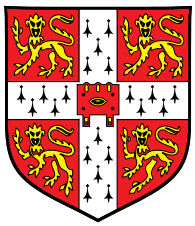

University of Cambridge
Department of Chemistry

Pathways and Energy Landscapes

A thesis submitted to the University of Cambridge in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Semen A. Trygubenko

Darwin College

January 2006

Declaration

The work described in this dissertation was carried out in the Department of Theoretical Chemistry at the University of Cambridge between October 2002 and January 2006. The contents are the work of the author and contain nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements, and have not previously been submitted for any degree or qualification at another institution. The number of words does not exceed 60 000.

Semen A. Trygubenko

Cambridge, January 2006

Acknowledgements

I am glad I have the opportunity to express my sincere gratitude to all the people that accompanied and supported me during my study.

I had the pleasure of being supervised by Dr. David Wales and would like to thank him for the high-quality guidance I received in the duration of the Ph.D. course.

The second person I gratefully acknowledge is Dr. Catherine Pitt who introduced me to the Debian GNU/Linux operating system and various scientific applications that were heavily used throughout the development of this work.

My appreciation also goes to Dr. Thomas Middleton for providing transition state databases for the binary Lennard-Jones system, Dr. David Evans for his Dijkstra-based utilities that locate shortest paths in connected databases of minima and transition states, Dr. Mark Miller for many KMC-related discussions as well as his programs for creating and manipulating disconnectivity graphs, and Dr. Dmytro Hovorun and Dr. Pavel Hobza who greatly influenced my decision to do a Ph.D. in this area.

The chain of my gratitude would definitely be incomplete without special thanks to Tetyana Bogdan, Dr. Joanne Carr and Tim James, for many stimulating discussions we had on various aspects of my research, the Life, the Universe and everything.

Lastly, I would like to thank Cambridge Commonwealth Trust, Cambridge Overseas Trust and Darwin College for financial support. Most of the calculations were performed using computational facilities funded by the Isaac Newton Trust.

To My
Teachers

Abstract

A modification of the nudged elastic band (NEB) method is presented that enables stable optimisations to be run using both the limited-memory quasi-Newton (L-BFGS) and slow-response quenched velocity Verlet minimisers. The performance of this new 'doubly nudged' DNEB method is analysed in conjunction with both minimisers and compared with previous NEB formulations. We find that the fastest DNEB approach (DNEB/L-BFGS) can be quicker by up to two orders of magnitude. Applications to permutational rearrangements of the seven-atom Lennard-Jones cluster (LJ_7) and highly cooperative rearrangements of LJ_{38} and LJ_{75} are presented.

Secondly, we propose new measures of localisation and cooperativity for the analysis of atomic rearrangements. We show that for both clusters and bulk material cooperative rearrangements usually have significantly lower barriers than uncooperative ones, irrespective of the degree of localisation. We also find that previous methods used to sample stationary points are biased towards rearrangements of particular types. Linear interpolation between local minima in double-ended transition state searches tends to produce cooperative rearrangements, while random perturbations of all the coordinates, as sometimes used in single-ended searches, has the opposite effect.

Thirdly, we report a new algorithm for constructing pathways between local minima that involve a large number of intervening transition states on the PES. A significant improvement in efficiency has been achieved by changing the strategy for choosing successive pairs of local minima that serve as endpoints for the next search. We employ Dijkstra's algorithm to identify the 'shortest' path corresponding to missing connections within an evolving database of local minima and the transition states that connect them.

Finally, we describe an exact approach for calculating the total transition probabilities in finite-state discrete-time Markov processes. All the states and the rules for transitions between them must be known in advance. We can then calculate averages over a given ensemble of paths for both additive and multiplicative properties in a non-stochastic and non-iterative fashion. In particular, we can calculate the mean first passage time between arbitrary groups of stationary points for discrete path sampling databases, and hence extract phenomenological rate constants. We present a number of examples to demonstrate the efficiency and robustness of this approach.

Contents

List of Abbreviations

BFGS	Broyden-Fletcher-Goldfarb-Shanno minimisation algorithm [1–3]
BFM	Bellman-Ford-Moore algorithm
BKL	Bortz-Kalos-Lebowitz algorithm
BLJ	Binary Lennard-Jones potential
CG	Conjugate Gradient
CHARMM	Chemistry at HARvard Molecular Modelling program [4]
CHARMM19	United-atom CHARMM force field [5]
CHARMM22	All-atom CHARMM force field [6]
CPU	Central Processing Unit
CS	Chain-of-States approach
DFP	Davidon-Fletcher-Powell optimisation method
DGT	Dense-optimised Graph Transformation method
DNEB	Doubly Nudged Elastic Band method [7]
DPS	Discrete Path Sampling method [8–10]
EEF1	Effective Energy Function 1 solvation potential [11]
EF	Eigenvector-Following method [12–24]
GT	Graph Transformation method
KMC	Kinetic Monte Carlo method
L-BFGS	Limited memory version of the BFGS algorithm [2]
LJ	Lennard-Jones potential
LST	Linear Synchronous Transit method
MB	Müller-Brown surface [25]
MCAMC	Monte Carlo with Absorbing Markov Chains [26]
MD	Molecular Dynamics method

MM	Matrix Multiplication method
NEB	Nudged Elastic Band method [27–30]
NMR	Nuclear Magnetic Resonance
NR	Newton-Raphson approach [31]
ORT	Overall Rotation and Translation
PDB	Protein Data Bank
PES	Potential Energy Surface
QST	Quadratic Synchronous Transit method
QVV	Quenched Velocity Verlet minimisation algorithm [27]
RMS	Root-Mean-Square
SDGT	A version of the GT method that is a combination of SGT and DGT
SGT	Sparse-optimised Graph Transformation method
SMM	Sparse-optimised Matrix Multiplication method
SQVV	Slow-response Quenched Velocity Verlet minimisation algorithm [7]
SRW	Simple Random Walk
VV	Velocity Verlet integrator [32]

Glossary of Symbols

Consistency is the last refuge of the unimaginative.

Oscar Wilde

$\langle x \rangle$	The mean value of variable x	81
$\emptyset$	Empty set	105
$\hat{}$	Denotes a unit vector; $\hat{\mathbf{x}} = \mathbf{x}/\lvert\mathbf{x}\rvert$	12
$\lvert\,\rvert$	A vector norm; $\lvert\mathbf{x}\rvert = \sqrt{\sum_{i=1}^{dim(\mathbf{x})} x_i^2}$; also, cardinality of a set	10
$\ominus$	Denotes symmetric difference of two sets	100
$\otimes$	Denotes vector direct product, a.k.a. dyadic; $\mathbf{a} \otimes \mathbf{b}^T = \mathbf{c}$, $c_{i,j} = a_i b_j$	18
$(\,,\,)$	A scalar product of two vectors, a.k.a. dot product; $(\mathbf{a}, \mathbf{b}) = \mathbf{a}^T \cdot \mathbf{b}$	12
$\{\,\}$	A set of objects; $\{\mathbf{x}_i\}_1^n = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$	10
$\mathbf{0}$	A matrix or a vector filled with zeros	15
A	One of the two superstates in two-state kinetic model	79
$\mathbf{A}$	Approximation to the inverse Hessian matrix, $\mathbf{H}^{-1}$	18
$Adj[i]$	Set of all the nodes adjacent to node i	94
$AdjIn[i]$	Set of all the nodes connected to node i via incoming edges	75
$AdjOut[i]$	Set of all the nodes connected to node i via outgoing edges	75
B	One of the two superstates in two-state kinetic model	79
BLJ_n	Binary Lennard-Jones liquid with n atoms in a periodic cell	65
$\mathbf{C}$	Square matrix, columns of which are eigenvectors	20
C_N	A chain graph with N nodes	85
D	Endpoint separation	57
$\Delta_i(j)$	ith displacement in magnitude of an atom between structures $j-1$ and j	61
E	A set of edges	85
$\mathcal{E}_\alpha^G$	Probability of escape from G starting from α in a single step	80
E_-	Energy barrier corresponding to the reverse reaction	43

E_+	Energy barrier corresponding to the forward reaction	43
F	Set of all minima connected to the final endpoint	40
$\mathcal{F}$	Frequency distribution function	55
G_N	An arbitrary graph with N nodes	89
$\mathbf{H}(\mathbf{x})$	Hessian matrix evaluated at point $\mathbf{x}$	15
I	A set containing all the minima that do not belong to $A \cup B$	79
K_N	A complete graph with N nodes	89
L	Lagrangian function	22
LJ_n	n-atom Lennard-Jones cluster	30
$\mho(\mathbf{x})$	The set of all possible values of the control variable $\mathbf{x}$	15
N	Number of atoms; number of nodes in a graph	55
N_c	Cooperativity index	55
N_f	Number of frames or points sampled along a path	54
N_i	Number of images in a band	10
N_p	Participation index	55
$\widetilde{N}$	Participation index evaluated using the endpoints alone	42
O_k	Displacement overlap evaluated for k atoms using displacements $d_i(j)$	58
$\mathcal{O}(\,)$	$f(n) = \mathcal{O}\Big(g(n)\Big)$ means $0 \leqslant f(n) \leqslant cg(n)$ holds for some constants $c > 0$	87
$\mathbf{P}$	Transition probability matrix	74
P_i^{eq}	Equilibrium occupation probability of state i	79
$P_i(t)$	Occupation probability of state i at time t	78
$P_{j,i}$	Probability of transition from state i to state j	75
$\mathcal{P}_\xi$	Pathway probability	81
R_N	A random graph with N nodes	102
$\mathbf{R}_\alpha$	$3N$-dimensional rotation matrix about axis α, $\alpha \in \{x, y, z\}$	21
$\mathbb{R}$	The set of all real numbers	20
S	Set of all minima connected to the starting endpoint	40
Σ_α^G	Total probability of escape from G if started at node α	80
$\mathcal{S}_{\alpha,\beta}^G$	Sum of weights of all pathways connecting α and β and confined to G	81
T	Temperature	74
Θ_k	Displacement overlap evaluated for k atoms using displacements $\Delta_i(j)$	61

$\mathcal{T}_i^G$	Mean escape time from graph G if started at node i	83
U	Set of all minima that do not belong to $S \cup F$	40
$\Upsilon(\mathbf{x}, \varepsilon)$	A set of feasible points contained in the neighbourhood ε of $\mathbf{x}$	15
V	Potential energy functional; also, a set of graph nodes	2
$\boldsymbol{\mathcal{V}}$	$3N$-dimensional vector of velocities*	17
$\widetilde{V}$	Spring potential	10
$W(a, b)$	Weight of the shortest path $\xi = a \leftarrow b$; $W(a, b) = -\ln(\mathcal{W}_\xi)$	134
$\mathbb{W}$	The set of whole numbers; $\mathbb{W} = \{0, 1, 2, \dots\}$	76
$\mathcal{W}_\xi$	Product of branching probabilities associated with path ξ	81
$\mathbf{X}$	$3N$-dimensional vector representing a point in configuration space	10
Ξ	Pathway ensemble	80
a	A state that belongs to a superstate A	79
α	Pathway nonlinearity index	57
b	A state that belongs to a superstate B	79
β	Energy barrier asymmetry index	43
$\mathbf{c}$	Eigenvector	20
$\det \mathbf{M}$	A determinant [33] (a scalar-valued function) of matrix $\mathbf{M}$	21
d_i	Integrated path length for atom i; also, degree of node i	55
$d_i(j)$	Displacement of atom i between structures $j-1$ and j	55
$e_{j,i}$	Directed edge that describes a transition from node i to node j	75
ϵ	A parameter in LJ potential (the depth of the potential energy well)	2
ε	Small positive parameter	15
η	Number of atomic degrees of freedom	12
$f(\mathbf{x})$	Objective function of a vector argument $\mathbf{x}$	15
$\mathbf{g}$	$3N$-dimensional gradient vector of the true potential	10
γ	Kurtosis of a distribution evaluated using moments about the mean	56
γ'	Kurtosis of a distribution evaluated using moments about the origin	57
$\widetilde{\mathbf{g}}^{\parallel}$	Spring gradient vector component parallel to the path	13
$\widetilde{\mathbf{g}}^{\perp}$	Spring gradient vector component perpendicular to the path	12
$\widetilde{\mathbf{g}}$	$3N$-dimensional gradient vector of the spring potential	10

*This vector and the other vectors defined here are column vectors.

$\mathbf{g}^{\parallel}$	True gradient vector component parallel to the path	10
$\mathbf{g}^{\perp}$	True gradient vector component perpendicular to the path	10
i, j, k	Indices; range and meaning may vary depending on the context	16
k_B	Boltzmann's constant	74
$k_{j,i}$	Rate constant for transitions from state i to state j	78
k_{spr}	Spring force constant	12
$l(\xi)$	Length of path ξ	82
λ	Eigenvalue	20
m	Atomic mass	17
m_n	nth moment of a distribution function about the mean	56
m'_n	nth moment of a distribution function about the origin	57
n	Time parameter of a discrete-time stochastic process	77
$o(\,)$	$f(n) = o\Big(g(n)\Big)$ means $0 \leqslant f(n) \leqslant cg(n)$ holds for all constants $c > 0$*	135
$\mathbf{p}$	Search direction vector	16
π	Path length asymmetry index	42
$\mathbf{r}_i(j)$	Three-dimensional Cartesian coordinates vector of atom i for structure j	54
s	Integrated path length	37
σ	A parameter in the LJ potential ($2^{1/6}\sigma$ is the pair equilibrium separation)	2
$\boldsymbol{\tau}$	$3N$-dimensional tangent vector	12
τ_i	Mean waiting time in state i before escape	78
t	Time	17
δt	Time integration step	17
^{T}	A matrix or vector transpose	15
ϖ	Step size	16
v_i	ith graph node	85
$w(u, v)$	Weight of the undirected edge connecting nodes u and v	44
x_i	ith component of vector $\mathbf{x}$	15
ξ	A pathway	81
$\mathbf{x}, \mathbf{y}, \mathbf{z}$	Vectors; dimensionality and meaning may vary depending on the context	15

*Otherwise known as an upper bound that is not asymptotically tight.