecting barriers. The resulting map allows investiga-

tion of all possible transformation (e.g., reaction) pathways and the identification of the lowest energy paths.

The main motivation for developing our approach was the lack of tools to visualize energetics of transformation pathways in complex systems. We expect that our energy graphs will be widely used by scientists working on porous materials as they provide new capabilities: (i) display diffusion paths and the corresponding energy barriers to gain insight into materials properties and to set up proceeding molecular simulations; (ii) display adsorption sites, which can be correlated with features in adsorption isotherms; and (iii) statistics describing our energy graphs can be used to compare different materials. Moreover, in studies on porous materials, free energy grids are routinely calculated in the course of their characterization. For that reason, our energy maps can be obtained at almost no additional cost. The usefulness of our approach in studies of other similar systems will have to be verified by chemists.

The two examples presented in Section 4, have involved datasets defined in up to four dimensions. Our visualization method can handle a much higher number of dimensions; however, the number is currently limited in practice to six to eight dimensions by the size and cost of calculating the energy defined on a reasonably discretized configuration space. The time required to calculate the energy at one point may vary between tenths of a second (classical force field) to weeks (accurate *ab initio* quantum mechanics (QM). We expect that one can generate three-dimensional energy grids using *ab initio* QM, four- to five-dimensional grids using semi-empirical QM, and six- to eight-dimensional grids using force fields.

We plan to further improve our approach to include additional information, e.g., indicate lowest energy pathways with distinct color. We also explore to further simplification for large and/or complex graphs. For example, (i) display only selected fragment of a graph; (ii) simplification of graphs, for example by partial clustering of nodes. Finally, we intend to develop a user interface to interactively adjust persistence and upper thresholds.

6. Acknowledgements

References

[ABG*02] APAYDIN M. S., BRUTLAG D. L., GUESTRIN C., HSU D., LATOMBE J.-C.: Stochastic roadmap simulation: An efficient representation and algorithm for analyzing molecular motion. In *Proceedings RECOMB'02* (2002), pp. 12–21. 4

[ACD04] AUERBACH S., CARRADO K., DUTTA P. K.: *Handbook of Zeolite Science and Technology*. Marcel Dekker, 2004. 8

[BHD*05] BACHORZ R., HARANCZYK M., DABKOWSKA I., RAK J., GUTOWSKI M.: Anion of the formic acid dimer as a model for intermolecular proton transfer induced by a π* excess electron. *Journal of Chemical Physics 122*, 20 (May 2005), 204304. 6

[BR63] BOYELL R. L., RUSTON H.: Hybrid techniques for real-time radar simulation. In *Proceedings of the 1963 Fall Joint Computer Conference* (1963), IEEE, pp. 445–458. 3

[CSA03] CARR H., SNOEYINK J., AXEN U.: Computing contour trees in all dimensions. *Computational Geometry – Theory and Applications 24*, 2 (Feb. 2003), 75–94. 3

[CSvdP04] CARR H., SNOEYINK J., VAN DE PANNE M.: Simplifying flexible isosurfaces using local geometric measures. In *IEEE Visualization 2004* (Oct. 2004), IEEE, pp. 497–504. 3

[EHNP03] EDELSBRUNNER H., HARER J., NATARAJAN V., PASCUCCI V.: Morse-smale complexes for piecewise linear 3-manifolds. In *Proceedings of the 19th ACM Symposium on Computational Geometry* (2003), pp. 361–370. 3

[EHZ03] EDELSBRUNNER H., HARER J., ZOMORODIAN A.: Hierarchical morse-smale complexes for piecewise linear 2-manifold. *Discrete & Computational Geometry 30* (2003), 87–107. 3

[FHSS07] FLAMM C., HOFACKER I. L., STADLER B. M. R., STADLER P. F.: Saddles and barrier in landscapes of generalized search operators. *Foundations of Genetic Algorithms 4436* (2007), 194–212. 3

[FHSW02] FLAMM C., HOFACKER I. L., STADLER P. F., WOLFINGER M. T.: Barrier trees of degenerate landscapes. *Z. Physical Chemistry 216* (2002), 155–173. 3

[Fod02] FODOR I.: *A Survey of Dimension Reduction Techniques*. Tech. rep., LLNL, 2002. 4

[Fre42] FREUDENTHAL H.: Simplizialzerlegung von beschrankter flachheit. *Ann. Math. 43* (1942), 580–582. 5

[FTS*] FRISCH M. J., TRUCKS G. W., SCHLEGEL H. B., SCUSERIA G. E., ROBB M. A., CHEESEMAN J. R., MONTGOMERY JR. J. A., VREVEN T., KUDIN K. N., BURANT J. C., MILLAM J. M., IYENGAR S. S., TOMASI J., BARONE V., MENNUCCI B., COSSI M., SCALMANI G., REGA N., PETERSSON G. A., NAKATSUJI H., HADA M., EHARA M., TOYOTA K., FUKUDA R., HASEGAWA J., ISHIDA M., NAKAJIMA T., HONDA Y., KITAO O., NAKAI H., KLENE M., LI X., KNOX J. E., HRATCHIAN H. P., CROSS J. B., BAKKEN V., ADAMO C., JARAMILLO J., GOMPERTS R., STRATMANN R. E., YAZYEV O., AUSTIN A. J., CAMMI R., POMELLI C., OCHTERSKI J. W., AYALA P. Y., MOROKUMA K., VOTH G. A., SALVADOR P., DANNENBERG J. J., ET AL.: Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004. 6

[GBPW10] GERBER S., BREMER P.-T., PASCUCCI V., WHITAKER R.: Visual exploration of high dimensional scalar functions. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov. 2010), 3, 4

[GNP*05] GYULASSY A., NATARAJAN V., PASCUCCI V., BREMER P.-T., HAMANN B.: Topology-based simplification for feature extraction from 3D scalar fields. In *Proceedings of the IEEE Visualization 2005 (VIS'05)* (2005), IEEE Computer Society, pp. 535–542. 3

[GPPA*07] GARCÍA-PÉREZ E., PARRA J. B., ANIA C. O., GARCÍA-SÁNCHEZ A., VAN BATEN J. M., KRISHNA R., DUBBELDAM D., CALERO S.: A computational study of co2, n2, and ch4 adsorption in zeolites. *Adsorption 13*, 5-6 (Sept. 2007), 469–476. 8

[GTS04] GARTH C., TRICOCHE X., SCHEUERMANN G.: Tracking of vector field singularities in unstructured 3d time-dependent data sets. In *Proceedings of IEEE Visualization 2004* (2004), pp. 329–336. 4

[HS10] HARANCZYK M., SETHIAN J. A.: Automatic structure analysis in high-throughput characterization of porous materials. *Journal of Chemical Theory and Computation 6* (2010), 3472–3480. 8

[HSF*06] HEINE C., SCHEUERMANN G., FLAMM C., HOFACKER I. L., STADLER P. F.: Visualization of barrier tree sequences. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (Sept./Oct. 2006), 781–788. 3

[JWR07] JAMES T., WALES D. J., ROJAS J. H.: Energy landscapes for water clusters in a uniform electric field. *The Journal of Chemical Physics 126* (2007), 054506. 3

[KGK*96] KEFFER D., GUPTA V., KIM D., LENZ E., DAVIS H. T., MCCORMICK A. V.: A compendium of potential energy maps of zeolites and molecular sieves. *Journal of Molecular Graphics and Modelling 14* (1996), 108–116. 3, 8

[KvB07] KRISHNA R., VAN BATEN J.: Using molecular simulations for screening of zeolites for separation of co2/ch4 mixtures. *Chemical Engineering Journal 133* (Sept. 2007), 121–131. 8

[Mil63] MILNOR J. W.: *Morse Theory*. Princeton University Press, Princeton, New Jersey, May 1963. 3

[ONIS09] OKUSHIMA T., NIIYAMA T., IKEDA K. S., SHIMIZU Y.: Graph-based analysis of kinetics on multidimensional potential-energy surface. *Physical Review E 80*, 3 (2009), 036112. 3

[Ree46] REEB G.: Sur les points singuliers d'une forme de pfaff complement intergrable ou d'une fonction numerique. *Comptes Rendus Acad. Science Paris 222* (1946), 847–849. 3

[SM08a] SMIT B., MAESEN T. L. M.: Molecular simulations of zeolites: Adsorption, diffusion, and shape selectivity. *Chemical Reviews 108* (Sept. 2008), 4125–4184. 8

[SM08b] SMIT B., MAESEN T. L. M.: Towards a molecular understanding of shape selectivity. *Nature 451* (Feb. 2008), 671–678. 8

[Ste89a] STEWART J. J. P. J.: Optimization of parameters for semiempirical methods i. method. *Journal of Computational Chemistry 10* (Mar. 1989), 209–220. 6

[Ste89b] STEWART J. J. P. J.: Optimization of parameters for semiempirical methods ii. applications. *Journal of Computational Chemistry 10* (Mar. 1989), 221–264. 6

[TFO09] TAKAHASHI S., FUJISHIRO I., OKADA M.: Applying manifold learning to plotting approximate contour trees. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1185–1192. 3, 5

[WBP07] WEBER G., BREMER P.-T., PASCUCCI V.: Topological landscapes: A terrain metaphor for scientific data. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1416–1423.

Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.