

Abstract of "Potential landscape perspectives on roaming: Insights on formaldehyde from geodesic paths" by D. Vale Cofer-Shabica, Ph.D., Brown University, May 2018.

Roaming is a novel and counter-intuitive mechanism of dissociation involving far-wandering fragments that return and abstract another component from the parent before dissociating. The mechanism is often an alternative route to closed-shell products in systems where both molecular and radical products are accessible and seems to be widespread. In the case of the photodissociation of formaldehyde (H_2CO), where roaming was first observed, a small, but non-trivial fraction of the molecular products are formed by roaming, bypassing the clear transition state that would otherwise be traversed. Barring a few reduced-degree-of-freedom examples where the dynamics can be treated analytically, most attempts at theories of roaming have entirely sidestepped the question of "why?" Transition state and associated theories take the inherently local perspective that the dynamics of a chemical reaction can be understood in terms of the various minima of a potential surface and transitions between them. But, such a perspective lacks explanatory power over the roaming mechanism. At issue is the manifest irrelevance of transition states. We suggest a global perspective instead: that roaming may be understood in terms of the geodesic (most efficient) pathways across the energy surface. To make such an analysis, we develop the techniques necessary to calculate and analyze geodesics on a small-molecule system, and then compute them on an established surface for formaldehyde. Even though roaming manifests on the flat, asymptotic region of the potential, the properties of geodesics in the region of the excited reactant well are quite informative. The distribution of lengths of geodesics leading to roaming is much broader than those leading directly to molecular or radical products. Indeed, the entropy of the roaming distributions is significantly higher. Measuring these entropies as a function of energy correctly predicts the decline in roaming as energy increases. It appears that the wider space of dynamical options available to the roaming mechanism gives it an entropic advantage over the other two channels. In light of these observations, we suggest that the presence of roaming is less about the existence of a localized "roaming transition state" and more about an entire region of the potential where the system's frustrated ability to navigate gives rise to multiple equivalent pathways.

Potential landscape perspectives on roaming: Insights on formaldehyde from geodesic
paths

by
D. Vale Cofer-Shabica
Sc. B., Brown University, 2009

A dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Chemistry at Brown University

Providence, Rhode Island
May 2018

© Copyright 2017, 2018 by D. Vale Cofer-Shabica

This dissertation by D. Vale Cofer-Shabica is accepted in its present form by
the Department of Chemistry as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date _____

Richard M. Stratt, Director

Recommended to the Graduate Council

Date _____

Peter M. Weber, Reader

Date _____

Brenda M. Rubenstein, Reader

Approved by the Graduate Council

Date _____

Andrew G. Campbell
Dean of the Graduate School

Vita

Dylan Vale Cofer-Shabica was born in Miami, Florida, in January of 1987. In addition to understanding and helping others to understand new things, he loves to bike, to cook, and to make music—most often by singing. He aspires to leave the world a better place than he found it.

Education

Ph.D. Physical Chemistry	Brown University, Providence, RI, 2012–2018
Sc.B. Chemical Physics	Brown University, Providence, RI, 2005–2009

Publications, Talks, & Posters

- D. V. Cofer-Shabica, R. M. Stratt. What is special about how roaming chemical reactions traverse their potential surfaces? Differences in geodesic paths between roaming and non-roaming events. *The Journal of Chemical Physics* **2017**, *146*, 214303
- J. M. Budarz, M. P. Minitti, D. V. Cofer-Shabica, B. Stankus, A. Kirrander, J. B. Hastings, P. M. Weber. Observation of femtosecond molecular dynamics via pump-probe gas phase x-ray scattering. *Journal of Physics B: Atomic Molecular and Optical Physics* **2016**, *49*
- D. V. Cofer-Shabica, R. M. Stratt, *The geometries of potential energy landscapes imply dynamical signatures for roaming reactions*, American Chemical Society, 250th National Meeting, Boston, MA, 2015, PHYS 554
- M. P. Minitti, J. M. Budarz, A. Kirrander, J. Robinson, T. J. Lane, D. Ratner, K. Saita, T. Northey, B. Stankus, V. Cofer-Shabica, J. Hastings, P. M. Weber. Toward structural

femtosecond chemical dynamics: Imaging chemistry in space and time. *Faraday Discussions* **2014**, *171*, 81–91
D. V. Cofer-Shabica, *Wandering Molecules*, Brown University, Research Matters, Providence, RI, **2014**. Invited Talk. <https://www.youtube.com/watch?v=X3xyMP9EAc0>

Awards

Brown University Department of Chemistry	Providence RI
Elaine Chase Award for Leadership and Service	2017
William T. King Prize for Teaching	2014
Teaching Fellow	2013, 2014, 2015

Service

Brown University	Providence, RI
Diversity and Inclusion Action Committee	2016–2018
WE Teach STEM Discussion Group	2015–2018
Stand Up for Graduate Student Employees	2013–2017
Blackstone Academy High School	Pawtucket, RI
Exhibition Night Judge	2013–2018

Teaching

Graduate Teaching Assistant	2012–2015
Brown University	Providence, RI
Math Teacher	2009–2011
Blackstone Academy Charter School	Pawtucket, RI
The Metropolitan Regional Career and Technical Center	Providence, RI

Affiliations

Member of the American Chemical Society	2015–Present
--	--------------

Acknowledgments

The completion of my doctorate marks the end of my second degree at Brown—the university for which I moved to and made a home of Providence, Rhode Island. This place is the only in which I have lived as an adult and, as such, it has shaped me greatly. More, even, than Brown or Providence, I have been shaped, mentored, loved, helped-along, and cared-for by many people. Without them I would not be writing nor would I have had the ability to complete my degree. To all those named below and to those who I do not name, I extend my great and heartfelt thanks.

My work with Richard Stratt made my graduate education. He drove me to render half-baked ideas concrete and to consider carefully the linkages in the chains of my arguments. Rich shared his commitments to teaching and, I hope, passed on some of the wisdom he has accumulated about how to tell a scientific story. How rare and how lucky to have a mentor who cares as much about word-choice and flow as I. I am grateful beyond measure for all that he has taught me.

All the members of Rich's group contributed to my education—providing me with ideas and with places to try out new thoughts. I am particularly indebted to Layne Frechette for his friendship and for his ideas during the early parts of my project.

The other members of my committee all played important parts in my development as scientist and teacher. Peter Weber took me in and employed me when I was wounded and wondering what I would do with myself. Brenda Rubenstein has been a much appreciated advocate and ally in the department. Matthew Zimmt listened and acted to make space for me when I told him of teaching assignments, "We can do better." Even though he passed away the year before I completed my degree, Jimmie Doll was a force throughout my education: he taught my first course in chemistry as well as my last three. I am grateful for his encouragement and his stories and his gentleness when I

asked him about results that were not even wrong.

The staff of the Chemistry department has done so much to make my decade-plus stay productive and warm. Ken Talbot and Randy Goulet gave my bike a place in the machine shop even after I became a theorist. Rose Barreira, Sheila Quigley, and Lynn Rossi all helped me put out fires in my personal and academic life. Margaret Doll, Carol Defeciani, and David Blair shepherded our cluster of computers. Ana Maria Fortes made sure that our offices were clean. I am especially grateful for my friendship with Eric Ferrara.

My relationships with all the members of Stand Up for Graduate Student Employees (SUGSE) have given me meaning and purpose during my time at Brown. Struggling with you to make Brown a better place to live and work for everyone has been an honor and privilege and taught me more about how to be in community and to work for change than I could have imagined.

I want to thank Kyle Trenshaw for his constant support and friendship, and for founding and inviting me to participate in WE-Teach-STEM—a much needed space for discussing the patriarchy and other systems of domination in the academy. I learned so much by listening.

Christopher Brook and Carolyn Trombi literally kept me in fighting shape and helped me hold to what I pass off as sanity in our bi-weekly Tae Kwon Do class. Your friendship has sustained me through many storms of the heart and mind.

The Providence Sacred Harp singers have given me community to call home and a place to sing. My soul soars with our voices. In addition to sharing songs, Jesse Polhemus, Laura Hodges, and Mike Richards have become close friends and confidants by sharing of themselves.

Tara Mulder and I loved each other; she supported me and made sure I completed my degree when I thought that I might leave. Cassie Houtz is an unexpected joy who fell into my life and made it possible to finish by feeding me bodily and emotionally.

I am immensely thankful for my family; without their encouragement I would be with nothing. My parents, Nancy and Stephen Cofer-Shabica, have supported me more than I can express since I left home. Knowing that I have their confidence is one of my deepest sources of strength. Molly Shabica and Rachel Shabica are each amazing sisters who taught me much and cheered me along the way. In the last three years of

my degree, my relationship with Betsy Cofer-Shabica, my youngest sister, grew-up and became one of mutual care. Thank you for always being ready for a phone-cry and for being a conduit to my playful and silly self.



This dissertation is dedicated to my grandparents. Grandpa was an organic chemist and a story-teller who fed my early love of science and of art. Grandma taught me to bake and gave me the voice with which I sing. Granny spoiled me rotten and knew my favorite dessert even as the penultimate of thirteen grandchildren. Granddaddy was a professor of geology, whose sharp mind continues to inspire and who has encouraged me to think broadly about how to play a part in the healing of the world. Thank you.

Contents

List of Tables	xii
List of Figures	xiii
1 Introduction	1
1.1 Organization of this document	2
1.2 Roaming and formaldehyde	3
1.2.1 Roaming observed in a broad class of systems	5
1.2.2 Previous theoretical efforts towards explaining roaming	7
1.3 Geodesics	9
1.3.1 An idea from condensed systems	10
1.3.2 Why geodesics for roaming formaldehyde?	11
2 Methods	13
2.1 Models	13
2.1.1 Formaldehyde	13
2.1.2 The potential energy landscape ensemble	16
2.2 Geodesic paths on a molecular energy surface	17
2.2.1 Center of mass conservation	22
2.2.2 Geodesic parameters	31
2.2.3 Trapped at a fixed point in 2 dimensions	34
2.3 Defining and sampling regions of interest	36
2.3.1 Monte Carlo for microcanonical reactants	37
2.3.2 Molecular dynamics for products	38
2.3.3 Monte Carlo random walk for the roaming region	39

2.3.4	Geodesic itineraries	42
2.4	Analyzing geodesics	43
2.4.1	Total lengths	43
2.4.2	Diatom Vibration	44
2.4.3	Diatom Rotation	45
2.4.4	Boundary fraction	46
2.5	Estimating geodesic lengths	47
2.5.1	Total Kinematic Length	47
2.5.2	Vibrational Kinematic Length	49
2.6	A search for saddles	50
2.6.1	Methods	50
2.7	Numerical methods	52
2.7.1	With respect to integration	52
2.7.2	Gradients	53
2.7.3	Hessian matrices & vibrational frequencies	54
2.8	Computing the entropy of a sampled distribution	58
3	Results and Discussion	60
3.1	An image of the formaldehyde potential	61
3.2	Locating a 2nd order saddle on the Bowman energy surface	63
3.2.1	Verification	63
3.2.2	Search results	64
3.2.3	No “roaming saddle” to be found	66
3.3	Observations from molecular dynamics	69
3.3.1	Bird’s nest	72
3.3.2	Molecular dynamics branching ratios	74
3.3.3	An estimate of the energy-dependent incubation period	78
3.3.4	Existence of high J_{CO} roamers	81
3.4	Geodesics are consistent with estimate	84
3.5	Geodesics are inherent dynamics	86
3.6	Observations from geodesics	90
3.6.1	Implications of the bird’s nest	91
3.6.2	Key observations at a single energy	92

3.6.3	The null effect of dynamical endpoint correlation	98
3.6.4	Trends with increasing landscape energy	100
3.6.5	Convergence of entropy of distributions with energy	108
3.7	Roaming as an entropically mediated phenomenon	112
4	Concluding remarks	113
4.1	New methods, new observations	113
4.2	Future directions	114
Bibliography		116
A	Derivation of path-integral form of the equation for free diffusion	125
A.1	Derivation	126
A.2	Unconditional minimization	129
B	Two more ways of preserving the center of mass during escape steps	133
B.1	Swatting a fly without moving: Follow the inverse-mass-weighted gradient	134
B.2	Laser-cannon: A general projection	137
C	Microcanonical sampling for the anharmonic oscillator	143
C.1	Normal Mode Hamiltonian	143
C.2	Construction of Normal Mode Coordinates	145
C.3	Corrections for Anharmonicity and Selection of Total Angular Momentum	147
D	Monte Carlo sampling the landscape ensemble	150
D.1	General requirements	151
D.2	Implementation specifics	152
E	Integrating externally-supplied FORTRAN code	157
E.1	Form of switching functions	164

- * Parts of this dissertation are reproduced from D. V. Cofer-Shabica, R. M. Stratt. What is special about how roaming chemical reactions traverse their potential surfaces? Differences in geodesic paths between roaming and non-roaming events. *The Journal of Chemical Physics* **2017**, *146*, 214303 with the permission of AIP Publishing.

List of Tables

2.1	Atomic masses	16
2.2	Comparison of step sizes in geodesic algorithm	32
2.3	Microcanonical and landscape energies	33
2.4	Roaming region constants	41
3.1	Properties of stationary points on the Bowman surface	65
3.2	Coordinates of second order saddle on Bowman's surface	67
3.3	Coordinates of global minimum on Bowman's surface	67
3.4	Coordinates of <i>cis</i> -HCOH minimum on Bowman's surface	67
3.5	Coordinates of <i>trans</i> -HCOH minimum on Bowman's surface	68
3.6	Coordinates of transition state for <i>cis-trans</i> isomerization	68
3.7	Coordinates of H ₂ CO → <i>trans</i> -HCOH transition state	68
3.8	Coordinates of tight transition state on Bowman's surface	68
3.9	Coordinates of proposed roaming saddle	69
3.10	Dispositions of molecular dynamics trajectories with energy	75
3.11	Fraction of roaming trajectories with high J_{CO}	83
3.12	Success rate of computed geodesics by channel and energy	91
3.13	Comparison to dynamically correlated (paired) geodesics	99
3.14	Mean total kinematic length for geodesics by channel and energy . . .	103
3.15	Portion of identically zero boundary fraction vs. energy	107
3.16	Mean boundary fraction vs. energy	107
3.17	Entropy of total length distributions vs. energy	108
D.1	Energy-specific acceptance ratios and seed configurations	156

List of Figures

1.1	Schematic representation of roaming	3
1.2	Geodesics on a simple potential energy surface	9
1.3	Geodesics on a simple potential energy surface	11
2.1	Energy scales for formaldehyde energy surface	15
2.2	The potential energy landscape ensemble	17
2.3	Free and escape steps wind around boundaries	20
2.4	Müller-Brown potential	35
2.5	Fixed point on the Müller-Brown potential	36
2.6	Hessian parameters	57
3.1	A projection of the potential surface	62
3.2	Orientation to the H ₂ CO surface	71
3.3	Six roaming trajectories	73
3.4	Molecular dynamics channels (reactive)	76
3.5	Molecular dynamics channels	77
3.6	Rates of reaction	79
3.7	Failure of statistical distribution of reaction times	80
3.8	A “cold” shoulder in the J_{CO} distribution	82
3.9	Roamers exhibiting high J_{CO}	84
3.10	Distribution of lengths over full geodesic paths	85
3.11	Inherent dynamics pictograph	87
3.12	Geodesic as inherent dynamics of formaldehyde	89
3.13	Total length distributions	94
3.14	Component length distributions	95

3.15	Boundary fraction	97
3.16	Total length distributions in the high-energy limit	101
3.17	Convergence of total length distributions	102
3.18	Component length distributions in the high-energy limit	104
3.19	Boundary fraction in the high-energy limit	105
3.20	Convergence of distribution of boundary fractions	106
3.21	Entropy of total length distributions	110
3.22	Entropy of component length distributions	111
A.1	Discretizing a Path	127
B.1	A projection	138

CHAPTER 1

INTRODUCTION

Roaming is a novel reaction mechanism [6] characterized by fragments that take long excursions, far from the reacting center before returning and dissociating. This strange dynamical feature turns out to occur in a large number of systems [7] and has generated considerable experimental and theoretical effort around the detection and explanation of the phenomenon in the last decade. Formaldehyde, H₂CO—being one of the simplest systems to exhibit roaming and being the first known to do so—provides us with an attractive arena within which to understand this novel dynamical phenomenon.

The geodesic formalism [8] is a tool to understand the *inherent dynamics* of a classical system by characterizing the most efficient pathways through the energy landscape. The inherent dynamics of a system are the dominant, most important modes of motion—instead of describing the fine details, the inherent dynamics give the overall behaviors and sequencing of events. Likewise, geodesics, rather than being concerned with the enumeration of minima and transition states between them, provide a global view of the potential surface.

As theorists, one of our primary tasks is to tell stories that provide durable intuition to those who would understand or reason about the systems we study. Oftentimes, the plots of those stories are driven by experimental observations and/or insights gained through analytic equations or numeric calculations. The work presented here is another attempt to tell the story of roaming as it manifests in formaldehyde. Experiments by van Zee and co-workers [9] first hinted at a hole in our understanding of formaldehyde’s photodissociation dynamics; extant theories of reaction dynamics, such as transition

state theory [10], neither account for their observations nor enable us to reason about the roaming mechanism. We use the theoretical framework provided by geodesics to make a new window into the inherent dynamics of roaming in formaldehyde.

1.1 Organization of this document

In the thesis that follows, we make a thorough investigation of roaming in the context of formaldehyde using geodesics as our primary lens. The present chapter resumes by providing background on roaming in the context of formaldehyde and other notable systems. We also survey the various theoretical thrusts towards understanding roaming in general. Moving on, we then discuss the group’s development of geodesics as a tool for understanding the inherent dynamics of condensed systems and make an argument for why geodesics in a small-molecule system might be applicable to the problem of roaming.

Chapter 2 details our methods of study and analysis, beginning with the models we use and assumptions we make about formaldehyde and roaming. We then move to a description of geodesics in the context of a small molecule, discussing new wrinkles and our resolutions to them. In Chapter 2, we also describe the molecular dynamics and Monte Carlo methods used to sample regions of the formaldehyde potential. The chapter concludes with a description of several of the numerical methods used throughout our studies.

In Chapter 3, we present and discuss our major results. We begin by looking through the frame of stationary points and molecular dynamics, but quickly exhaust their utility and move on to geodesics. Our discussion first takes up a single energy, as in our paper [1], and then moves to an energy-dependent investigation. Our observations lead us to conclude that the roaming mechanism is an *entropically* mediated phenomenon in formaldehyde. We discuss the implications of this conclusion.

The final chapter, 4, is summative. We recapitulate the major findings as well as the technical developments that enable our study. We also make suggestions about future directions for research.

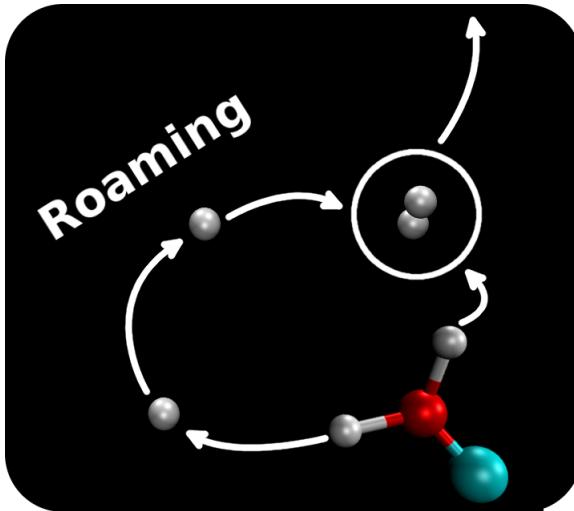


Figure 1.1: Schematic representation of roaming

Roaming is a route to products that involves a fragment taking long excursions from a dissociating parent before *returning* and abstracting another component (often, but not exclusively, an atom). In the case of formaldehyde, roaming provides another route to close-shelled molecular products when excited near the radical dissociation threshold. A single hydrogen may “roam” up to several angstroms away from the formyl parent for several picoseconds before returning and abstracting another hydrogen: $\text{H}_2\text{CO} \longrightarrow \text{HCO}\cdots\text{H} \longrightarrow \text{H}_2 + \text{CO}$.

1.2 Roaming and formaldehyde

Because of its relative simplicity, ease of handling, and relevance to fundamental reaction mechanics, formaldehyde, H_2CO , has served as a test bed for theorists and experimentalists interested in fragmentation dynamics for many years; indeed by the early 1980s, formaldehyde was already considered a well beaten horse [11]. The great amount of scrutiny to which the four-atom molecule has been subjected makes it perhaps surprising that a totally unknown reaction mechanism lay hidden amongst its dynamics¹ for so long. In the mid 1990s, van Zee and co-workers presaged roaming after they observed a strange shift as a function of energy in the distribution of the product rotational states of CO when photolyzing H_2CO under collisionless conditions near

¹We will address the following point later in this chapter, but it is worth indicating now that in most of the experiments treating formaldehyde, H_2CO is excited to its first electronic state, S_1 , which then non-radiatively decays to the ground state, S_0 , where vibrationally hot, close-shelled products are formed.

the radical dissociation threshold [9]. Dissociation via a slightly energetically *higher* vibrational state, yielded a fraction of substantially *slower* CO rotation. The “cold shoulder” that appeared in the distribution of J_{CO} product states when exciting near the radical threshold suggested to them the opening of a second channel to closed-shell molecular products.

It was not until over a decade later, in a joint theoretical-experimental effort, that Townsend and colleagues [6] revealed how bizarre the origin of the low-energy bump in van Zee’s distribution was. Near the radical dissociation threshold, one of the hydrogens may be ejected from the parent, but not actually depart. This roaming hydrogen may linger several angstrom’s away for up to several picoseconds before returning and abstracting another hydrogen. Figure 1.1 shows this schematically. In departing by this route, the H₂ is typically highly vibrationally excited and, as a consequence of energy conservation, the CO is left with much less internal energy. Because the excited vibronic states of CO are not heavily populated, most of the energy comes from rotation; hence the greater population in low J_{CO} states.

The authors of that first paper characterizing roaming [6] immediately recognized that the phenomenon was not at all mediated by the well-known tight transition state to molecular products². Aside from the great strangeness of the mechanism, a major piece of evidence was that molecular dynamics studies initiating trajectories from the transition state to molecular products revealed no evidence of roaming [13, 14]. If roaming did proceed via the transition state, surely these studies would have encountered the mechanism.

A note about energy surfaces and scales: After excitation to the first excited electronic state, H₂CO decomposes into either molecular (scheme 1.1) or radical products (scheme 1.3). Molecular products are formed on the electronic ground state, S_0 , and radical products on the S_0 and T_1 (first excited triplet) states [15]. Roaming to molecular products (scheme 1.2) occurs on the S_0 surface [6]. Roaming, however, is not the primary channel for the formation of molecular products in formaldehyde. Estimates range from 15% to 40% as a function of energy [16]; but, roaming remains an ill-defined

²We refer to a “tight” transition state in the sense that it involves the formation of an activated complex with an energetic barrier [12].

process and these estimates should be taken as order-of-magnitude only. Experimentally, the onset of roaming is somewhere between 100 cm^{-1} and 200 cm^{-1} below the radical dissociation threshold [17]. The relevant experimental energy scales³ for processes on the S_0 surface are given here:



Townsend and colleagues [6] demonstrated that roaming in formaldehyde occurs on a single ground electronic state under classical dynamics and microcanonical initial conditions with total angular momentum of 0. Therefore, even though quantum effects surely play a role in dynamics of energy transfer leading up to roaming, roaming can be understood from a purely classical perspective. This classical viewpoint will be our frame of reference through the investigation.

1.2.1 Roaming observed in a broad class of systems

While roaming was first observed and characterized in formaldehyde, researchers quickly sought and found many other systems where roaming occurs [7]; within the first decade of observation, over 20 systems were reported to roam [19]. Many early investigations identified roaming by multi-component product state distributions as in formaldehyde. Out of all the systems, a collection of characteristics of roaming fragments came together: large amplitude, wandering motion on a flat region of potential and time-scale separation between that motion and the other vibrations of the system. Roaming has been reported in a wide range of unimolecular as well as bimolecular reactions. Here we briefly describe a few of the systems exhibiting roaming to give a sense of the great plurality of systems that roam.

Acetaldehyde, CH_3HCO —which is simply formaldehyde with a hydrogen replaced by a methyl group, CH_3 —was a natural second place to find roaming [20]. Here the relatively massive methyl group roams away from the formyl before abstracting another

³After this chapter, our investigation will be exclusively based on a high-accuracy, full-configuration-space, model potential for formaldehyde. The experimental values given here are connected to energies on that surface by the addition of the zero point energy; see Section 2.1.1.

hydrogen (though rarely a hydrogen will roam and abstract the methyl). Roaming seems to be the *dominant* channel for formation of molecular products in acetaldehyde and can account for as much as 80% of the flux [21].

Roaming on multiple electronic surfaces was first reported on the D_0 and D_1 surfaces of NO_3 [22, 23, 24]. These authors also report that the $\text{NO}_3 \longrightarrow \text{NO} + \text{O}_2$ reaction proceeds *exclusively* by roaming [23].

It seems that many roaming systems also feature radical-radical abstraction pathways [7] and at least one author has speculated that wherever there exists such a barrierless route to abstraction leading to excited, closed-shell products, roaming will be present too [25]. This idea was indeed borne out in a large number of $\text{H} + \text{C}_n\text{H}_{2n+1}$ alkane reactions [26].

Even though roaming can be described in purely classical terms, much speculation has been made about the quantum manifestations of roaming. There are not enough studies to address this point in generality, but the roaming mediated reaction, $\text{H} + \text{MgH} \longrightarrow \text{Mg} + \text{H}_2$ offers at least one finding: Just below break-up energy of the complex, the vibrational wave functions of bound states seemed to be highly delocalized [27]. This is not at all surprising given the great variety of roaming pathways seen in other systems [17].

“Roaming” or something like it has even been reported in the liquid phase [28]. Mereshchenko *et al.* claim that CHBr_3 exhibits a roaming-mediated isomerization to $\text{BrHCB}-\text{Br}$ while solvated. In this system, “roaming” occurs on a 100 fs time-scale, with large amplitude motion and on a flat region of the potential energy surface. While 100 fs is an order of magnitude faster than roaming in most of the other systems described here, the other properties: flat potential and unpredictable wandering are certainly consistent.

Roaming has been presented as a failure of transition state theory and a challenge to our understanding of reaction dynamics in general [7, 19]. Indeed roaming does present such an opportunity and has rightly inspired considerable interest in a theoretical framework for understanding the mechanism—the present work is no exception. However, we would be remiss if we failed to point out that roaming is only another example in a long history of systems the dynamics of which are not well understood in terms of barriers and extrema [25, 29]. We suspect that a broader perspective on the

energy landscape is necessary to provide insight into the dynamics of these systems.

1.2.2 Previous theoretical efforts towards explaining roaming

Previous attempts to provide a theoretical explanation for roaming can be broadly divided into two camps based on whether they take a local or global perspective on the configuration- or phase-space. Local descriptions often rely on the identification of a reaction coordinate and/or a dividing surface and then make some kind of transition state calculation to compute a rate or branching ratio. Global methods seek to apprehend the large scale (phase- or configuration-space) structures that guide the dynamics of roaming. In this section we will give a brief exposition of the major ideas advanced in each camp.

“Roaming transition states”

Klippenstein, Harding, and colleagues have done considerable work to advance a statistical, transition state theory of roaming [30]. They were the first group to identify a “roaming” transition state, or a saddle point resembling roaming, in formaldehyde and acetaldehyde [31]. In both cases, the saddles were located 30–70 cm⁻¹ below the radical dissociation threshold. They present a multi-path transition state theory with transition states constructed to take into account dynamical effects as well as the features of the energy surface. Their predictions compare favorably to reduced dimension trajectories in acetaldehyde [32] at energies less than 350 cm⁻¹ above the radical asymptote, but suffer substantially above this limit, where roaming is still prevalent [21, 33].

In the same family of theories, Andrews, Kable and Jordan [34] present a two parameter model coupled to a variational transition state theory calculation that successfully fits roaming vs. radical branching ratios over a narrow range of energies for H₂CO, CH₃HCO, and NO₃. Unfortunately, they are not able to speak to the portion of the molecular channel arising from the tight transition state or provide much insight into how to think about the mechanism.

The Bowman group has also weighed in on the subject of roaming saddles [35] and contends that, at least in the aldehydes studied, the saddles are insufficiently

energetically separated from the tight transition state to be treated independently in a statistical theory (this judgment is disputed [36]). They also have observed [19] that in all attempts to define intrinsic reaction coordinates [37] by way of roaming saddles [38, 39], the dynamics show massive deviations from these paths, euphemistically called “large amplitude motion” in the vicinity of the saddles. Such deviations are perhaps at the root of the terminally frustrated attempts at a statistical theory for the prediction of branching ratios.

Phase space structures governing roaming dynamics

In a recent series of papers [40, 41, 42, 43, 44], the Bristol-Cornell-Crete collaboration offers a rather different perspective on roaming in the language of non-linear dynamical systems theory. In a variety of model systems, the authors identify phase-space structures that guide the dynamics of roaming. In their scheme, roaming trajectories are given their bounds by phase-space dividing surface constructed by computing the normally hyperbolic invariant manifolds of the space [41]. They also identify periodic orbits—which function like attractors—and suggest that the origins of roaming behavior are in a family of periodic orbits within what they identify as the roaming region of phase-space. Of particular interest is the authors’ observation of the fractal nature of phase-space bands of initial conditions leading to different products, which suggests that the subspace leading to roaming is highly variegated [42].

The collaboration also identifies these periodic orbits and unstable manifolds in a reduced-dimensional model of formaldehyde [43]. They propose that there exists a particular periodic orbit associated with roaming that bounds all roaming (but not radical) trajectories and guides them back to molecular products [43, 44]. While the results they achieve do much to address the “whys” of roaming, their approach is unfortunately limited to two degree of freedom systems because of technical (though not insurmountable) hurdles to computing normally hyperbolic invariant manifolds in larger systems.

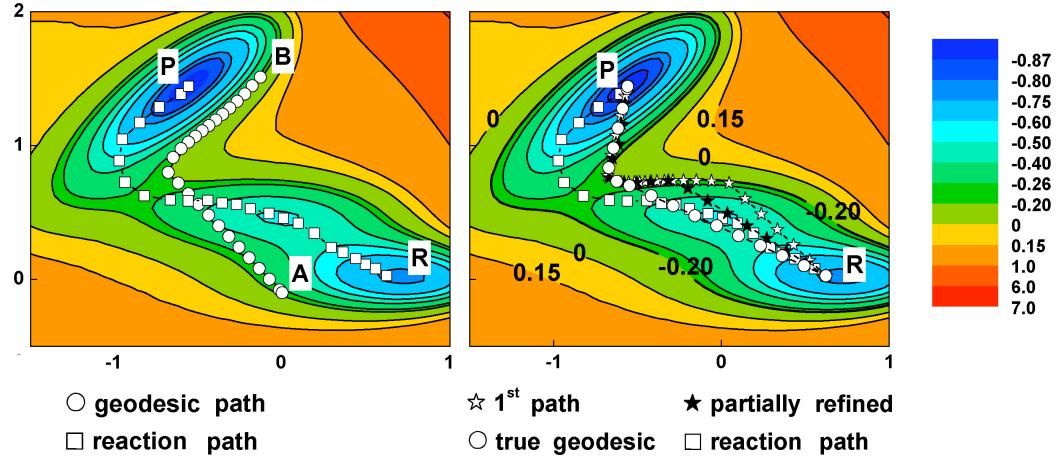


Figure 1.2: Geodesics on a simple potential energy surface

In the figures above, **R** and **P** correspond respectively to the reactant and product wells of the Müller-Brown potential [45]. A steepest-descent reaction path connecting them is shown in the squares (\square). We also show geodesic paths, circles (\circ), between the reactants and products (right) and points labeled **A** and **B** (left). In addition to showing how the geodesic path differs from the reaction path, the figure on the right also shows a geodesic in several stages of refinement using the algorithms discussed in Section 2.2. Figure modified from [8].

1.3 Geodesics

Similar in spirit to the global methods employed by the Bristol-Cornell-Crete collaboration, our geodesic formalism [8, 46] is a means for studying the large-scale structures of a dynamical system. We seek to understand the *inherent dynamics*—the most important components of the dynamics, smoothed and stripped of noise. Geodesics are decidedly *not* concerned with the basins, transition states, or stationary points of a potential surface, but with the properties of the major thoroughfares by which a system traverses configuration space⁴. Figure 1.2 shows the contrast between a geodesic path and a steepest-descent reaction path connecting the basins.

We will formalize our definition of geodesics later, in Chapter 2, but looking ahead, our notion of *shortest paths* (geodesics) as *most important* paths comes from a formal

⁴That geodesics are constructed entirely in the space of configurations rather than the phase-space of the system is another distinction between our methods and those of the Bristol-Cornell-Crete collaboration. We are still able to pose fruitful questions of the potential surface so long as we probe the properties of paths rather than of points.

rendering of a solution to the Fokker-Planck equation for free diffusion as a path integral:

$$G(\underline{R}_0 \rightarrow \underline{R}_f, t) = \int_{\underline{R}_0, 0}^{\underline{R}_f, t} \mathcal{D}[\underline{R}(\tau)] \exp \left[\frac{-1}{4D} \int_0^t \left(\frac{d\underline{R}}{d\tau} \right)^2 d\tau \right] \quad (1.4)$$

$G(\underline{R}_0 \rightarrow \underline{R}_f, t)$ is the probability density for diffusing from \underline{R}_0 to \underline{R}_f in a time t and $\mathcal{D}[\underline{R}(\tau)]$ indicates an integral over the space of all possible *paths*, $\underline{R}(\tau)$. D is the diffusion constant. The insight comes from the integral term in the exponential. When diffusion is slow, D is small and the only paths that will contribute are those that minimize the integral over $(d\underline{R}/d\tau)^2$, i.e., those that minimize the force-free action. Those same paths are also the shortest paths, or geodesics of the system. There is a quantitative connection between the rate of diffusion in a system and the length of its geodesics; as they elongate, the dynamics slow down [8].

Geodesic theory is related conceptually and in aim to transition path theory [47], a scheme for understanding rare-events in potential systems by deriving average trajectory flows. The techniques provide a prescription for constructing, among other things, the “current” of reactive trajectories between regions of configuration space. The flow lines of such currents are filtered objects, with unimportant motions washed out, much like our geodesics and notion of inherent dynamics. In contrast, however, where the methods of transition path theory are difficult to apply to systems with many degrees of freedom, there are simple algorithms [8] for computing geodesics in systems with 2 to 1000s of degrees of freedom.

1.3.1 An idea from condensed systems

The initial ideas for geodesics and our notions of most efficient paths were developed in the context of condensed systems [8, 46, 49, 50, 51, 52]. Geodesics have enabled a number of remarkable successes including quantitative prediction of the diffusion-percolation transition [8], revealing the underlying dynamics of preferential solvation in a binary mixture of competing solvents [49], and, more recently, providing insight into the glass transition in hard spheres [51].

What all of these phenomena have in common is that they manifest in systems where small variations in their energy landscapes are unimportant. Indeed, in the extreme case of hard spheres, the potential has no fine variation—it only takes on two values: 0 or ∞ .

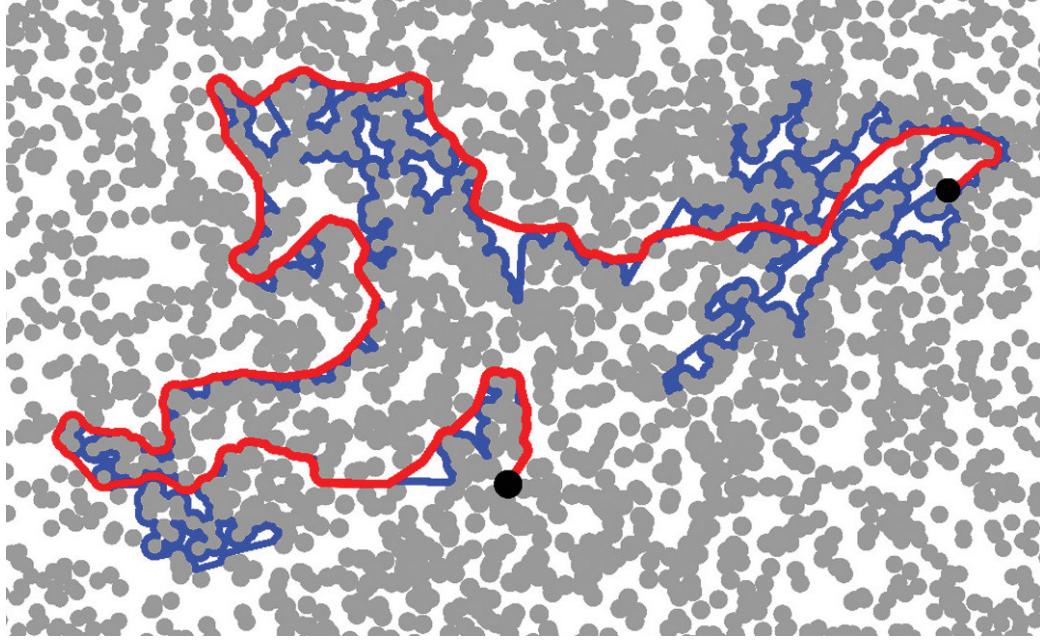


Figure 1.3: Geodesics on a simple potential energy surface

Geodesic path (solid red) between endpoints (filled, black circles) in a simple model of a liquid. In this two-dimensional example, it is apparent how geodesics are like the shortest-path solutions to a maze. Figure modified from [48].

Even in the case of ordinary liquids at room temperature, the well-depths of most of the minima are far below the few $k_B T$ per degree of freedom ($\sim 200 \text{ cm}^{-1}$) available to the system. Rather than being mediated by gathering enough energy into a special coordinate and surmounting a barrier, these processes are driven by the system's ability to navigate a disordered landscape (see Figure 1.3). These entropic considerations require us to probe the global structure of energy landscape instead of searching for special stationary points.

1.3.2 Why geodesics for roaming formaldehyde?

Much like the perspective taken by the members of the Bristol-Cornell-Crete collaboration [43, 44], we are firmly of the view that understanding the peculiar dynamics of roaming will not come from the properties of stationary points on the energy surface. We also have a different goal than many of the other theoretical projects in Section 1.2.2. Rather than fiddling with branching ratios, we seek to understand the essential properties

of the formaldehyde energy landscape that give rise to roaming.

In deploying geodesics to understand roaming, we implicitly assume that the minor variations in the potential are not central to the dynamics—indeed, we take the minima to be irrelevant. Our perspective is in stark contrast to the basin-driven lens offered by Wigner’s transitions state theory [10] and even Wales’ network-of-basins view on energy landscapes [53].

We hypothesize that such disregard for local features of the surface is not only appropriate, but essential because photolysis injects such enormous amounts of energy into the system. For many regions of the landscape, fluctuations in the potential will be small in comparison to the kinetic energy—the system will skate high above any dips or rises. Given this great energetic excess, it would be impossible to pin down the backbone of the dynamics in terms of just minima and barriers. In the particular case of roaming, the potential in the region of the landscape where roaming occurs is quite flat and the roaming hydrogen strays far from the reacting center. The global exploration of the landscape suggested by its wanderings is just the kind of problem to which geodesics are suited.

Another way of distinguishing our perspective is by thinking in terms of low-energy versus high-energy limits on the energy landscape. In the low-energy limit, it makes sense to prioritize stationary points because excursions away from them are rare. In contrast, in the high-energy limit, the system’s dynamics will be driven by global connectivity, which renders the details of stationary points unimportant. In the case of roaming formaldehyde, we are clearly in the high energy limit and wide exploration of the space is called for; the holistic mode of analysis offered by geodesics, insensitive to the details of the potential, fits the problem well.

In the next chapter, we develop the tools to unambiguously specify geodesics on a small-molecule potential surface. We discuss the landscape ensemble, the space within which geodesics are defined, in Section 2.1.2, and then give a description of how they are computed in Section 2.2.

CHAPTER 2

METHODS

In this chapter we describe our models and the assumptions that underlie them. We then proceed to a description of our algorithms, sampling procedures, and other methods. The initial sections deal with computing geodesics on a small-molecule potential energy surface. This task is related to but different from previous work done by the group in several key ways. The remainder treat the search for a saddle on such a surface, numeric methods, and a technique for estimating the entropy of an unknown distribution.

2.1 Models

2.1.1 Formaldehyde

To treat roaming in the photodissociation of Formaldehyde (H_2CO), we use a single classical, ground-state potential surface. This is reasonable at the energy scales of interest because formaldehyde dissociates on the ground state after excitation to and relaxation from the first excited singlet state [15].

The potential surface we use was generously provided to us by Joel Bowman, of Emory University and is described in detail by Zhang *et al.* [54]. A summary of the development and features of the surface follows. The surface is global in that it covers the space of all configurations at energies from the ground state all the way up to triple fragmentation ($\text{H}_2\text{CO} \longrightarrow \text{H} + \text{H} + \text{CO}$). In addition to explicitly treating molecular dissociation, the authors constructed it with an eye towards studying the abstraction

reaction $\text{H} + \text{HCO} \longrightarrow \text{H}_2 + \text{CO}$; this because of its importance to roaming. The surface is assembled from independently constructed fits to 6 different regions of configuration space in internal coordinates. Regions expected to be well described by single reference methods were computed with coupled clusters [CCSD(T)]. Where these were expected to fail, multi-reference, configuration-interaction (MR-CI) with complete active space, self-consistent field (CASSCF) methods were used (with 6 active orbitals and 6 active electrons). Roughly 170,000 single point energy calculations went into the fits, which were then joined by quintic switching functions¹. Fits were validated against the *ab initio* calculations, adjusted, and generally found to be in quite good agreement. The authors note that the surface is suitable for dynamics studies, but not for those demanding “spectroscopic accuracy” because of generally modest energetic deviations (there was, however, a large, 3.23%, one at a single stationary point) and discrepancies in computed frequencies. Likewise, the surface is sufficient for our purposes.

The most important property of the surface is that it exhibits roaming; indeed it is this surface that first contributed to the discovery of roaming by Townsend and co-workers [6]. As part of our validation of the potential, we implemented the molecular dynamics scheme described in Section 2.3.2 and successfully reproduced the product distributions reported in their paper.

That roaming is exhibited on this surface is a point not to be minimized or underestimated. As our investigation targets the inherent dynamics of roaming and photodissociation in formaldehyde, we do not need a perfectly accurate potential. We require only one that has the essential features that give rise to roaming. Classical treatment of the Bowman surface meets this criteria. While quantum effects may play a role in the photodissociation of formaldehyde (and they certainly do guide energy flow and deposition through the electronic states), we know that roaming manifests classically on this surface and so we pursue a purely classical treatment.

For the purposes of making contact with experimental energies, the zero-point energy of formaldehyde must be added to the energy of any excitation laser. Zhang and co-workers give a harmonic estimate of the zero-point energy [55] as:

$$E_{\text{zpe}} = 5844 \text{ cm}^{-1} \quad (2.1)$$

The potential is defined in a 6-degree-of-freedom space of inter-atomic distances, but

¹The published form deviates from that used in their source code; see Appendix E.1.

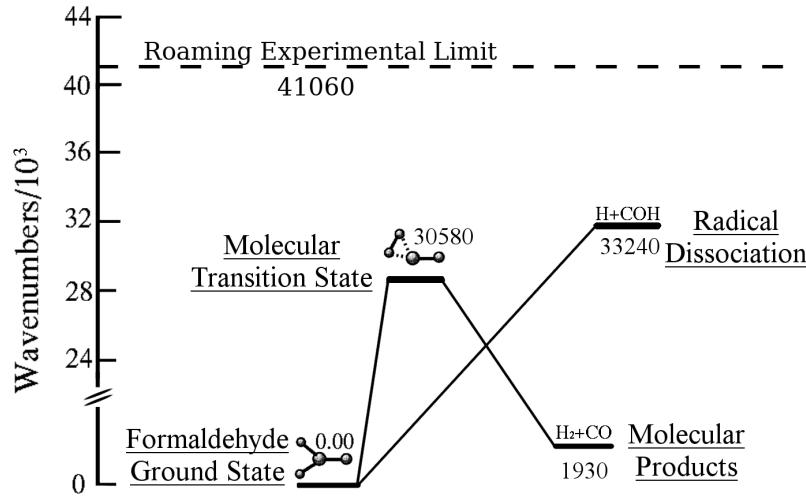


Figure 2.1: Energy scales for formaldehyde energy surface

This scale is for the fitted energy surface used in our investigations. Plotted also is the experimental limit of detection for roaming. All energies are given in wavenumbers. Figure modified from [54].

all of our simulations (and path-finding routines) are carried out in the 12-dimensional, Cartesian lab frame. A detailed description of the way we interface with the FORTRAN code provided by the Bowman group is in Appendix E.

The relevant energy scales for the problem and the energy surface are given in figure 2.1. Our investigations probed energies from 35000 cm^{-1} to 45000 cm^{-1} , corresponding to excitation wavelengths² between 349 nm and 255 nm; see Table 2.3 for all energies used. This range starts just above the radical dissociation threshold ($E_{\mu}^{\text{rad}} > 33240 \text{ cm}^{-1}$) allows us to compare the three channels of interest: direct dissociation to molecular products, roaming dissociation to molecular products, and radical dissociation. Roaming is not observed experimentally above 41060 cm^{-1} [17].

A natural system of units for the study of phenomena in the energy range of electronic excitations is that of Hartree atomic units and we use them throughout our investigation. The values of the electron's rest mass, m_e , its charge, e , and the reduced Planck constant, \hbar , are all set to unity. The fundamental unit of length is the Bohr radius, $a_0 = 0.529 \text{ \AA}$;

²The upper end of our energy range is non-physical as the same electronic pathways are not accessed at this wavelength. We can, however, make useful comparison between molecular dynamics and geodesic results on this surface even at experimentally inaccessible energies.

Element	Mass / m_e^*
H	1837.15
C	21874.66
O	29156.95

Table 2.1: Atomic masses

* m_e : rest mass of the electron

the unit of energy is the Hartree, $E_H = 27.211 \text{ eV} = 2.1947 \times 10^5 \text{ cm}^{-1}$; and time passes in units of $\hbar/E_H = 0.0242 \text{ fs}$. The masses of the atoms relevant to the simulation are given in Table 2.1.

2.1.2 The potential energy landscape ensemble

What we call the *inherent dynamics* of a system can be thought of as the system's classical dynamics stripped of noise (e.g. high frequency vibrations or rotations) that are not essential to the motions of a system. They describe the dominant thoroughfares of configuration space—the major pathways followed during a chemical or physical transformation. In a sense, the inherent dynamics are the most efficient routes through configuration space. To understand the nature of roaming, we take formaldehyde as a model system and examine the topology and connectivity of its energy landscape. There are certainly many ways one could study the space of allowed paths, but a conceptually simple route is a binary partition of configuration space into allowed and disallowed regions: those where the high energy of photo-excitation makes the variegation of the potential irrelevant and those where the potential is so large as to be practically inaccessible. We do this formally by introducing the *potential energy landscape ensemble* [46].

The landscape ensemble is a statistical mechanical ensemble like the canonical or microcanonical ensembles. However, instead of a fixed temperature or energy, its thermodynamic variable of control is a fixed maximum potential energy, called the landscape energy, E_L . Within this ensemble, all configurations such that the potential is less than the landscape energy,

$$V(\mathcal{R}) \leq E_L \quad (2.2)$$

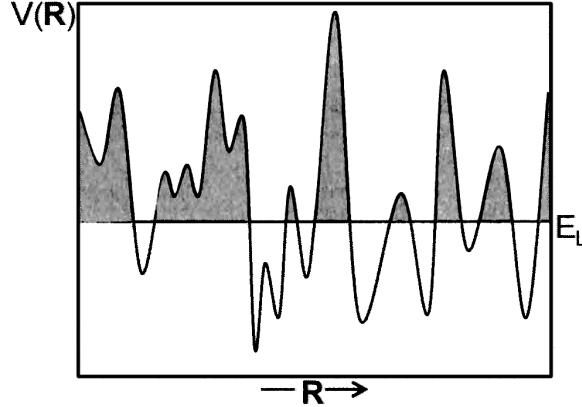


Figure 2.2: The potential energy landscape ensemble

The landscape energy, E_L , provides the thermodynamic variable of control in the landscape ensemble. This fixed maximum potential energy plays a similar role to the microcanonical energy or the canonical temperature. In the figure, clear valleys are accessible, $V(\underline{R}) \leq E_L$, while the shaded mountains are forbidden. Figure from [46].

are admitted with equal probability; see Figure 2.2. Equiprobability is a result of the same entropy maximization arguments that lead to equiprobable states in the microcanonical ensemble. In the thermodynamic limit, the landscape ensemble gives the same configuration space probability densities any other ensemble (e.g. micro-canonical, canonical) gives. In some ways, this binary division is the simplest partitioning that one could make given a space with a scalar field on it³. Such a divided space gives us a frame within which to locate the most efficient paths.

2.2 Geodesic paths on a molecular energy surface

The group's initial development of the ideas behind efficient paths occurred in the context of diffusing liquids [8]. Writing the solution for the Fokker-Planck equation for free diffusion as a path integral⁴ provides our guiding principle.

$$G(\underline{R}_0 \rightarrow \underline{R}_f, t) = \int_{\underline{R}_0, 0}^{\underline{R}_f, t} \mathcal{D}[\underline{R}(\tau)] \exp \left[\frac{-1}{4D} \int_0^t T(\underline{R}(\tau)) d\tau \right] \quad (2.3)$$

³Rather, it is the simplest partitioning of a scalar field that leaves us with any chance of the space remaining connected. We could have also demanded $V(\underline{R}) = E_L$, but this space would have volume of zero measure and would surely be hopelessly disconnected for all but the simplest potentials.

⁴See Appendix A for a full derivation.

Here, $G(\underline{R}_0 \rightarrow \underline{R}_f, t)$ is the probability density for diffusing from \underline{R}_0 to \underline{R}_f in a time t and $\mathcal{D}[\underline{R}(\tau)]$ indicates an integral over the space of all possible paths, $\underline{R}(\tau)$, respecting whatever boundary conditions we impose. T is the kinetic energy and D is the diffusion constant.

We examined and were then motivated by eq. 2.3 with the goal of understanding the *inherent dynamics* or most efficient paths of systems under study. There are many ways that we could formulate a notion of most efficient path, but an expedient one comes from combining eq. 2.3 with the potential energy landscape ensemble discussed in Section 2.1.2.

The landscape ensemble partitions configuration space into allowed and dis-allowed regions and provides boundary conditions for eq. 2.3. When diffusion is slow (D is small), the paths that dominate are those that minimize the force-free action in the exponential.

$$S[\underline{R}(\tau)] = \int_0^t d\tau T(\underline{R}(\tau)) \quad (2.4)$$

We require that we this functional be minimized in the space of paths respecting the landscape criterion, where $V(\underline{R}(\tau)) \leq E_L$ for all τ . These same paths minimizing the action are also *shortest* paths or geodesics, which minimize the kinematic length (to be introduced in Section 2.4.1):

$$\ell = \int_0^t d\tau [2T]^{1/2} \quad (2.5)$$

Note the dimension of T is [Energy] and so the integral in eq. 2.5 has dimension of [Length] \times [Mass] $^{1/2}$. The paths that minimize the functional are therefore technically mass-weighted shortest paths. We will confront this distinction in Section 2.2.1.

From the landscape perspective, the dynamics of a system are not classical paths in phase space, but much simpler objects, geodesic paths in configuration space. These geodesics provide information about the connectivity and structure of the energy landscape. This is our main result: that the inherent dynamics are given by geodesics in landscape ensemble.

Casting the diffusion equation in the path-integral manner of eq. 2.3 is so powerful because it frees us from having to solve what is otherwise a horribly complicated and specific problem—diffusion over a potential is a different (possibly insoluble) differential equation for every potential. We integrate over all possible paths, composed of many

infinitesimally short steps. Because the potential does not vary dramatically over such *short* scales, we can use the equation for free diffusion. But, the impact of the potential does not go away; its presence is encoded in the, perhaps quite stringent, boundary conditions imposed by the landscape criterion, $V(\underline{R}) \leq E_L$.

The condition in eq. 2.4 has form of a non-holonomic constraint.⁵ Finding solutions to problems with non-holonomic constraints is often difficult, but the form of such solutions is specified by the Kuhn-Tucker theorem [56]. The theorem holds that the desired solution will be the union of paths that unconditionally minimize the functional in eq. 2.4 and those where strict equality holds. The Kuhn-Tucker theorem points to the need for two components of an algorithm for computing geodesics: free-steps that unconditionally minimize the path length and steps that trace along the boundary of the obstacles defined by $V(\underline{R}) = E_L$ contours. For an N atom system, the kinetic energy can be written in terms of the lab-frame, Cartesian coordinates for the centers of each atom:

$$T = \frac{1}{2} \dot{\underline{R}}^T \mathbf{M} \dot{\underline{R}} \quad (2.6)$$

Where $\underline{R}^T = (x_1, y_1, z_1, x_2, y_2, \dots, z_N)$ is the total configuration-space vector with $\dot{\underline{R}}$ its time derivative, and \mathbf{M} is the mass matrix, given by:

$$(\mathbf{M})_{i\mu,j\nu} = \delta_{i,j}\delta_{\mu,\nu}m_i \quad i, j = 1 \dots N \quad ; \quad \mu, \nu = x, y, z \quad (2.7)$$

With m_i is the mass of the i th particle in the N -particle system.

The path between a target vector, \underline{R}_f , and a starting position, \underline{R}_0 , that unconditionally minimizes eq. 2.4 in time t for such a kinetic energy is then the straight-line path given by:

$$\underline{R}(\tau) = \left(1 - \frac{\tau}{t}\right) \underline{R}_0 + \left(\frac{\tau}{t}\right) \underline{R}_f \quad (2.8)$$

Discretizing this path is straight-forward⁶:

$$\underline{R}^{i+1} = \underline{R}^i + \delta_s \frac{\underline{R}_f - \underline{R}^i}{|\underline{R}_f - \underline{R}^i|} \quad (2.9)$$

⁵A *holonomic* constraint is one that can be expressed as an *equality* relationship between a constant and a function of the parameters of its domain, e.g. coordinates and time, as in: $0 = f(q_1, q_2, \dots, t)$.

⁶Consider $\underline{R}(\tau + \Delta\tau)$

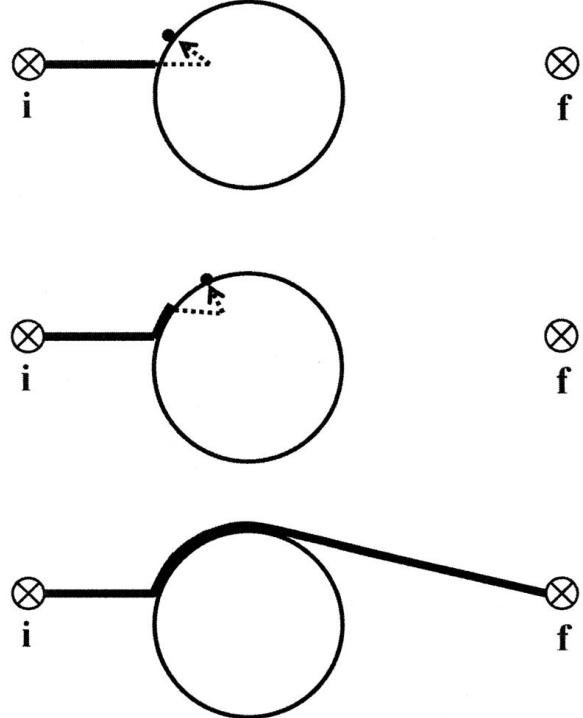


Figure 2.3: Free and escape steps wind around boundaries

From top to bottom: A free step causes system to enter forbidden region, which it then escapes to the boundary. The repeated action of free steps into the forbidden region and escape steps out trace the boundary of the obstacle in accordance with the Kuhn-Tucker theorem. Figure from [8].

where δ_s is a step length, chosen to be small relative to the features of the potential⁷. This is the same procedure used previously by the group [8]. After each step we check to make sure that the landscape criterion, $V(\underline{R}) \leq E_L$, is satisfied. If it is, we accept the step. If it is not, we are in a forbidden region and must escape. The repeated combination of the free, straight-line, steps and escape steps will cause the geodesic to wind around the obstacles; see Figure 2.3.

Typically, the group has used a Newton-Raphson root search to find the nearest solution to $V(\underline{R}) = E_L$.

$$\underline{R}^{n+1} = \underline{R}^n - \zeta \nabla V|_{\underline{R}^n} \quad (2.10)$$

⁷See Section 2.2.2.

where

$$\zeta = \frac{V(\underline{R}^n) - E_L}{\left| \nabla V|_{\underline{R}^n} \right|^2} \quad (2.11)$$

By repeated application of eq. 2.10, we would take steps following the gradient to the edge of the $V(\underline{R}) = E_L$ boundary. However, in the present case, there is an additional complication. Formaldehyde is a system that has heterogeneous masses and so moves along the gradient (as the above, naïve Newton-Raphson step would take) will displace the center of mass. This problem had never been encountered and its exposition is treated alone in Section 2.2.1. Resolution is found in a projected Newton-Raphson root search that prevents center of mass perturbation via a uniform back-translation. Escape steps are then found by iteratively following the projected gradient until the landscape criterion is satisfied:

$$\underline{R}^{n+1} = \underline{R}^n - \zeta (\mathbb{1} - \mathbb{M}) \cdot \nabla V|_{\underline{R}^n} \quad (2.12)$$

where ζ is defined as before in eq. 2.11 and the projection, $(\mathbb{1} - \mathbb{M})$, is a $3N \times 3N$ matrix, to be derived in Section 2.2.1.

Geodesics between endpoints are therefore found on the formaldehyde surface by applying eq. 2.9 and, as necessary, eq. 2.12 until the system is within $2\delta_s$ of the target.

Optimization While the geodesics found by the deterministic portion of our algorithm satisfy the Kuhn-Tucker theorem, they are not necessarily shortest paths—the theorem establishes necessary but not sufficient conditions of the form of the result. We therefore apply a simple optimization scheme to reduce the lengths. During the optimization phase, we construct geodesics using the previously developed algorithm and so all the properties of those geodesics hold here as well. The optimization round proceeds as follows:

1. A random configuration along the $\underline{R}_0 \rightarrow \underline{R}_f$ path is selected: \underline{R}^i , not one of the end-points
2. \underline{R}^i is perturbed:

$$\underline{R}^{i'} = \underline{R}^i \pm \delta_p (\mathbb{1} - \mathbb{M}) \cdot \hat{E}$$

Where the sign is chosen randomly between + and – with equal probability, $\delta_p = \delta_s$ is the length of a small, perturbative step, and \hat{E} is a configuration-space

unit vector along a randomly chosen, atomic, lab-frame axis. $(\mathbb{1} - \mathbb{M})$ is the same projection we use to enforce a constant center of mass (see Section 2.2.1).

3. If $V(\underline{R}^{i'}) > E_L$, the configuration is forbidden; a new \underline{R}^i is then chosen and the process repeated until $V(\underline{R}^{i'}) \leq E_L$ and $\underline{R}^{i'}$ is in the allowed region.
4. Two new geodesics, $\underline{R}_0 \rightarrow \underline{R}^{i'}$ and $\underline{R}^{i'} \rightarrow \underline{R}_f$, are computed.
5. If the total $\underline{R}_0 \rightarrow \underline{R}^{i'} \rightarrow \underline{R}_f$ path is shorter than the original, the new path replaces the old; otherwise the old path is retained.

The above procedure is repeated many times until the path length is converged; explicit termination criteria are given in Section 2.2.2.

2.2.1 Center of mass conservation

In systems with heterogeneous masses, a gradient-following Newton-Raphson root search erroneously introduces center of mass translation. Here we explore the ramifications of this and derive a projected root search that does not. This topic is also taken up in Appendix B, where two other methods are described. The results derived in this section allow us to construct geodesics that are guaranteed not to displace the center of mass while traversing configuration space. This is of particular importance to the present work, where the center of mass motion contributed a non-trivial excess to the total kinematic length. We considered enforcing other conservation properties, such as the Eckart conditions for rotations [57]. But the molecular-dynamics-derived endpoints do not obey these conditions⁸ and so there is no way to impose them on the geodesics.

Notation

In the analysis of molecular systems with N atoms in d physical dimensions, it is sometimes desirable to picture the system as specified by a single vector in the Nd -dimensional configuration space and at other times as N spatial vectors. We use the following definitions to transition between representations.

⁸nor should they, though they do conserve angular momentum.

- Let $\vec{x}_i \in \mathbb{R}^d$ be the column vector specifying the spatial coordinates of the i th atom; $i = 1 \dots N$.
- Let $\underline{R} \in \mathbb{R}^{dN}$ be the column vector representing the total configuration of the system; $\underline{R}^T = (\vec{x}_1^T, \vec{x}_2^T, \dots, \vec{x}_N^T)$
- Let m_i be the mass of the i th atom.

We can extract spatial vectors from the configuration space vector by defining the $d \times dN$ matrix, $\vec{\zeta}_i$ such that

$$\vec{x}_i = \vec{\zeta}_i \cdot \underline{R} \quad (2.13)$$

For instance, in 3 spatial dimensions,

$$\vec{\zeta}_1 = \begin{bmatrix} 1 \\ & 1 \\ & & 1 \\ & & & \dots \end{bmatrix}$$

In general,

$$(\vec{\zeta}_i)_j = \delta_{i,j} \overset{\leftrightarrow}{1} \quad , \quad j = 1 \dots N \quad (2.14)$$

Where $\overset{\leftrightarrow}{1}$ is the $d \times d$ unit matrix. In the other direction we have:

$$\underline{R} = \sum_{i=1}^N \vec{\zeta}_i^T \cdot \vec{x}_i \quad (2.15)$$

We compute the center of mass of the system as follows:

$$\begin{aligned} \vec{x}_{cm} &= \frac{1}{M_{\text{total}}} \sum_{i=1}^N m_i \vec{x}_i \\ &= \frac{1}{M_{\text{total}}} \sum_{i=1}^N m_i \vec{\zeta}_i \cdot \underline{R} \\ &= \left(\frac{1}{M_{\text{total}}} \sum_{i=1}^N m_i \vec{\zeta}_i \right) \cdot \underline{R} \\ &\equiv \underline{\vec{M}} \cdot \underline{R} \end{aligned} \quad (2.16)$$

where $M_{\text{total}} = \sum_{i=1}^N m_i$ is the total mass of the system and $\vec{\mathcal{M}}$ is a $d \times dN$ dimensional, block-diagonal matrix composed of N ordered, $d \times d$ blocks with m_i ($i = 1 \dots N$) along the diagonals.

$$\left(\vec{\mathcal{M}}\right)_i = \frac{m_i}{M_{\text{total}}} \overset{\leftrightarrow}{1} \quad , \quad i = 1 \dots N \quad (2.17)$$

Should the Center of Mass be Conserved?

Our intuition is that center of mass translation *should* increase the length of a path between two regions of configuration space which have the same center of mass; this is indeed the case as proved below.

The length we seek to minimize (eq. 2.5) is

$$\ell = \int d\tau \sqrt{2T(\tau)} \quad (2.18)$$

where $T(\tau)$ is the kinetic energy as a function of progress along the path, defined (equivalently to eq. 2.6) as:

$$2T(\tau) = \sum_i^N m_i \left(\frac{d\vec{x}_i}{d\tau} \right)^2 \quad (2.19)$$

Discretizing the integral in eq. 2.18 (see eq. 2.117 in Section 2.7.1) gives the contribution from a single small step,

$$\Delta\ell = \sqrt{\sum_i^N m_i \Delta\vec{x}_i^2} \quad (2.20)$$

where $\Delta\vec{x}_i$ is the displacement of the i th center in a time $\Delta\tau$. Suppose the $\{\Delta\vec{x}_i\}$ are decomposed into 2 components:

$$\Delta\vec{x}_i = \Delta\vec{x}_i^0 + \Delta\vec{x} \quad (2.21)$$

where $\Delta\vec{x}$ corresponds to net center of mass translation and the $\{\Delta\vec{x}_i^0\}$ preserve the center of mass. That is:

$$\vec{0} = \sum_i^N m_i \Delta\vec{x}_i^0 \quad (2.22)$$

Inserting eq. 2.21 into eq. 2.20 yields:

$$\begin{aligned}
 \Delta\ell &= \sqrt{\sum_i^N m_i (\Delta\vec{x}_i^0 + \Delta\vec{x})^2} \\
 &= \sqrt{\sum_i^N m_i \Delta\vec{x}_i^{02} + \Delta\vec{x}^2 \sum_i^N m_i + \left(\sum_i^N m_i \Delta\vec{x}_i^0 \right) \cdot \vec{x}} \\
 &= \sqrt{\sum_i^N m_i \Delta\vec{x}_i^{02} + M \Delta\vec{x}^2}
 \end{aligned} \tag{2.23}$$

where the last term of the middle line is 0 by eq. 2.22. Identifying,

$$\Delta\ell_0 = \sqrt{\sum_i^N m_i \Delta\vec{x}_i^{02}} \tag{2.24}$$

as the length in the absence of center of mass translation, we then have:

$$\Delta\ell = \sqrt{\Delta\ell_0^2 + M \Delta\vec{x}^2} > \Delta\ell_0 \tag{2.25}$$

Thus completing the proof that net center of mass translation increases the kinematic length.

Is the Center of Mass Conserved?

Having established that to be “shortest”, paths through configuration space must preserve the center of mass in real space, we turn to the algorithm to compute geodesics in the potential energy landscape ensemble presented in the previous section. There are two components of the algorithm, which treat “free” and “escape” steps respectively. We now analyze each in turn.

Free Steps We give eq. 2.9 for computing free (straight-line) steps between the current position, \underline{R}^t , and the target, \underline{R}_f :

$$\underline{R}^{t+1} = \underline{R}^t + \delta R \frac{\underline{R}_f - \underline{R}^t}{|\underline{R}_f - \underline{R}^t|} \tag{2.26}$$

Collapsing the scalars into the constant γ yields:

$$\underline{R}^{t+1} = \underline{R}^t + \gamma (\underline{R}_f - \underline{R}^t) \quad (2.27)$$

We can impose the requirement that our boundaries have the same center of mass (recall eq. 2.16) by demanding:

$$\vec{M} \cdot \underline{R}^0 = \vec{M} \cdot \underline{R}_f \quad (2.28)$$

And this is indeed the case for endpoints generated by molecular dynamics. We then ask if \underline{R}^1 , computed by eq. 2.27 will have the same center of mass as \underline{R}^0 .

$$\begin{aligned} \vec{M} \cdot \underline{\underline{R}}^0 &\stackrel{?}{=} \vec{M} \cdot \underline{\underline{R}}^1 \\ &= \vec{M} \cdot \left[\underline{R}^0 + \gamma (\underline{R}_f - \underline{R}^0) \right] \end{aligned} \quad (2.29)$$

Cancelling the common term and the scalar leaves:

$$\vec{0} = \vec{M} \cdot (\underline{R}_f - \underline{R}^0), \quad (2.30)$$

which is a restatement of the limits on our boundary conditions, eq. 2.28. This proves, therefore, that the first step does not perturb the center of mass. By induction, we conclude:

$$\vec{M} \cdot \underline{R}^0 = \vec{M} \cdot \underline{R}^t = \vec{M} \cdot \underline{R}_f \quad (2.31)$$

for all t . And so the center of mass is not perturbed by the free steps of the algorithm.

Escape Steps When the system enters a region where the potential is greater than the landscape energy, $V(\underline{R}) > E_L$, An escape step is required. Usually this entails a Newton-Raphson root search—iteratively following the gradient until the landscape energy criterion is satisfied:

$$\underline{R}^{n+1} = \underline{R}^n - \frac{V(\underline{R}^n) - E_L}{\left| \nabla V \right|_{\underline{R}^n}^2} \cdot \nabla V \Big|_{\underline{R}^n} \quad (2.32)$$

Such steps do no preserve the center of mass; here we show how. As in the case of the free steps, we collapse all the scalars into a constant term, ζ , leaving:

$$\underline{R}^{n+1} = \underline{R}^n - \zeta \nabla V \Big|_{\underline{R}^n} \quad (2.33)$$

Again our question is, *Does the iterated map preserve the center of mass?* We encode this as (recall eq. 2.16):

$$\begin{aligned} \tilde{\vec{M}} \cdot \tilde{R}^n &\stackrel{?}{=} \tilde{\vec{M}} \cdot \tilde{R}^{n+1} \\ &\stackrel{?}{=} \tilde{\vec{M}} \cdot \left(\tilde{R}^n - \zeta \tilde{\nabla} V \Big|_{\tilde{R}^n} \right) \end{aligned} \quad (2.34)$$

Cancelling the common term and the scalar leaves us with the following expression:

$$\vec{0} \stackrel{?}{=} \tilde{\vec{M}} \cdot \tilde{\nabla} V \Big|_{\tilde{R}^n} \quad (2.35)$$

which doesn't look too good.

Does it Hold? Proving that eq. 2.35 holds in general would be quite difficult⁹, but invalidating it requires only one counter-example. All the problems treated by the group's analysis of geodesics to date, involve translationally invariant potentials, so we confine our attention to a particularly simple one: the pair-potential between two mutually interacting centers, labeled 1 and 2, in 3 dimensions. The form of our potential is then:

$$V = u(\vec{x}_1 - \vec{x}_2) \quad (2.36)$$

Taking $\vec{x} = \vec{x}_1 - \vec{x}_2$ We compute the derivative in eq. 2.35, arriving at:

$$\tilde{\nabla} V = \begin{pmatrix} + \frac{\partial u}{\partial \vec{x}} \\ - \frac{\partial u}{\partial \vec{x}} \end{pmatrix} \quad (2.37)$$

Inserting eq. 2.37 into eq. 2.35 gives

$$\begin{aligned} \tilde{\vec{M}} \cdot \tilde{\nabla} V \Big|_{\tilde{R}^n} &= m_1 \frac{\partial u}{\partial \vec{x}} - m_2 \frac{\partial u}{\partial \vec{x}} \\ &= \frac{\partial u}{\partial \vec{x}} (m_1 - m_2) \\ &\stackrel{?}{=} \vec{0} \end{aligned} \quad (2.38)$$

Showing that eq. 2.35 fails for $m_1 \neq m_2$. Therefore, a simple, gradient-following root search does *not* preserve the center of mass with heterogeneous masses, which

⁹and would require mistakes, as it does not.

erroneously lengthens the path. On the flip-side, eq. 2.35 *does* hold when the masses are the same. It can be shown that eq. 2.35 holds for a sum of such pair potentials with equal masses and therefore the group's previous results are safe from this issue.

Enforcing Center of Mass Conservation Analytically

In this section, we derive a transformation to displacements in configuration space, which leaves them unmodified except for removing any center of mass translation they would have introduced. We effectively accomplish this *via* a uniform back-translation. This general result is then verified. The method presented here is the one used in our work though is only one of several. Two others, of staggeringly different complexity, are presented in Appendix B.

Suppose we have a general displacement in configuration space, $\underline{\Delta}R$, and would like to remove center of mass motion. We therefore seek \underline{P} such that:

$$\vec{0} = \vec{M} \cdot (\underline{\Delta}R + \underline{P}) \quad (2.39)$$

This leads to the under-determined relation:

$$\vec{M} \cdot \underline{P} = -\vec{M} \cdot \underline{\Delta}R \quad (2.40)$$

There are many solutions to this equation (including $\underline{P} = -\underline{\Delta}R$, which wouldn't help us much!), but we can choose the following convenient approach to constructing a useful solution: Let \underline{P} be the configuration space vector which displaces each center by the *same* quantity such that eq. 2.40 is satisfied. This quantity, of course, will be the negative of the net center of mass displacement introduced by $\underline{\Delta}R$ and is given by:

$$\vec{\Delta}x_{cm} = -\vec{M} \cdot \underline{\Delta}R \quad (2.41)$$

We can build \underline{P} from $\vec{\Delta}x_{cm}$ using eq. 2.15 as follows:

$$\underline{P} = \left(\sum_{i=1}^N \vec{C}_i^T \right) \cdot \vec{\Delta}x_{cm} \quad (2.42)$$

Defining the summed matrices as $\vec{\bar{C}}^T$,

$$\left(\vec{\bar{C}}^T \right)_i = \overset{\leftrightarrow}{1} \quad , \quad i = 1 \dots N \quad (2.43)$$

we can write

$$\underline{P} = -\vec{\underline{C}}^T \cdot \vec{\underline{M}} \cdot \underline{\Delta R} \quad (2.44)$$

Further, defining

$$\underline{M} = \vec{\underline{C}}^T \cdot \vec{\underline{M}} \quad (2.45)$$

we can write

$$\underline{P} = -\underline{M} \cdot \underline{\Delta R} \quad (2.46)$$

So the configuration space representation of $\underline{\Delta R}$, which includes no center of mass translation is:

$$\underline{\Delta R}^* = \underline{\Delta R} + \underline{P} = (\mathbb{1} - \underline{M}) \cdot \underline{\Delta R} \quad (2.47)$$

Where $\mathbb{1}$ is the $dN \times dN$ identity.

The $dN \times dN$ matrix \underline{M} has a band structure defined by:

$$\underline{M}_{i,j} = \frac{m_j}{M_{\text{total}}} \overset{\leftrightarrow}{1} \quad , \quad i, j = 1 \dots N \quad (2.48)$$

For instance, in three spatial dimensions, $d = 3$, we have:

$$\underline{M} = \frac{1}{M_{\text{total}}} \begin{bmatrix} m_1 & 0 & 0 & m_2 & 0 & 0 & \cdots & m_N & 0 & 0 \\ 0 & m_1 & 0 & 0 & m_2 & 0 & \cdots & 0 & m_N & 0 \\ 0 & 0 & m_1 & 0 & 0 & m_2 & \cdots & 0 & 0 & m_N \\ m_1 & 0 & 0 & m_2 & 0 & 0 & & & & \\ 0 & m_1 & 0 & 0 & m_2 & 0 & & & & \vdots \\ 0 & 0 & m_1 & 0 & 0 & m_2 & & & & \\ & \vdots & & & \vdots & & & & & \end{bmatrix} \quad (2.49)$$

Armed with this result we can re-write the equation for escape steps (eqs. 2.32 and 2.33) as:

$$\underline{R}^{n+1} = \underline{R}^n - \zeta (\mathbb{1} - \underline{M}) \cdot \nabla V|_{\underline{R}^n} \quad (2.50)$$

where

$$\zeta = \frac{V(\underline{R}^n) - E_L}{\left| \nabla V|_{\underline{R}^n} \right|^2} \quad (2.51)$$

as before.

Verification

We would like to verify that our expression (eq. 2.50) conserves the center of mass. We proceed as before, using our new result.

$$\begin{aligned}\vec{\underline{M}} \cdot \underline{R}^n &\stackrel{?}{=} \vec{\underline{M}} \cdot \underline{R}^{n+1} \\ &\stackrel{?}{=} \vec{\underline{M}} \cdot \left(\underline{R}^n - \zeta (\mathbb{1} - \mathbb{M}) \cdot \nabla V \Big|_{\underline{R}^n} \right)\end{aligned}\quad (2.52)$$

After, rearranging and canceling the common term as well as the scalar, we have:

$$\vec{\underline{M}} \cdot \nabla V \Big|_{\underline{R}^n} \stackrel{?}{=} \vec{\underline{M}} \cdot \mathbb{M} \cdot \nabla V \Big|_{\underline{R}^n} \quad (2.53)$$

Inserting the definition for \mathbb{M} (eq. 2.45) gives:

$$(\vec{\underline{M}}) \cdot \nabla V \Big|_{\underline{R}^n} \stackrel{?}{=} (\vec{\underline{M}} \cdot \vec{\mathbf{C}}^T \cdot \vec{\underline{M}}) \cdot \nabla V \Big|_{\underline{R}^n} \quad (2.54)$$

which can only hold if

$$\vec{\underline{M}} = \vec{\underline{M}} \cdot \vec{\mathbf{C}}^T \cdot \vec{\underline{M}} \quad (2.55)$$

We can see that this does indeed hold by inserting the component definitions of $\vec{\mathbf{C}}^T$ (eq. 2.43) and $\vec{\underline{M}}$ (eq. 2.17):

$$\begin{aligned}[\vec{\underline{M}} \cdot \vec{\mathbf{C}}^T] \cdot \vec{\underline{M}} &= \left[\sum_i^N \vec{\underline{M}} \cdot \vec{\mathbf{c}}_i^T \right] \cdot \vec{\underline{M}} \\ &= \left[\sum_i^N \frac{m_i}{M_{\text{total}}} \vec{1} \cdot \vec{1} \right] \cdot \vec{\underline{M}} \\ &= \left[\vec{1} \left(\frac{1}{M_{\text{total}}} \right) \left(\sum_i^N m_i \right) \right] \cdot \vec{\underline{M}} \\ &= [\vec{1}] \cdot \vec{\underline{M}} \\ \vec{\underline{M}} \cdot \vec{\mathbf{C}}^T \cdot \vec{\underline{M}} &= \vec{\underline{M}}\end{aligned}\quad (2.56)$$

What of the Newton-Raphson Root Search? From the above, eq. 2.50 clearly preserves the center of mass. However, the point of eq. 2.32 was a root search to ensure $V(\underline{R}) = E_L$. Does the sequence of projected Newton-Raphson steps retain the property

that they locate the $V(\mathcal{R}) = E_L$ boundary of the forbidden region? Yes it has; the construction of \mathbb{M} began with \mathcal{P} in eq. 2.42. In which we specified displacing all centers by the same amount, $\vec{\Delta x}_{cm}$. This amounts to a net translation, under which our potential is invariant. In Appendix B.2, we take up the general problem of projected root searches. If our problem contained a field and the potential was not translationally invariant or if the boundary configurations did not have the same center of mass, all of this would go out the window. In particular if it were not the case that the potential was translationally invariant, we would not be able to rationalize there being something “wrong” with center of mass translation in the geodesic as we did earlier.

2.2.2 Geodesic parameters

With respect to step sizes, we take steps during the optimization phase of length equal to the steps during the deterministic phase, $\delta_p = \delta_s = 5 \times 10^{-2} a_0 = 2.645 \times 10^{-2} \text{ \AA}$. We chose this value because we expected variations in the potential to be small on this scale and smaller values markedly increased computation time without substantively increasing the length of the path found. Lengthening of the path corresponds to skipping fewer obstacles and/or more closely tracing curves—thereby more accurately exploring the landscape. See Table 2.2 for a comparison of lengths over a range of step sizes.

With regards to termination criteria, we aggressively optimize of the length of the paths: requiring 10^3 trials resulting in no shortening. This means that on average, optimizing a single geodesic requires computing order 10^4 paths. In most cases, the optimization phase offers modest (5%) reductions in length of the computed geodesics, however there are some extreme cases where it yields length reductions of order one half.

There are also times when the relative positions of endpoints prevent our algorithm from finding a connecting path between them¹⁰—it is quite possible that one does not even exist. Pairs of endpoints were excluded from analysis when the algorithm failed to find a path within a “reasonable” (orders of magnitude greater than typical) amount of time. Specifically, if the algorithm required more than 10^2 more steps than

¹⁰For details and rates of failure, see Table 3.12 on page 91.

δ_s / a_0	Computed Path Length / a_0	Execution Time / s
5×10^{-1}	16.6827	0.06
5×10^{-2}	17.2255	0.33
5×10^{-3}	17.2570	3.51
5×10^{-4}	17.2620	35.81

Table 2.2: Comparison of step sizes in geodesic algorithm

Comparison of Euclidean path lengths (without optimization) and execution times to compute geodesics for different step sizes on the formaldehyde potential with $E_L = 34928.7 \text{ cm}^{-1}$. In this example, the Cartesian distance between the products and reactants is $15.923 \text{ } a_0$. An attempt was made with $\delta_s = 5 \times 10^{-5}$, but it did not terminate on its own within 1000 seconds. We postulate that this is because the step size was smaller than the scale used in computing the gradient (see Section 2.7.2), which rendered the algorithm inaccurate.

an unobstructed path would require, the search was abandoned:

$$i_{\max} = \left\lfloor 10^2 \cdot \frac{|R_f - R_0|}{\delta_s} \right\rfloor \quad (2.57)$$

where $\lfloor \cdot \rfloor$ is the floor function. We also set the cutoff at 10^3 , but this served only to lengthen execution time and did not appreciably decrease the rate of failure.

Selecting a landscape energy for a small molecule system

Another concern of importance is the selection of the landscape energy. In the group's previous work, which has focused on macroscopic systems, the average value of the potential over the course of a simulation was used [8, 49]. Indeed, in the thermodynamic limit, $\langle V \rangle = E_L$ exactly. In the present case, we are not in the thermodynamic limit and such a solution would be of little use. The system spends much of its time in the reactive well, where the majority of the internal energy is kinetic. The moment of passage through or around any transition state is but fleeting and will not meaningfully contribute to the average.

Another technique is simply to select some probable, "thermodynamic" value for the energy. This is what we did in our first paper treating formaldehyde [1]: the landscape energy was taken from the peak of the distribution of values of the potential while a hydrogen was far-removed from the formyl over many molecular dynamics simulations.

E_μ / cm^{-1}	excitation wavelength / nm
34500	349.0
*35000	343.0
35500	337.2
*36000	331.6
36223	329.2
37000	321.0
*37600	314.9
37688	314.0
38000	311.0
38688	304.5
*38814	303.3
*41010	284.4
<hr/>	
*45000	255.4

Table 2.3: Microcanonical and landscape energies

These energies were selected to span the range of energies over which roaming manifests in formaldehyde and to allow maximal overlap with the studies of Houston *et al.* [58]. Energies above were used for sampling reactant well and for molecular dynamics studies. An experimental realization of these energies would correspond a laser with energy $E_\mu - E_{\text{Zpe}}$; $E_{\text{Zpe}} = 5844 \text{ cm}^{-1}$, see Section 2.1.1. An asterisk (*) indicates energies for which geodesics were computed. The horizontal line shows the approximate limit at which roaming is observed experimentally (41060 cm^{-1} [17]).

While this is fine, it is still somewhat arbitrary in that it depends on the choices for “far-removed”. It could also conceivably forbid paths that are indeed classically energetically accessible on the energy surface. In small systems, with few degrees of freedom, it is the partitioning of energy into different degrees of freedom that controls the particulars of the dynamics. The key observation is this: the microcanonical energy, E_μ , already sets a limit on the maximum potential energy, $V(\mathcal{R}) \leq E_\mu$. For rare processes, such as roaming, energy may need to be partitioned among the various degrees of freedom very specifically. By taking the landscape energy equal to the microcanonical energy, $E_L = E_\mu$, all such partitionings and transformations between them which are classically accessible, are accessible to the geodesics we construct in the landscape ensemble. Desiring such properties, this is just what we do.

For this work we compute geodesics at the energies given in Table 2.3; see Figure 2.1 for comparison to the relevant energies in formaldehyde. While all the major conclusions

described in Chapter 3 rest upon geodesics constructed with $E_L = E_\mu$, we do, at times, appeal to analyses conducted for our previous work [1]. In these cases, $E_\mu = 37600 \text{ cm}^{-1}$ and $E_L = 34928.7 \text{ cm}^{-1}$. In that earlier work, we also repeated our analysis of direct and roaming geodesics with the landscape energy set slightly *below* the radical dissociation threshold¹¹, $E_L = 33238.7 \text{ cm}^{-1} < 33240 \text{ cm}^{-1} = E_\mu^{\text{rad}}$. We did not repeat the sampling procedure for reactants or products described below in Section 2.3, but did for the roaming region. Our results and conclusions were virtually identical for this lower bound on the landscape energy.

2.2.3 Trapped at a fixed point in 2 dimensions

As part of getting up to speed with the group’s methods, we re-implemented Wang and Stratt’s early work on geodesics [8]. In the process we encountered a previously unreported failure mode of the geodesic algorithm described in Section 2.2. The failure is only of practical concern in systems with two degrees of freedom and presumably lay undetected because the group typically treats problems 1 to 3 orders of magnitude larger.

The Müller-Brown potential [45] is a 2-dimensional model surface that captures the key features of typical chemically-relevant energy landscapes. It has two major basins, an intermediate, and two saddle points—these correspond to reactant, product, intermediate, and transition states between them. Formed as a superposition of Gaussians, it is simple to compute and the surface is often used as test case for new optimization or path-finding algorithms. So it was used by our group. The potential is shown in figure 2.4.

The major surprise while implementing the algorithm came during the testing phase. When finding geodesics from the upper (“reactant”) basin to the lower (“product”) basin, the program became stuck in an infinite loop for landscape energies below -34.06. In the opposite direction, the same thing happened at an even higher energy: -29.93. Both of these energies are comfortably above the highest saddle connecting the basins, -40.7, so there should have left plenty of space for a transition.

The error was *not* in the implementation, it is a feature of the algorithm when

¹¹Radical geodesics cannot be computed with landscape energy below the threshold for radical dissociation.

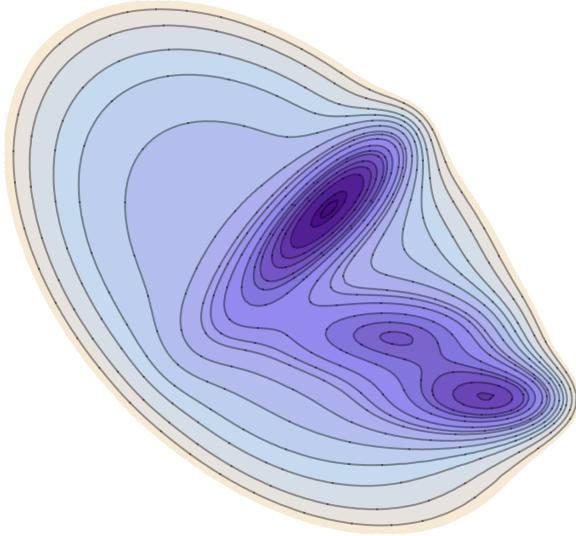


Figure 2.4: Müller-Brown potential

In the arbitrary units of the original paper, this depiction spans an energy range of roughly -150 to +150. Basins have depths of -146.7, -80.8, and -108.1 (diagonally from upper left) and saddles are at energies of -40.7 and -72.3 (diagonally from upper left).

operating in two dimensions. The failure is more likely to occur the longer the algorithm is confined to a basin with concave boundaries. *En route* to escape, the trial point, \underline{R}^t , winds around the edge of the basin. As it does, the unit vector pointing at the target smoothly rotates in one direction. At the same time, the direction of steepest descent, the unit gradient, rotates smoothly in the *opposite* direction. If the particle does not escape the basin in time, at some iteration, the two are parallel and point directly at each other. Then the algorithm encounters a fixed point, and will forever take steps back-and-forth in the same location: $\underline{R}_n^{t+1} = \underline{R}^t$; see Figure 2.5. The fixed point is stable because following the gradient in the Newton-Raphson search (eq. 2.12) on one side and the trial step towards the target (eq. 2.9) on the other will drive the system back towards the problem location and thus the errors associated with computing in finite precision will be of no help in escape.

The problem manifests for two degrees of freedom, but not for more because the rotating unit vectors trace out the unit circle, S^1 . Given long enough time and if they rotate in “opposite directions”, they are guaranteed to be coincident. On unit spheres of greater dimension, there is more than one angular degree of freedom and the target and gradient vectors may wander without encountering each other. Thus for systems

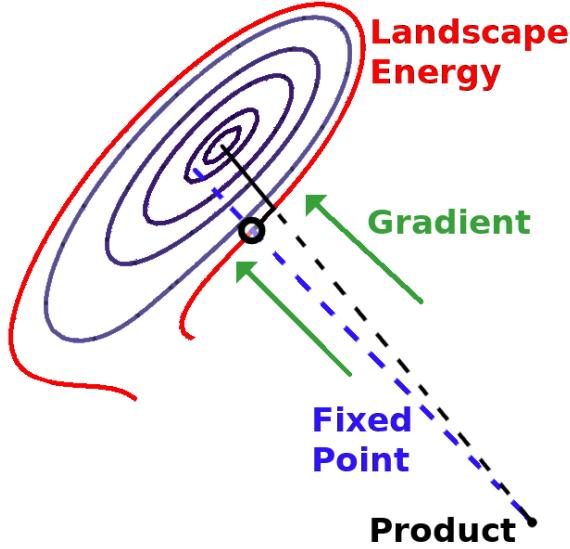


Figure 2.5: Fixed point on the Müller-Brown potential

The highlighted contour is at the landscape energy. The initial path of the algorithm is shown in black. The dashed lines point to the product, R_{Product} , from the current position, $R^{(t)}$. The gradient is shown with arrows. A fixed point occurs at the **O** when the vector pointing at the product is parallel to the gradient:

$$R_f - R^t \parallel \nabla V(R)|_{R^t}$$

with $n > 2$ degrees of freedom, the probability of coincidence on S^{n-1} is infinitesimal and the problem does not occur.

2.3 Defining and sampling regions of interest

Our algorithm generates paths that give us information about the connectivity between regions of configuration space. Rather than being concerned with the fine details of minima, we sample the volumes of configuration space corresponding to the species whose inter-conversion we wish to study: reactants, molecular and radical products, and roaming intermediates. As geodesics are paths between points, we need a way to sample these volumes of configurations space.

The following section details our methods of drawing points from the regions

of configuration space we explore. To compute geodesics, we select endpoints by randomly drawing points from the appropriate distributions. Some of our sampling techniques produce points in *phase* space (including kinetic degrees of freedom). As our algorithm only requires points in *configuration* space as input, we simply project out the momenta when necessary.

2.3.1 Monte Carlo for microcanonical reactants

To microcanonically sample region of the reactants we use the method of normal mode approximation with anharmonic corrections described by Peslherbe and coworkers [59]; we describe the high-points here and in more detail in Appendix C. First the potential is approximated by normal modes constructed from a diagonalization of the Hessian matrix at the global minimum. Energy is then deposited in these modes according the microcanonical distribution for a collection of harmonic oscillators with total energy, E_μ . In our studies, E_μ takes on the values given in Table 2.3 on page 33.

As we desire to fix total angular momentum to 0, excess angular momentum is then removed by subtracting the net instantaneous rotation of the molecule. First, we form the spurious angular momentum (this is simply the total angular momentum of the molecule),

$$\vec{j}_s = \sum_{i=1}^N \vec{r}_i \times m_i \vec{v}_i \quad (2.58)$$

Where the sum is over atomic centers and \vec{r}_i , \vec{v}_i , and m_i give the Cartesian position, velocity, and mass of the i th atom. To correct for this term we add $\vec{j} = -\vec{j}_s$ to the molecule by computing the vector angular velocity:

$$\boldsymbol{\omega} = \mathbf{I}^{-1} \vec{j} \quad (2.59)$$

where \mathbf{I}^{-1} is the inverse of the moment of inertia tensor [60]. We then add this angular velocity to each center:

$$\vec{v}'_i = \vec{v}_i + \boldsymbol{\omega} \times \vec{r}_i \quad (2.60)$$

As we deposit energy in the normal modes, there will automatically be no net translational momentum. Finally, to correct for anharmonicity of the potential, the coordinates are iteratively displaced and the momenta scaled until the phase space point reflects the desired total microcanonical energy (see Appendix C.3).

Using this method, 4×10^4 such phase space points were generated and then used as inputs to molecular dynamics described in the next section. They were also used as initial configurations for our algorithm for constructing geodesics on the energy landscape. Of note: only those configurations, whose phase space points lead to molecular products were used as input to the geodesic algorithm. We have no reason to believe that such dynamical correlation would impact our results; see Section 3.6.3.

2.3.2 Molecular dynamics for products

To generate configurations corresponding to molecular and radical products, the reactant phase space points are propagated via molecular dynamics. At each time step forces act along the negative of the gradient. A 4th order Runge-Kutta integrator [61, equation 25.5.10] implemented in the GNU Scientific Library [62] advances the state of the system according to Hamilton's equations for the potential. This point in phase space is propagated forward in steps of at most¹² $5 \cdot \hbar/E_h = 0.121$ fs until one of the following occurs:

1. The center of mass distance between H₂ and CO is greater than 12 Bohr ($12 a_0 = 6.350 \text{ \AA}$), indicating dissociation (molecular or radical), or
2. 12.1 ps of simulation time ($5 \times 10^5 \hbar/E_h$) have elapsed without the above occurring, indicating no reaction.

During the trajectory, the state of the system is recorded every $100\hbar/E_H = 2.42$ fs. These termination conditions and time step choices are commensurate with those used by Bowman and coworkers [55] when working on the same surface.

Once a reactive trajectory has been identified, the H–H bond distance is probed to assign the trajectory to radical or molecular products. We assign molecular products for bond distances less than $6 a_0$ (3.18 Å) and radical products otherwise. We do not address the triple-fragmentation channel as it is energetically inaccessible.

For each of the energies listed in Table 2.3 on page 33, 4×10^4 such trajectories were simulated; their final dispositions are discussed in Section 3.3. The final states of

¹²If the implementing library deems the accuracy (estimated through step-doubling [63]) with the given time-step insufficient, the time-step is reduced by 20%. This is repeated to a minimum of $0.84\hbar/E_h$ at which point the trajectory is discarded as a failure.

the reactive trajectories were then used as endpoints for geodesics¹³.

2.3.3 Monte Carlo random walk for the roaming region

One of the difficulties associated with studying roaming is unambiguously deciding when it happens. Given the high vibrational excitation necessary to elicit roaming in H₂CO, C–H bonds may often be quite stretched during a molecular dynamics trajectory and a hydrogen might often “look like” it is or is about to roam. The situation is in little danger of improving as there is no consensus in the literature defining roaming.

Our notion of roaming is based on two components. The first is a *static*, configuration space requirement. Like other workers [30, 42], we make use of a roaming region, which is described below. The second is a *dynamical* feature of a trajectory, namely that after excitation, the roaming hydrogen visits this region and then returns to abstract the other H from the formyl and to dissociate to close-shelled molecular products.

Because roaming is a rare event, the molecular dynamics studies described in the previous section do not produce enough configurations to adequately sample the roaming region at a given energy. In this section, we also describe the random walk Monte Carlo method used to sample the roaming region in the landscape ensemble.

Defining the roaming region

A succinct definition for the roaming region remained elusive for some time. Key to our effort adding a force-based constraint to the otherwise geometric considerations. The force criterion allows us to ensure that a dissociating hydrogen is indeed on a potential energy plateau (small forces if and only if small gradients), characteristic of roaming dynamics. For the geometric criteria, we err on the side of over-constraint so as to avoid mis-classifying states of radical dissociation as roaming.

A configuration-space point is in the roaming region per our definition if the following criteria are satisfied; see Table 2.4 for constants:

¹³While the ends of all reactive trajectories were used as endpoints for geodesics, only those initial phase space points that resulted in *molecular* products were used as start-points. We discuss the impacts of dynamical correlation in detail in Section 3.6.3, but suffice it to say that we expect this particular choice to have no impact on our results.

1. The restoring force on the roaming hydrogen is of small magnitude¹⁴:

$$F_{H^{(0)}} < F_{\max} \quad (2.61)$$

2. The hydrogens are separated by more than a minimum and less than a maximum distance:

$$d_{\min} < |\vec{r}_{HH^{(0)}}| < d_{\max} \quad (2.62)$$

The roaming hydrogen, designated $H^{(0)}$, is uniquely identified in configuration space as the hydrogen with the greatest Euclidean separation from the total center of mass. We define the restoring force on the roaming hydrogen as:

$$F_{H^{(0)}} = -\hat{r}_{H^{(0)}-CM} \cdot \vec{\nabla}_{H^{(0)}} V(\mathcal{R}) \quad (2.63)$$

where

$$\vec{\nabla}_{H^{(0)}} = \left(\frac{\partial}{\partial x_{H^{(0)}}}, \frac{\partial}{\partial y_{H^{(0)}}}, \frac{\partial}{\partial z_{H^{(0)}}} \right) \quad (2.64)$$

and

$$\hat{r}_{H^{(0)}-CM} = (\vec{r}_{CM} - \vec{r}_{H^{(0)}}) / \|\vec{r}_{CM} - \vec{r}_{H^{(0)}}\| \quad (2.65)$$

As usual, $V(\mathcal{R})$ is the formaldehyde potential energy function and $\vec{r}_{H^{(0)}}$ and \vec{r}_{CM} are the real-space coordinates of the roaming hydrogen and total center of mass respectively. One may notice that under the above definitions, attractive forces are positive and repulsive forces are negative (eq. 2.65 results in this inversion of convention). We hope that our choice allows us to express the bounds on the force more intuitively as it is most important the restoring force be *small* in magnitude¹⁴.

We place a maximum on the force because in roaming trajectories, there are large, persistent H–HCO separations, which require small forces on the wayfaring hydrogen. Bounds on hydrogen-separation are required because the force criterion alone would mis-classify equilibrium and transition state configurations, clearly not representative of roaming, as such. We use the hydrogen separation as opposed to H–HCO separation because large hydrogen-formyl separation can also correspond to

¹⁴If the restoring force is to be attractive, it must be small otherwise the hydrogen is not on a plateau and will plunge back into the reactive center. For simplicity, we place no bound on the extent to which the force is repulsive; the geometric and energetic constraints already cover this case.

Parameter	Value
F_{\max}	-103 cm ⁻¹ /Å
d_{\min}	3.10 Å
d_{\max}	4.76 Å

Table 2.4: Roaming region constants

dissociation to molecular products¹⁵. The upper bound excludes radical dissociation, which otherwise comprises a majority of the space.

One benefit of the criteria we employ is the ability to distinguish a wide variety of roaming behavior. Had we simply relied on a dynamical signature like “low rotational quantum number after dissociation”, e.g. $J_{\text{CO}} \leq 16$, we would have missed trajectories that manifestly roam with J_{CO} much higher; see Section 3.3.4 for more.

Sampling the roaming region

To sample the roaming region we use a Monte Carlo random-walk technique adapted from Brady and coworkers [64], which was originally designed to facilitate micro-canonical sampling of Lennard-Jones clusters. Here, we use it to sample configurations in the potential energy landscape ensemble for E_L with the additional requirement that the configurations be in the roaming region described above. The details are given in Appendix D.

Our probability density is a function which equals 1 if the trial configuration is in the roaming region *and* meets the landscape criterion and otherwise equals 0. Because the codomain of this density is binary, the acceptance probability is as well—there are no marginally better or worse moves, only allowed and disallowed.

To generate trial moves for the random walk, we displace a single random atom, j , along a random coordinate, by a random amount, $\vec{\Delta}_j = \hat{\mu}_j(\xi - 0.5)\delta_x$. Where ξ is a uniformly distributed, random variable on $[0, 1]$, δ_x is a scaling factor and $\hat{\mu}_j \in \{\hat{x}, \hat{y}, \hat{z}\}$ is along a randomly selected Cartesian axis of the j th atom. After each such step, an additional move to preserve the center of mass is made by moving *all* atoms by the

¹⁵Consider, for distantly separated H₂ and CO, the relative centers of mass of a single hydrogen and a formally defined “formyl” composed of the other hydrogen and the CO.

same, appropriate amount. Thus the total move for each atom i is:

$$\Delta \vec{r}_i = \hat{\mu}_j(\xi - 0.5)\delta_x \left(\delta_{i,j} - \frac{m_j}{M} \right) ; \quad i = 1 \dots N \quad (2.66)$$

where $M = \sum_{i=1}^N m_i$ is the total mass of the system.

To select δ_x , we computed several “burn-in” sequences of 10^3 steps at $E_L = 37600 \text{ cm}^{-1}$ during which δ_x was scaled up and down in 5% steps until we achieved an acceptance ratio between 0.1 and 0.9. During the data collection phase, δ_x was pinned to its final value of $0.27562 a_0$, which yielded acceptance ratios between 0.80 and 0.95. To seed the random walk we took configurations satisfying both the landscape and roaming criteria from molecular dynamics trajectories that went to molecular products (roaming trajectories); this yielded between 200 and 9000 configurations at each energy. Using each configuration as a different starting position, we generated random walks of length 10^5 steps, recording a configuration every 100 moves. This gave a list of $10^5\text{--}10^7$ points in the roaming region, which were then shuffled and used as inputs to the geodesic algorithm. For energy specific details, see Table D.1.

2.3.4 Geodesic itineraries

Once points in each of the regions were generated at each of the landscape energies, we construct itineraries for the geodesics as follows:

- A reactant configuration is drawn from the those phase space points microcanonically sampled from the reactant well as described in Section 2.3.1 that later dissociates to closed shell products.
- A *radical product* and a *molecular product* endpoints are each selected from the products of the molecular dynamics simulation described in Section 2.3.2
- An intermediate, *roaming point*, is drawn from points sampled via a random walk over the roaming region and the landscape ensemble as described in Section 2.3.3.

Geodesics are then computed as described in Section 2.2 between points as given here:

direct: reactant \rightarrow *molecular product*

roaming: $(\text{reactant} \rightarrow \text{roaming point}) \cup (\text{roaming point} \rightarrow \text{molecular product})$

radical: reactant \rightarrow *radical product*

The union (\cup) specifies that roaming geodesics are the concatenation of two geodesics. Tens of thousands of unique routes are thus planned and executed at each energy.

2.4 Analyzing geodesics

Having computed some hundreds of thousands of geodesics, we would like to analyze them in aggregate. Our goal is to extract the key features of the inherent dynamics of the of each of the dissociation channels—and in particular to understand the genesis of roaming.

2.4.1 Total lengths

Given that the geodesics are characterized by a minimum kinematic length, measuring the lengths of the paths is a natural place to start. The total kinematic length of a path gives us information about the total amount of motion required along the geodesic. This length is defined as follows:

$$\ell_{\text{Total}} = \int_0^t d\tau [2 \cdot T(\tau)]^{1/2} \quad (2.67)$$

where T is the kinetic energy, τ is a progress variable on $[0, t]$, specifying the limits of integration are over the entire path. To find the total kinematic length, we insert $T = \frac{1}{2} \dot{\mathcal{R}}^T \mathbf{M} \dot{\mathcal{R}}$ and have:

$$\ell_{\text{Total}} = \int_0^t d\tau [\dot{\mathcal{R}}^T \mathbf{M} \dot{\mathcal{R}}]^{1/2} \quad (2.68)$$

where $\dot{\mathcal{R}} = \frac{d\mathcal{R}}{d\tau}$ is the rate of change of the configuration space vector \mathcal{R} and \mathbf{M} is the diagonal mass matrix:

$$\mathbf{M} = \begin{pmatrix} m_1 & & & & \\ & m_1 & & & \\ & & m_1 & & \\ & & & m_2 & \\ & & & & \ddots \\ & & & & & m_N \end{pmatrix} \quad (2.69)$$

or

$$(\mathbf{M})_{i\mu,j\nu} = \delta_{i,j}\delta_{\mu,\nu}m_i \quad i,j = 1 \dots N \quad ; \quad \mu, \nu = x, y, z \quad (2.70)$$

Where m_i is the mass of the i th particle in the N -particle system. (Recall Table 2.1 for the relevant masses.) In Section 2.7, we will discuss the finite difference expressions used to calculate the integral in eq. 2.68. The reduction provided by eq. 2.113 yields, for n steps:

$$\ell_{\text{Total}} = \sum_{i=0}^{n-1} [\delta \mathcal{R}^i]^T \mathbf{M} \delta \mathcal{R}^i]^{1/2} \quad (2.71)$$

where $\delta \mathcal{R}^i = \mathcal{R}^{i+1} - \mathcal{R}^i$. The expression for the length is simply the mass weighted sum of the steps along the path.

As the kinetic energy can be partitioned between many degrees of freedom, we can interrogate kinematic lengths corresponding to a subset of them by constructing the appropriate component of the kinetic energy (e.g.: rotation or vibration) and integrating as before. In doing so we can probe the contributions of different kinds of motion to the inherent dynamics. In general, we can construct arbitrary kinematic lengths provided that we can construct a kinetic energy, T_α , such that:

$$H = \sum_{\alpha} T_{\alpha} + V$$

$$T_{\alpha} = T_{\alpha}(\mathcal{R}, \dot{\mathcal{R}})$$

Though the significance of such lengths will be limited to our ability to interpret them physically. In the case of the diatomic rotations and vibrations we treat, the motions are well-defined and physically significant even if the species are not chemically isolated.

2.4.2 Diatomic Vibration

If we wish to understand the vibrating motions of a pair of atoms in our system, we can construct the vibrational kinematic length. The vibrational kinetic energy for 2 particles, α and β is:

$$T_{\text{Vib}} = \frac{1}{2} \mu \dot{r}^2 \quad (2.72)$$

Where $r = \|\vec{r}\| = \|\vec{r}_\alpha - \vec{r}_\beta\|$ is the separation between the centers and $\mu = \frac{m_\alpha \cdot m_\beta}{m_\alpha + m_\beta}$ is the reduced mass. In analogy with eq. 2.67, we insert eq. 2.72 into the expression for

the length (eq. 2.67) and have:

$$\ell_{\text{Vib}} = \int_0^t d\tau [\mu \dot{r}^2]^{1/2} \quad (2.73)$$

$$= \sqrt{\mu} \int_0^t d\tau \|\dot{r}\| \quad (2.74)$$

The resulting sum is:

$$\ell_{\text{Vib}} = \sqrt{\mu} \sum_{i=0}^{n-1} \|\delta r^i\| \quad (2.75)$$

where $\delta r^i = r^{i+1} - r^i$.

2.4.3 Diatomic Rotation

We can perform a similar calculation for the rotational degrees of freedom. The rotational kinetic energy for a pair of particles, α and β , is:

$$T_{\text{Rot}} = \frac{1}{2} I \vec{\omega}^2 \quad (2.76)$$

$$= \frac{1}{2} \mu r^2 \left(\frac{d\hat{r}}{d\tau} \right)^2 \quad (2.77)$$

Where $\vec{\omega}$ is the angular velocity, I is the moment of inertia, and r and μ are defined as above. The unit orientation vector is $\hat{r} = \vec{r}/r$ and it is true that $\vec{\omega}^2 = \left(\frac{d\hat{r}}{d\tau} \right)^2$ [60]. Again, we insert the expression for kinetic energy into eq. 2.67 and have:

$$\ell_{\text{Rot}} = \sqrt{\mu} \int_0^t d\tau \left[r^2 \left(\frac{d\hat{r}}{d\tau} \right)^2 \right]^{1/2} \quad (2.78)$$

Now we compute:

$$\frac{d\hat{r}}{d\tau} = \frac{d}{d\tau} \left[\frac{\vec{r}}{\|\vec{r}\|} \right] \quad (2.79)$$

$$= \frac{\dot{\vec{r}}}{r} - \frac{\vec{r}}{r^2} \frac{d}{d\tau} [\|\vec{r}\|] \quad (2.80)$$

$$= \frac{\dot{\vec{r}}}{r} - \frac{\vec{r}}{r^2} \left(\frac{\vec{r} \cdot \dot{\vec{r}}}{r} \right) \quad (2.81)$$

With some re-arrangement, we arrive at a more compact form:

$$\frac{d\hat{r}}{d\tau} = \frac{\dot{\vec{r}}}{r} - \hat{r} \left(\hat{r} \cdot \frac{\dot{\vec{r}}}{r} \right) \quad (2.82)$$

$$= \frac{1}{r} \left[\dot{\vec{r}} - \hat{r} \left(\hat{r} \cdot \dot{\vec{r}} \right) \right] \quad (2.83)$$

$$= \frac{\dot{\vec{r}}}{r} \cdot [\mathbf{1} - \hat{r}\hat{r}] \quad (2.84)$$

where $\mathbf{1}$ is the identity and we have used the dyadic identity $\vec{x}(\vec{x} \cdot \vec{y}) = \vec{y} \cdot \vec{x}\vec{x}$.

With $\frac{d\hat{r}}{d\tau}$ in hand, we can compute the argument to the square root in our integral:

$$r^2 \left(\frac{d\hat{r}}{d\tau} \right)^2 = r^2 \left[\frac{\dot{\vec{r}}}{r} \cdot (\mathbf{1} - \hat{r}\hat{r}) \right]^T \cdot \left[\frac{\dot{\vec{r}}}{r} \cdot (\mathbf{1} - \hat{r}\hat{r}) \right] \quad (2.85)$$

$$= \dot{\vec{r}}^T \cdot (\mathbf{1} - \hat{r}\hat{r})^T \cdot (\mathbf{1} - \hat{r}\hat{r}) \cdot \dot{\vec{r}} \quad (2.86)$$

$$= \dot{\vec{r}}^T \cdot (\mathbf{1} - \hat{r}\hat{r}) \cdot \dot{\vec{r}} \quad (2.87)$$

The tensor terms in the penultimate step can be reduced by simple distribution and the identity: $(\hat{x}\hat{x})^2 = \hat{x}\hat{x}$, which holds for all unit vectors \hat{x} ; see eq. B.21 for derivation. The reduction also follows from the idempotency of projections.

Algebra out of the way, we can write a final expression for the rotational kinematic length:

$$\ell_{\text{Rot}} = \sqrt{\mu} \int_0^t d\tau \left[\dot{\vec{r}}^T \cdot (\mathbf{1} - \hat{r}\hat{r}) \cdot \dot{\vec{r}} \right]^{1/2} \quad (2.88)$$

which is then reduced to the sum:

$$\ell_{\text{Rot}} = \sqrt{\mu} \sum_{i=0}^{n-1} \left[\delta \vec{r}^i{}^T \cdot (\mathbf{1} - \hat{r}^i \hat{r}^i) \cdot \delta \vec{r}^i \right]^{1/2} \quad (2.89)$$

where $\delta \vec{r}^i = \vec{r}^{i+1} - \vec{r}^i$ and $\hat{r}^i = \vec{r}^i / r^i$.

2.4.4 Boundary fraction

In the geodesic approach, the potential energy surface enters into the dynamics only as a constraint on otherwise straight-line paths. It manifests as obstacles in the configuration space, which the geodesics then trace around. In our previous work, studying systems

with many hundreds of degrees of freedom, we found that the geodesics hugged these obstacles for the near-entirety (>99%) of the path [8]. Formaldehyde, with only six internal degrees of freedom, is sufficiently low-dimensional that the obstacles are sparse and geodesics ricochet between them. By measuring the fraction of path spent along the boundaries of these obstacles, we can understand how tortured and constrained the routes taken by geodesics are.

We define the boundary-fraction of a geodesic, $\langle \Theta \rangle$, as the length-fraction of the path that is very near the landscape energy, E_L :

$$\langle \Theta \rangle = \frac{\int_0^t d\tau [2T_{\text{total}}]^{1/2} \Theta[\mathcal{R}(\tau); E_L]}{\int_0^t d\tau [2T_{\text{total}}]^{1/2}} \quad (2.90)$$

where:

$$\Theta[\mathcal{R}; E_L] = \begin{cases} 1 & E_L - V(\mathcal{R}) < \delta E_L \\ 0 & \text{otherwise} \end{cases} \quad (2.91)$$

for suitably small $\delta = 10^{-4}$, Θ is finite only for configurations “on the boundary”. With $\delta = 10^{-5}$ our results are indistinguishable. While we used the total kinematic length as our weighting function, we could have easily used any of the other kinetic energies: diatomic rotation or vibration. When computing the boundary fraction in this other way, the distributions are nearly to within statistical noise.

2.5 Estimating geodesic lengths

To verify our calculations it is desirable to make an analytic estimate of the sizes of the kinematic lengths involved. Here we use a simple model to make such an estimate. These estimates should be considered lower bounds for a given pair of endpoints because they assume no potential obstacles—that the landscape energy is effectively infinite.

2.5.1 Total Kinematic Length

We conceive of our system as having only 2 components: a mass of $m_{\text{H}_2} = 2 \cdot m_{\text{H}}$ located at \vec{r}_{H_2} and a mass of $m_{\text{CO}} = m_{\text{C}} + m_{\text{O}}$ located at \vec{r}_{CO} . Further we will content ourselves to deal only with the translational motions of these fragments and therefore

restrict ourselves to a line. Finally, we assume that all motion is barrier-free—that no potential obstacles intervene, or that for a straight-line path, $V(\vec{R}) \leq E_L$.

Are these reasonable simplifications? They are if the contribution to the kinematic length from translation is much greater than from internal rotation or vibration of H₂ and CO. In particular, we consider only direct dissociation paths, which are relatively unhindered.

Recalling our expression for kinematic length (eq. 2.67), the required kinetic energy of the simplified system is:

$$2T = m_{H_2} \dot{\vec{r}}_{H_2}^2 + m_{CO} \dot{\vec{r}}_{CO}^2 \quad (2.92)$$

If we assume constant velocity during dissociation and a total time t , we can write:

$$2T = m_{H_2} \left| \frac{\Delta \vec{r}_{H_2}}{t} \right|^2 + m_{CO} \left| \frac{\Delta \vec{r}_{CO}}{t} \right|^2 \quad (2.93)$$

where $\Delta \vec{r}_{CO}$ and $\Delta \vec{r}_{H_2}$ are the total distances traveled by CO and H₂ respectively. Fixing our total momentum to zero (implying zero net motion of the center of mass) allows us to write:

$$0 = \vec{p}_{H_2} + \vec{p}_{CO} \quad (2.94)$$

which implies:

$$m_{CO} \frac{\Delta \vec{r}_{CO}}{t} = -m_{H_2} \frac{\Delta \vec{r}_{H_2}}{t} \quad (2.95)$$

and therefore:

$$\Delta \vec{r}_{CO} = -\Delta \vec{r}_{H_2} \frac{m_{H_2}}{m_{CO}} \quad (2.96)$$

Inserting this relation into our expression for $2T$, eq. 2.93 yields:

$$2T = m_{H_2} \left(\frac{\Delta \vec{r}_{H_2}}{t} \right)^2 \left(1 + \frac{m_{H_2}}{m_{CO}} \right) \quad (2.97)$$

Now we only need an expression for $\Delta \vec{r}_{H_2}$, which we can find in terms of $\Delta \vec{r}_{H_2-CO}$, the change in the center of mass separation of H₂ and CO.

$$\Delta \vec{r}_{H_2-CO} = (\vec{r}_{H_2} - \vec{r}_{CO})_f - (\vec{r}_{H_2} - \vec{r}_{CO})_i \quad (2.98)$$

$$= \Delta \vec{r}_{H_2} - \Delta \vec{r}_{CO} \quad (2.99)$$

$$= \Delta \vec{r}_{H_2} \left(1 + \frac{m_{H_2}}{m_{CO}} \right) \quad (2.100)$$

where we have used eq. 2.96 to eliminate $\Delta\vec{r}_{CO}$. Using eq. 2.100, we eliminate $\Delta\vec{r}_{H_2}$ from $2T$ in eq. 2.97, which yields:

$$2T = \left(\frac{\Delta\vec{r}_{H_2-CO}}{t} \right)^2 \frac{m_{H_2}}{\left(1 + \frac{m_{H_2}}{m_{CO}} \right)} \quad (2.101)$$

$$= \left(\frac{\Delta\vec{r}_{H_2-CO}}{t} \right)^2 \mu_{H_2-CO} \quad (2.102)$$

where μ_{H_2-CO} is the reduced mass for H_2 and CO , defined in the usual way. Now our expression for the total length is:

$$\ell_{Total} = \int_0^t d\tau \left(\frac{\|\Delta\vec{r}_{H_2-CO}\|}{t} \right) \sqrt{\mu_{H_2-CO}} \quad (2.103)$$

which is readily integrated to yield:

$$\ell_{Total} = \|\Delta\vec{r}_{H_2-CO}\| \sqrt{\mu_{H_2-CO}} \quad (2.104)$$

Using the appropriate masses from table 2.1 gives $\mu_{H_2-CO} = 3427.5 m_e$. We estimate $\|\Delta\vec{r}_{H_2-CO}\|$ to be of order $10 a_0$ because the molecular dynamics trajectories used to produce endpoints for the geodesics were terminated when the centers were separated by $12 a_0$. Taken together, we estimate the lower bound on total kinematic length to be in the range of:

$$\ell_{Total} \gtrsim 585 a_0 \sqrt{m_e} \quad (2.105)$$

2.5.2 Vibrational Kinematic Length

An identical analysis can be preformed for vibrational degrees of freedom. Using the expression for T_{Vib} from eq. 2.72 we have:

$$2T = \mu \dot{r}^2 \quad (2.106)$$

Again using the assumption of constant velocity yields:

$$\dot{r} = \frac{\Delta r}{t} \quad (2.107)$$

Which we can combine with the expression for vibrational kinematic length, eq. 2.73:

$$\ell_{Vib} = \int_0^t d\tau \left(\frac{\|\Delta r\|}{t} \right) \sqrt{\mu} \quad (2.108)$$

The expression is readily integrated to give:

$$\ell_{\text{Vib}} = \|\Delta r\| \sqrt{\mu} \quad (2.109)$$

That this result for diatomic vibration is quite similar to the estimate we made for total kinematic length, eq. 2.104, is no accident. The previous analysis of the total kinematic length effectively treated formaldehyde as single large dissociating diatomic. Such a simplification was reasonable because we expect the internal motions to be of secondary importance relative to the overall motion of the dissociating fragments.

In the specific case of H_2 vibration, we have $\mu = 918.56 m_e$. We posit $\|\Delta \vec{r}\|$ —the total straight-line motion of the vibrating H_2 —is of order $1 a_0$ and therefore estimate the vibrational length as:

$$\ell_{\text{Vib}} \gtrsim 30 a_0 \sqrt{m_e} \quad (2.110)$$

2.6 A search for saddles

The existence of a S_0/S_1 conical intersection for formaldehyde was first reported by Araujo *et al.* in 2008 [65] and brought to our attention by one of Harding and Kilpenstein’s papers [36] investigating the putative “roaming” saddle. We were curious if this conical intersection would be evident in the fitted ground state potential surface for formaldehyde provided to us by Joel Bowman [54]. If it were, we would expect to find a second order saddle (a saddle with 2 “downhill” directions or 2 negative force constants) in its vicinity. The saddle does indeed exist; in this section, we detail the methods of its finding and of our unsuccessful search for the roaming saddle on the Bowman potential. We describe the characterization of the saddles by frequency analysis in Section 2.7.3.

2.6.1 Methods

There a variety of methods for locating saddles of arbitrary order on potential energy surfaces. The more sophisticated schemes such as Miller’s “eigenvector following” [66] rely on inverting the Hessian matrix. This poses difficulties when working in Cartesian coordinates due to the zero-frequency modes associated with translation. These can be worked-around using projection operators [67], but complicates the matter further.

A method for finding saddles, which in general is *bad*—but in certain, limited cases may be appropriate—is to simply minimize the square of the gradient of the potential. This works at all because $|\nabla V(\underline{R})|^2 = 0$ for all stationary points of a potential, $V(\underline{R})$. This is a bad idea for large, global searches because there may be local minima of $|\nabla V(\underline{R})|^2$, which interfere with locating the global extrema, all identically 0. In our case, the idea isn't terrible because we expect to start with a good guess for the location of the saddle. It even turns out to work.

Minimization Function & Algorithm

The object of investigation is the square of the gradient of the potential function provided by Joel Bowman, but we lack an analytic form of this fit or its gradient. To compute the gradient, we proceed via finite-differencing as described in Section 2.7.2.

To minimize such a finite-differenced function as $|\nabla V(\underline{R})|^2$, it is wise to avoid a method requiring a further gradient. This avoids accumulation of errors by multiple numeric differencing schemes. Nelder & Mead's simplex algorithm [68] is a good, general purpose algorithm which does not require a gradient. We employ the faithful implementation of their algorithm in the GNU Scientific Library [62].

While the simplex method of minimization is performant and was effective, after the fact, I noticed the following, perhaps-useful relation. Some straightforward vector algebra gives the gradient of the square of the gradient of a function as the simple product of the gradient and the Hessian:

$$\nabla \left[(\nabla V)^2 \right] = \nabla \left[\nabla V \cdot \nabla V \right] = 2 \nabla \nabla V \cdot \nabla V \quad (2.111)$$

One can see that $\nabla \nabla V$ is the Hessian by interpreting $\nabla \nabla$ as a dyad. Of course this requires estimating the Hessian at each step, which isn't cheap. And while this is better than twice applying a finite difference, it is still the (noisier) product of two noisy (finite-differenced) objects.

Points of departure

To locate the second order saddle associated with Araujo's conical intersection, we applied the simplex minimization scheme described in the previous section starting from the published coordinates (given in Table 3.2). To search for the roaming saddle,

we generated 10^5 starting coordinates by taking randomly oriented steps of random length uniformly distributed between 0 and $3 a_0$ away from the proposed location of the roaming saddle (given in Table 3.9) and then applied the same search to the given location and the 10^5 other trial points. Section 3.2 details our findings and reports on the newly located saddle.

2.7 Numerical methods

As is typical of numeric calculations on machines with finite precision, comparison operations were made to within specified tolerances. In all cases the condition for equality was less than or equal to one part in 10^6 . Double precision (64-bit) floats were used throughout. Vector operations were performed using the BLAS linear algebra subroutines via the GNU Scientific Library [62] in the C language with double precision on an x86_64 machine. Data analysis was implemented in a mixture of C, Python (with NumPy [69]), and Haskell [70] (using the `hmatrix` package for vector operations). Wherever required, random numbers were produced using the MT19937 variant [71] of the Mersenne Twister algorithm implemented in the GNU Scientific Library. Parallel data processing was aided by the use of GNU Parallel [72]. A considerable amount of the source code has been placed in the Brown University Digital Repository [73]. The remainder of this section gives the details of the finite-difference calculus employed throughout the study.

2.7.1 With respect to integration

Computing properties of the geodesic often requires computing integrals over functions of the kinetic energy. For example, the total kinematic length, ℓ_{total} , in terms of the total kinetic energy, T_{total} , is:

$$\ell_{\text{total}} = \int_0^t d\tau [2T_{\text{total}}]^{1/2} \quad (2.112)$$

where τ is an arbitrary progress variable. Because of the algorithm we use to compute the geodesics (Section 2.2), we have configurations of the system, \vec{x} , for unevenly spaced

$\{\tau_i\}$. All such integrals were performed using the following Riemann sum:

$$\int_{\tau_0}^{\tau_n} d\tau f \left[\vec{x}(\tau), \dot{\vec{x}}(\tau) \right] \approx \sum_{i=0}^{n-1} \delta\tau_i f \left[\vec{x}_i, \dot{\vec{x}}_i \right] \quad (2.113)$$

where $\delta\tau_i = \tau_{i+1} - \tau_i$ and $\dot{\vec{x}}_i = \frac{d\vec{x}}{d\tau} \approx (\vec{x}_{i+1} - \vec{x}_i)/(\tau_{i+1} - \tau_i) = \delta\vec{x}_i/\delta\tau_i$ is the rate of change of the configuration space vector \vec{x} computed via simple forward difference. The geometric picture of this simple technique corresponds to summing the distances between adjacent configuration space points with the appropriate mass weighting. While higher order techniques for integration and the required derivatives exist, we have directly verified that the errors introduced by this method are several orders of magnitude smaller than the statistical noise associated with averaging over many paths. Additionally, these “higher order” methods loose the simple geometric significance of summing over mass-weighted path lengths.

Given that there is no physical content in τ , properties of the geodesics should be independent of it. In all integrals for the kinematic length, $\delta\tau_i$ drops out. We show this explicitly in the case of the total kinematic length. Insert $T_{\text{total}} = \frac{1}{2} \dot{\vec{R}}^T \mathbf{M} \dot{\vec{R}}$ and eq. 2.112 becomes:

$$\ell_{\text{total}} = \int d\tau \left[\dot{\vec{R}}^T \mathbf{M} \dot{\vec{R}} \right]^{1/2} \quad (2.114)$$

where \mathbf{M} is the diagonal mass matrix given in eq. 2.69. Expressing our integral (eq. 2.114) as a sum (via eq. 2.113) yields:

$$\ell_{\text{Total}} = \sum_{i=0}^{n-1} \delta\tau_i \left[\dot{\vec{R}}_{\sim i}^T \mathbf{M} \dot{\vec{R}}_{\sim i} \right]^{1/2} \quad (2.115)$$

$$= \sum_{i=0}^{n-1} \delta\tau_i \left[\frac{\delta R_i^T}{\delta\tau_i} \mathbf{M} \frac{\delta R_i}{\delta\tau_i} \right]^{1/2} \quad (2.116)$$

$$= \sum_{i=0}^{n-1} \left[\delta R_i^T \mathbf{M} \delta R_i \right]^{1/2} \quad (2.117)$$

And we are indeed left with an expression independent of τ .

2.7.2 Gradients

While the potential described in Section 2.1.1 has an analytic form, extracting it from the source code would have been prohibitively tedious. We therefore treat it as a black

box and compute gradients numerically. Partial derivatives are computed along each of the 12 Cartesian, lab-frame coordinates defining the configuration space. We use the 5-point formula from Abramowitz and Stegun [61, eq. 25.3.6]:

$$\frac{\partial f}{\partial x_i} \Big|_{\vec{x}} = \frac{1}{12} (f_{-2} - 8f_{-1} + 8f_{+2} - f_{+2}) + \mathcal{O}(h^5) \quad (2.118)$$

where $f_j = f(\vec{x} + jh\hat{x}_i)$. We take the inter-abscissa spacing, $h = 10^{-4} a_0$.

2.7.3 Hessian matrices & vibrational frequencies

While Abramowitz and Stegun give a 5th order expression for $\partial^2 F / \partial x^2$, the off diagonal terms are not treated with the same level of precision. David Eberly gives a clear description of finite differencing schemes and their derivation for uni- and multi-variate functions [74], which was followed to construct the approximation we use here. The key result is that repeated application of orthogonal finite differencing methods (e.g., first along x then y) give rise to expressions where the coefficients are the outer products of the coefficients for each approximation.

$$\frac{\partial^2 F}{\partial x \partial y} = \mathcal{O}(h^4) + \frac{1}{144h^2} \sum_{i,j=-2}^2 (\mathbf{C}^{1,1})_{ij} F_{ij} \quad (2.119)$$

where $F_{i,j} = F(x + ih, y + jh)$ and the coefficients are given by:

$$\mathbf{C}^{1,1} = \begin{bmatrix} 1 & -8 & 0 & 8 & -1 \\ -8 & 64 & 0 & -64 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 8 & -64 & 0 & 64 & -8 \\ -1 & 8 & 0 & -8 & 1 \end{bmatrix} \quad (2.120)$$

Indexes on $\mathbf{C}^{1,1}$ are relative to the center. As hinted, the coefficients for this mixed derivative expression arise from the outer product of those for a first derivative: $\mathbf{C}^{1,1} = \mathcal{C}^1 \otimes \mathcal{C}^1$ where $\mathcal{C}^1 = [1, -8, 0, 8, -1]$. And indeed, to 5th order,

$$\frac{dF}{dx} = \mathcal{O}(h^5) + \frac{1}{12h} \sum_{i=-2}^2 (\mathcal{C}^1)_i F_i$$

with $F_i = F(x + ih)$.

The diagonal elements of the Hessian are taken from Abramowitz and Stegun, eq. 25.3.24 [61]:

$$\frac{\partial^2 F}{\partial x^2} = \mathcal{O}(h^4) + \frac{1}{12h} \sum_{i=-2}^2 (\zeta^2)_i F_i \quad (2.121)$$

with $F_i = F(x + ih, y)$ and $\zeta^2 = [-1, 16, -30, 16, -1]$.

To calculate frequencies, we diagonalize the mass weighted Hessian given by,

$$\frac{\partial^2 V}{\partial r_{k\alpha} \partial r_{l\beta}} \frac{1}{m_k^{1/2} m_l^{1/2}} , \quad k, l = 1 \dots N \quad \alpha, \beta = x, y, z \quad (2.122)$$

where $r_{k\alpha}$ and $r_{l\beta}$ are the lab-frame, Cartesian coordinates of the N -atom system. The eigenvalues are the mass-weighted force constants, k_i , and can be converted to frequencies in wavenumbers, $\tilde{\nu}_i$, by the relation

$$\tilde{\nu}_i = \frac{\sqrt{k_i}}{2\pi c}$$

where c is the speed of light. For formaldehyde, there are six vibrational modes and so the six frequencies with largest magnitude are assigned to vibrational modes. We assign the lowest three to net translation and the intervening three to net rotation.

We use the 4th order expressions for all elements of the Hessian matrix given in eqs. 2.119 and 2.121. A step-size of $h = 7 \times 10^{-3}$ was empirically determined to minimize the magnitude of translational and rotational modes (see Figure 2.6). If we were trying to be still more careful, we would choose h such that it could be represented exactly in base-2 (and therefore exactly representable on a binary computer). The implementation of these methods was verified against analytic calculations for Morse and harmonic oscillators with a variety of frequencies.

Switching functions?

To stitch together the multi-reference and single-reference portions of their fit to the formaldehyde potential, the Bowman group employed switching functions of the form¹⁶:

$$w(x) = \begin{cases} 1 & x \leq 0 \\ 1 - 10x^3 + 15x^4 - 6x^5 & 0 < x < 1 \\ 0 & 1 \leq x \end{cases}$$

where x is a dimensionless variable corresponding to the scaled distance of the farthest hydrogen from its nearest neighbor (see [54] for details). At $x = 0, 1$, $w(x)$ is only differentiable to second order and so higher derivatives of V are discontinuous across the hyper-planes defined by $x = 0$ and $x = 1$. Accordingly, approximations to derivatives of V relying on such continuities (e.g. eqs. 2.119 and 2.121) will fail if these planes intersect the sampling region of the approximation, a hyper-cube of side length $4h$ centered on the point where the derivative is to be approximated. The regions where we compute frequencies are not expected to be near the boundaries of the switching function. If we were computing the hessian in the roaming region (perhaps in the vicinity of a roaming saddle), we would need to be more careful.

In one instance, computed frequencies differed slightly (< 2.5%) from the published values (see Section 3.2.1). To rule out interference from the switching function, we repeated the calculation evaluating the elements of the Hessian by a 2nd order approximation from Abramowitz and Stegun [61] and found the computed frequencies to be within 0.01% of the first.

¹⁶See Appendix E.1 for a note on a mistake in their published version.

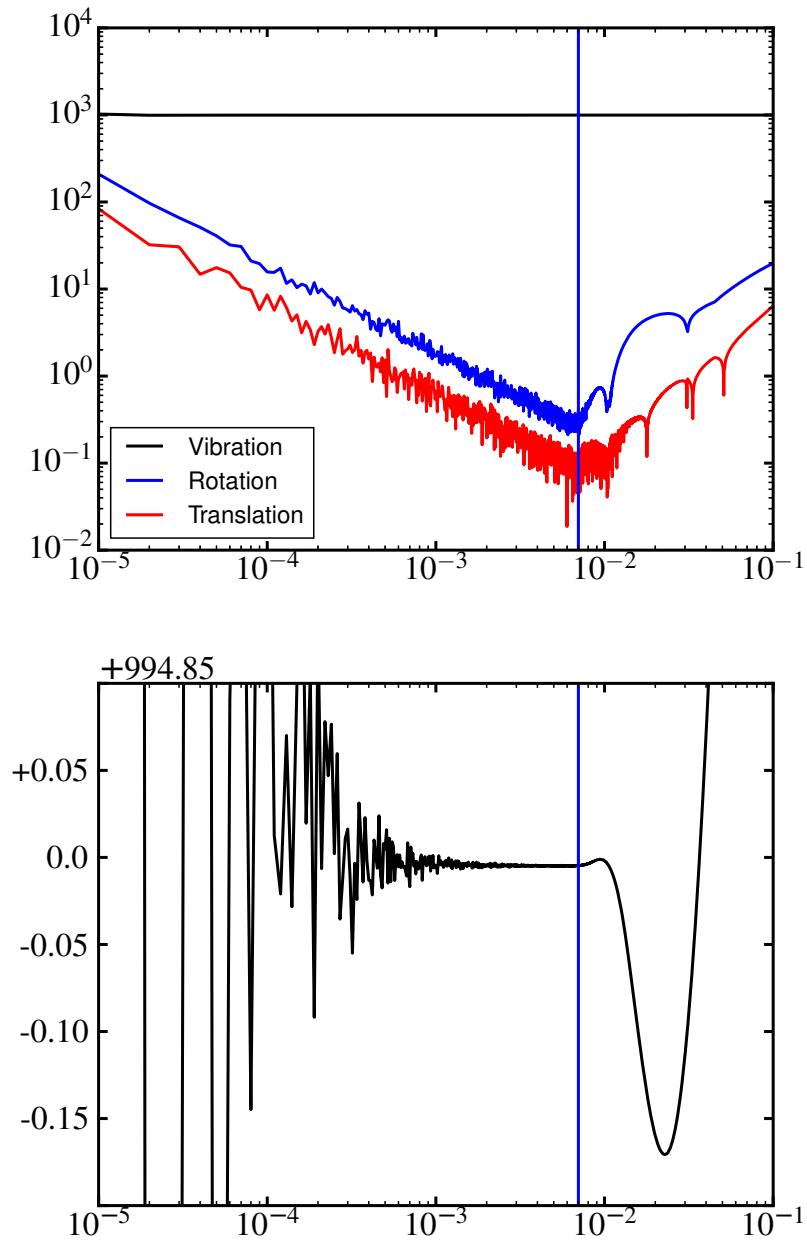


Figure 2.6: Hessian parameters

Results from computing normal modes at the second order saddle via the Hessian, eqs. 2.119 and 2.121. Shown are geometric averages of the vibrational, rotational, and translational normal modes (in wavenumbers) as a function of abscissa-spacing, h in a_0 . The vertical line shows the selected value of $h = 7.00 \times 10^{-3} a_0$, which is used throughout the investigation. The lower panel is a very tight zoom on the vibrational modes and shows that the vibrational modes are converged and stabilized for this value of h . Recall, from the upper panel, that the vibrational modes are grossly well-behaved over the whole range of values.

2.8 Computing the entropy of a sampled distribution

Part of our analysis will involve computing the entropy of continuous distributions of unknown form, but from which we can draw samples. The entropy of a continuous distribution is given by [75]:

$$S/k_B = - \int dx \rho(x) \ln [\rho(x)] \quad (2.123)$$

Supposing that one had no access to the distribution, $\rho(x)$, but could sample it, there are a variety of ways that one could imagine constructing an estimate of S/k_B from the samples $\{x_i\}$. One general approach is to construct a model of the distribution (via histogramming or kernel density estimation) and then to compute a numeric integral for eq. 2.123. This is attractive because it yields the intermediate representation of the distribution itself. However, these methods require tuning of parameters (bin or kernel width) and the computed entropy is sensitive to them. The estimate will also be sensitive to the particulars of the numeric integration employed.

An alternative approach, which directly constructs the entropy from the samples, is given by Henri Theil [76]. Given a set of n observations of real numbers, $\{x_i\}$, it is possible to construct the distribution satisfying a maximum entropy principle for the observations. That is, the distribution with maximum entropy consistent with the observations. Qualitatively, this maximum entropy distribution will be composed of many intervals of uniform density. The entropy of this maximum entropy distribution can then be had analytically as:

$$S_0/k_B = \frac{2 - 2 \ln 2}{n} + \frac{1}{n} \sum_{i=1}^n \ln \left[\frac{n}{2} \left(x^{i+1} - x^{i-1} \right) \right] \quad (2.124)$$

where $x^0 = x^1$, $x^{n+1} = x^n$ and $x^1 < x^2 < \dots < x^i < \dots < x^n$ is the ordered sequence (or order statistics) of the set $\{x_i\}$. The method is only appropriate for samples drawn from a continuous distribution.

The Theil method, along with numeric integration over histograms and kernel density estimates were compared to the analytic entropies for densely sampled Gaussian and uniform distributions of many thousands of widths. The method given in eq. 2.124 had lower relative error in all cases and is the method we use to compute the entropies of length distributions throughout. It should be noted that eq. 2.124 takes the log of

differences of adjacent order statistics. If any two samples are identical, the result will be divergent¹⁷. We use double precision numbers throughout this work and with 64 bits of precision, the probability of violating the strict inequality requirement is infinitesimal.

For any dimensioned distribution, $\rho(x)$ will require normalization in the \ln of eq. 2.123. The change of dimension will introduce an unknown constant offset in the entropy. This requirement is related to the reason that there is no absolute entropy scale classically—without quantized observables, there is no absolute “smallest” division and therefore no way to count up states. As a consequence, the absolute entropies themselves will not be meaningful, but differences will be.

¹⁷Such divergence, of course, is consistent with the maximum entropy principle used to construct the estimate.

CHAPTER 3

RESULTS AND DISCUSSION

In this chapter we describe the results of our investigations into the dynamics of roaming on the formaldehyde potential energy surface. The chapter is split sharply between two different perspectives: a local view grounded in classical trajectories and stationary points, and a global view, aiming to reveal the inherent dynamics in terms of geodesics and their properties.

The first section reports on our search for several stationary points of the surface. We successfully locate a previously unreported second order saddle. However, we could not locate another, previously reported stationary point—the “roaming saddle”. We discuss the implications of their respective presence and absence. We then move on to observations provided by molecular dynamics techniques, which serve as a point of departure and comparison for understanding the global information carried by geodesics.

In Section 3.4, we begin our discussion of geodesics by validating some of their basic properties and showing how the information they contain aligns with our notion of inherent dynamics. The geodesics themselves provide novel insight to the roaming problem on the formaldehyde surface. Our major result has to do with the entropy associated with the each of the dynamical channels available to the fragmenting molecule. From the relative entropy of these channels we are able to predict the decline in the incidence of roaming at high energy.

3.1 An image of the formaldehyde potential

To orient ourselves to the space, in Figure 3.1, we plot a slice of the potential surface along with several stationary points. We constructed this particular projection in terms of a set of Jacobi coordinates: $\{r_{\text{H-H}}, r_{\text{H}_2\text{-CO}}, r_{\text{C-O}}, \theta_1, \theta_2, \phi\}$ where $r_{\text{H-H}}$, $r_{\text{H}_2\text{-CO}}$, and $r_{\text{C-O}}$ are H–H, H₂–CO, and C–O separation respectively, θ_1 is the angle between the vectors $\vec{r}_{\text{H-H}}$ and $\vec{r}_{\text{H}_2\text{-CO}}$, θ_2 is the angle between the vectors $\vec{r}_{\text{C-O}}$ and $\vec{r}_{\text{H}_2\text{-CO}}$, and ϕ is the out-of-plane torsional angle. To compute the image shown, we varied $r_{\text{H-H}}$ and $r_{\text{H}_2\text{-CO}}$ and fixed the other coordinates at their equilibrium values: $r_{\text{C-O}} = 2.2883 \text{ \AA}_0$, $\theta_1 = \pi/2$, $\theta_2 = \pi$, $\phi = 0$.

Note that because of our choices for the projection, some features of the surface are not visible. For instance, this particular “shadow” of the 6-dimensional object shows neither the *cis* nor *trans* isomers of HCOH even though they are present on the surface. For a similar reason, it appears that some of the locations of the stationary points do not match up with the potential. The tight transition state to molecular products, \square , is displaced slightly to the left of the saddle clearly visible on the surface. In two dimensions, saddles of degree two appear as the peaks of hills, and the shadow of the second order saddle we locate in Section 3.2, is in fact the hill to the right of its marker, \diamond . The Global minimum, \bigcirc , appears accurate to the potential contours because all coordinates are fixed to the global minimum at that point. The putative roaming saddle, \triangle , is located in an extremely flat region—the spacing between contours is an order of magnitude closer there than on the rest of the surface.

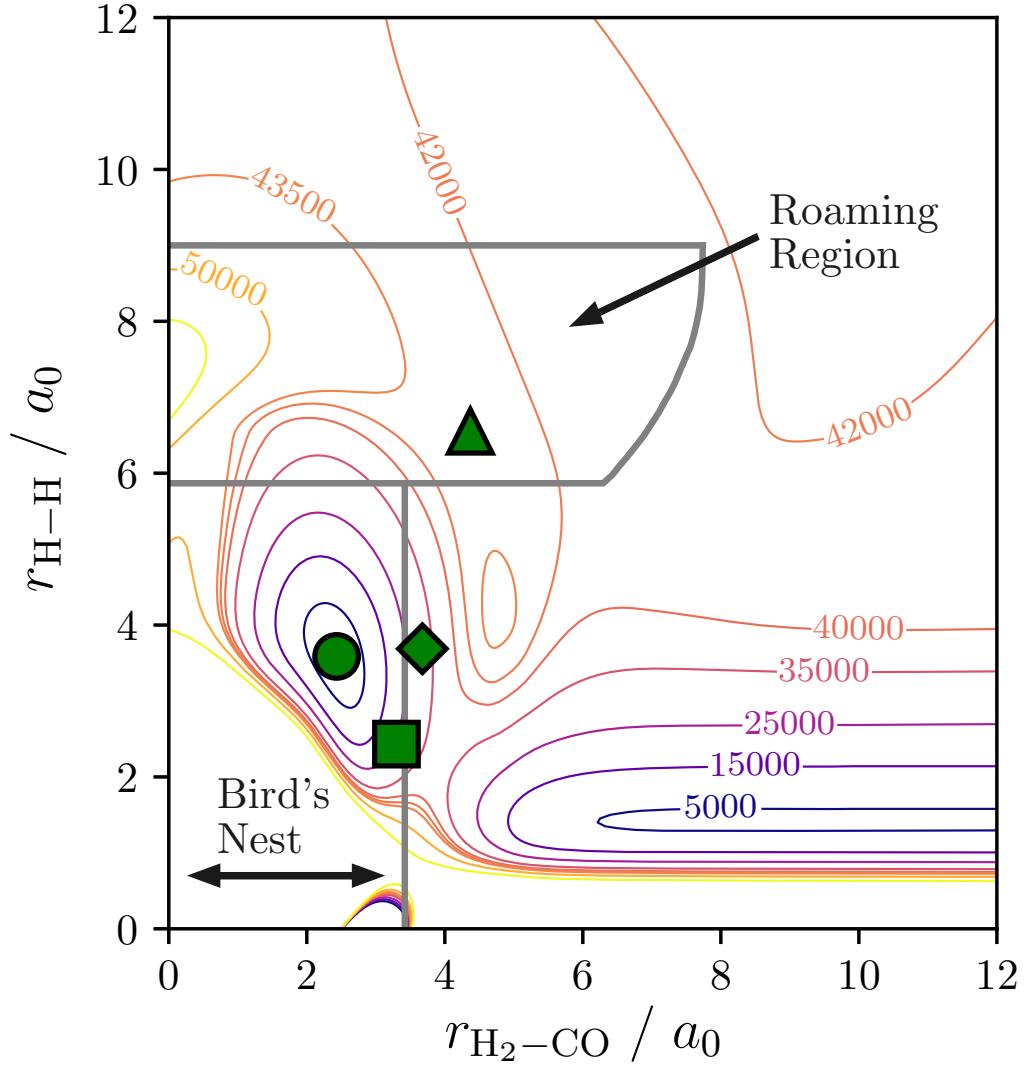


Figure 3.1: A projection of the potential surface

Vertical and horizontal axes are respectively H–H and H₂–CO separation in atomic units. The circle, \circ , is the global minimum of the surface. The square, \square , is the tight transition state to molecular products from the global minimum. The triangle, \triangle , is the location of the putative roaming saddle. The diamond, \diamond , is the newly-located saddle of degree two associated with the S_0/S_1 conical intersection; see Section 3.2. The colors of the contours do not change much in the upper right quadrant; this is intentional because the potential is so flat in this region. The roaming region is defined in Section 2.3.3; the bird’s nest will be defined and discussed in Section 3.3.1.

3.2 Locating a 2nd order saddle on the Bowman energy surface

In Section 2.6, we describe the simplex-minimization-based technique to search for saddles on the Bowman potential energy surface. Our targets are two saddles that had not been previously located on the potential:

1. We posit the existence of and search for a saddle of degree two in the vicinity of the S_0/S_1 conical intersection¹ first reported by Araujo and co-workers [65].
2. We also search for the so-called “roaming saddle”, reported by Harding *et al.* [31], which some believe (see Section 1.2.2) governs the rates of roaming reaction in the usual fashion of saddles in transition state theory. If such a saddle is essential to roaming, we expect to find one on the Bowman surface, upon which roaming certainly manifests.

We seek the “roaming saddle” because of the significance it takes in some theories of roaming (again, see Section 1.2.2). If it is not present on the Bowman surface, such a feature cannot be essential to our understanding. We search for the second order saddle because, if it exists, we expect it to form part of ridge bounding the bird’s nest (see Section 3.3.1).

3.2.1 Verification

We are fortunate to have available for comparison, literature values for the properties several stationary points on the Bowman surface. The description of the surface from Zhang *et al.* [54] lists stable minima and first order saddles with their energies and normal frequencies. For each of the stationary points given in their paper, we confirm the location by way of the minimization scheme described in Section 2.6 and we give published and computed values for frequencies and energy. As measures of

¹Conical intersections break the degeneracy of the intersecting states along two coordinates [77]. When constructing a high-fidelity adiabatic energy surface for the lower state, the intersection may manifest as a stationary point with two imaginary frequencies (two “down hill” directions). In a 1- or 2-degree-of-freedom system, such a feature would appear as a maximum, but in our case, there are 4 other internal degrees of freedom. These will have real frequency and so the stationary point is a saddle. See [36] for more in the case of formaldehyde.

accuracy, we also give the magnitude of the gradient and the root sum squared atomic displacements between our initial guess and the final location, for each stationary point. We define,

$$\Delta R = |\mathcal{R} - \mathcal{R}_0| = \sqrt{\sum_{i,\mu} (R^{i\mu} - R_0^{i\mu})^2} \quad \mu = \{x, y, z\}, i = 1 \dots N \quad (3.1)$$

where \mathcal{R}_0 is the published configuration from which the search was initiated and \mathcal{R} is the point we located. ΔR thereby provides a measure of the distance from the location of the start of the search to the stationary point which was found.

Table 3.1 gives the named stationary points (St.P) along with computed and published [54] values for comparison. In each pair of rows, the upper row holds the computed results and the lower row holds the published values (in parentheses). In all instances where computed frequencies differ from those published (excepting for TS_3 , see caption), we suspect it is because of rounding; had we simply truncated the values, they would be identical.

While the frequency deviations for the transition state to molecular products are small (< 2.5%), we were still perplexed and investigated two possible sources of error. It is possible that the choice for inter-abscissa spacing, $h = 7 \times 10^{-3}$, was improper for TS_3 on the potential surface. However, we discarded this possibility after constructing graphs similar to those in Figure 2.6 to test for convergence—the frequencies at TS_3 were even *better* converged with our choice of h than for the second order saddle. It is also possible that there would be interference from the switching function (see page 56) if the hyper-volume within which the function was evaluated contained a discontinuity in the 4th derivative (as exists at the boundaries). We also dispensed with this theory after using a 2nd order approximation to evaluate the elements of the Hessian; the computed frequencies were within 0.01% of the first. We suspect that the frequencies we provide in Table 3.1 are correct for the surface and the published values are in error.

3.2.2 Search results

Beginning our search from the location of the S_0/S_1 conical intersection reported by Araujo *et al.* [65] and minimizing $|\nabla V(\mathcal{R})|^2$, our simplex minimization method unambiguously locates a second order saddle point. We converge such that the vertices

Stationary Point	$V(R)/\text{cm}^{-1}$	$ \nabla V(R) /(\text{cm}^{-1}/\text{\AA})$	$\Delta R/\text{\AA}$	$\{\tilde{\nu}_i\}/\text{cm}^{-1}$
Global Min	3.36×10^{-6} (0.00)	2.05×10^{-3}	7.62×10^{-6}	$1165, 1337, 1624, 1807, 2811, 2950$ (1164, 1337, 1623, 1806, 2810, 2949)
trans-HOCH	17983.37 (17983.37)	0.595×10^{-3}	3.66×10^{-4}	$1108, 1282, 1369, 1715, 2927, 4099$ (1108, 1281, 1369, 1714, 2927, 4098)
cis-HOCH	19517.27 (19517.27)	1.14×10^{-3}	1.12×10^{-3}	$1221, 1343, 1440, 1609, 2741, 3839$ (1221, 1342, 1440, 1608, 2741, 3839)
TS₁	27335.32 (27335.33)	1.63×10^{-3}	5.82×10^{-4}	$1074i, 967, 1208, 1465, 2542, 3531$ (1074i, 968, 1208, 1465, 2541, 3531)
TS₂	30140.00 (30140.00)	1.17×10^{-3}	1.02×10^{-3}	$2068i, 726, 1260, 1454, 2599, 2993$ (2068i, 725, 1259, 1454, 2598, 2992)
TS₃	30583.66 (30583.66)	0.778×10^{-3}	7.30×10^{-4}	$1825i, \mathbf{612}, \mathbf{659}, 1357, 1864, \mathbf{2994}$ (1824i, 598, 652, 1351, 1861, 3039)
2nd order saddle	33777.20	2.42×10^{-3}	0.560	541.2i, 1165.3i, 313.0, 935.9, 1862.0, 2818.5

Table 3.1: Properties of stationary points on the Bowman surface

Comparison of various published [54] and computed properties (see text) of stationary points on Bowman's surface. The first row, are our results. The second rows give are published values within parentheses. The configurations of the stationary points were published [54] in terms of angles and bond lengths to a precision of one-tenth degree and one-ten-thousandth Angstrom. TS₁ is between the cis and trans isomers; TS₂ is between the global minimum and the *trans* isomer; and TS₃ is between the global minimum and molecular products, the "tight" transition state. TS₃ is the only example of departure from the expected values; deviations larger than 1 % are **bolded**. Energies all agree with published values to 6 significant figures. We give $|\nabla V(R)|$ as a measure of the accuracy with which we have located the stationary points; it is quite good in all cases. The location of the conical intersection from which we found the second order saddle was published in [65]; this stationary point is the only for which the search distance, ΔR (eq. 3.1), is non-trivial.

of the simplex are within, $10^{-14} a_0$, of their centroid². At the saddle, the gradient's magnitude is: $\nabla V(\underline{R}) = 5.83 \times 10^{-9} E_h/a_0 = 2.42 \times 10^{-3} \text{ cm}^{-1}/\text{\AA}$. This value is consistent with our findings for other, previously published stationary points on the surface (see Table 3.1). The presence of the second order saddle in the vicinity of a conical intersection is inherently a multi-electronic-surface result that manifests under the use of the Born-Oppenheimer approximation in the construction of our adiabatic potential surface. It is non-trivial that the Bowman group's fit [54] is able to capture the feature. The saddle's energy, gradient, and frequencies are listed in Table 3.1, which shows there are indeed two imaginary frequencies. The other six normal modes, corresponding to net -translation and -rotation, all have frequency of magnitude less than $1.0 \text{ cm}^{-1}/\text{\AA}$, which is of the same magnitudued as the rotational and translational modes for the other, previously published stationary points. The coordinates of the second order saddle are given in Table 3.2 along with those of the conical intersection. The coordinate system is chosen such that all of the free variables are 0. We place C at the origin, O along the y axis, and orient such that HCO is in the x – y plane. The hydrogen excluded from the formyl is the farthest away from the total center of mass. Distances are given in Bohr. Coordinates for all other stationary points of interest are given in the same coordinate system in Tables 3.3–3.9.

3.2.3 No “roaming saddle” to be found

Using the same methods, we searched the Bowman surface extensively for the reported “roaming saddle” [31]. We initiated searches from the proposed location [35] of the roaming saddle and perturbed its position by taking random steps in a 12-dimensional ball of up to several a_0 from the starting position (see page 51). However, even in trials of 10^5 such searches, we were unable to converge the gradient to less than $10^2 \text{ cm}^{-1}/\text{\AA}$, which we deemed insufficiently flat—indeed, the gradients at the other stationary points we located have magnitude 5 orders of magnitude smaller. Such searches also suffered in the following ways:

- Of those points with “shallowly” sloping gradient, none unambiguously possessed

²The measure of vertex displacement from centroid is an indicator of the degree of convergence of our algorithm—in this case, quite high relative to features of any molecular energy landscape. This value is consistent with those we found for the other stationary points on the surface.

2nd order saddle				
	x	y	z	
$\vec{r}_{H^{(1)}}$ =	1.672984	-1.508401	3.683196	
\vec{r}_O =	0.0	2.222303	0.0	
\vec{r}_C =	0.0	0.0	0.0	
$\vec{r}_{H^{(2)}}$ =	1.485300	-1.472777	0.0	
S_0/S_1 conical intersection				
	x	y	z	
$\vec{r}_{H^{(1)}}$ =	0.935657	-1.296518	3.332712	
\vec{r}_O =	0.0	2.257969	0.0	
\vec{r}_C =	0.0	0.0	0.0	
$\vec{r}_{H^{(2)}}$ =	1.238146	-1.679863	0.0	

Table 3.2: Coordinates of second order saddle on Bowman's surface

Given first are the coordinates of the saddle, which we located located on Bowman's surface [54]. Below, are the published coordinates [65] (rotated and translated into the same reference frame) of the conical intersection from which the search for the saddle was initiated.

	x	y	z
$\vec{r}_{H^{(1)}}$ =	-1.793899	-1.124608	0.0
\vec{r}_O =	0.0	2.288201	0.0
\vec{r}_C =	0.0	0.0	0.0
$\vec{r}_{H^{(2)}}$ =	1.793899	-1.124608	0.0

Table 3.3: Coordinates of global minimum on Bowman's surface

Location refined and coordinates transformed from the published configuration [54].

	x	y	z
$\vec{r}_{H^{(1)}}$ =	2.030988	-0.502804	0.0
\vec{r}_O =	0.0	2.513517	0.0
\vec{r}_C =	0.0	0.0	0.0
$\vec{r}_{H^{(2)}}$ =	1.726049	3.153715	0.0

Table 3.4: Coordinates of *cis*-HCOH minimum on Bowman's surface

Location refined and coordinates transformed from the published configuration [54].

	x	y	z
$\vec{r}_{\text{H}^{(1)}}$ =	-2.070234	-0.327514	0.0
\vec{r}_{O} =	0.0	2.513931	0.0
\vec{r}_{C} =	0.0	0.0	0.0
$\vec{r}_{\text{H}^{(2)}}$ =	1.747948	2.997965	0.0

Table 3.5: Coordinates of *trans*-HCOH minimum on Bowman's surface

Location refined and coordinates transformed from the published configuration [54].

	x	y	z
$\vec{r}_{\text{H}^{(1)}}$ =	0.208602	-0.724582	2.013021
\vec{r}_{O} =	0.0	2.510365	0.0
\vec{r}_{C} =	0.0	0.0	0.0
$\vec{r}_{\text{H}^{(2)}}$ =	1.516407	3.562945	0.0

Table 3.6: Coordinates of transition state for *cis-trans* isomerizationAlso on the Bowman Surface; in their papers, referred to as TS₁. Location refined and coordinates transformed from the published configuration [54].

	x	y	z
$\vec{r}_{\text{H}^{(1)}}$ =	-1.893152	-0.852520	0.0
\vec{r}_{O} =	0.0	2.490956	0.0
\vec{r}_{C} =	0.0	0.0	0.0
$\vec{r}_{\text{H}^{(2)}}$ =	1.920528	1.378915	0.0

Table 3.7: Coordinates of H₂CO → *trans*-HCOH transition stateAlso on the Bowman surface; in their papers, referred to as TS₂. Location refined and coordinates transformed from the published configuration [54].

	x	y	z
$\vec{r}_{\text{H}^{(1)}}$ =	2.909270	-1.080449	0.0
\vec{r}_{O} =	0.0	2.212928	0.0
\vec{r}_{C} =	0.0	0.0	0.0
$\vec{r}_{\text{H}^{(2)}}$ =	0.638885	-1.953390	0.0

Table 3.8: Coordinates of tight transition state on Bowman's surfaceIn Bowman's papers, referred to as TS₃; location refined and coordinates transformed from the published configuration [54].

		x	y	z
$\vec{r}_{\text{H}^{(1)}}$	=	2.331282	-0.475906	6.482761
\vec{r}_{O}	=	0.0	2.235156	0.0
\vec{r}_{C}	=	0.0	0.0	0.0
$\vec{r}_{\text{H}^{(2)}}$	=	1.720165	-1.206485	0.0

Table 3.9: Coordinates of proposed roaming saddle

Coordinates transformed from the published configuration [35]; saddle first reported by Harding and co-workers [31]. No such saddle could be located on Bowman's surface.

a *single* imaginary vibrational frequency (most often there were 2 imaginary frequencies in the top 6 modes).

- Further, the vibrational and rotational modes were frequently difficult to separate, there being 4–5 modes of similar magnitude (of order 10 cm^{-1} ; this was the case over several orders of magnitude of h when calculating the Hessian). It is unclear if these were indeed zero-frequency modes, which would have indicated that the algorithm had located a ridge.

It is possible that more sophisticated methods such as eigen-vector following [66] would succeed in locating such a roaming saddle. It is also possible, and this is our conjecture, that the Bowman surface simply does not capture this feature. If nothing else, it is clear that roaming does not require a well-defined saddle. Attempts to define a roaming transition state³ in terms of such a feature are dubious at best.

3.3 Observations from molecular dynamics

In this section, we discuss the results of the molecular dynamics simulations described in Chapter 2. Before we begin, Figure 3.2 provides an instructive map of space we explore. Formaldehyde has 6 intramolecular degrees of freedom⁴, but this particular 2-dimensional projection allows us to see distinctions between each of the three channels

³“Transition state” used here in the sense of a minimum-flux dividing surface in configuration or phase space for an RRKM rate calculation [78].

⁴Recall that even though there are 3 real-space coordinates for each of the 4 atoms, the potential is invariant under net rotation (about 3 axes) and translation (along 3 axes). The system therefore has $3 \times 4 - 3 - 3 = 6$ intramolecular degrees of freedom.

in which we are interested: direct dissociation, radical dissociation, and roaming-mediated dissociation to molecular products. We show three trajectories, one from each pathway. The coordinate along the abscissa is separation between the H₂ center-of-mass and CO center-of-mass. Motion to the right of the figure indicates the separation of H₂ from CO; one can indeed see that the **direct dissociation** (dotted red) and roaming (solid black) pathways terminate in the lower right, as separated molecular products. The ordinate gives the separation between the two hydrogens in the system. The **radical dissociation** trajectory (dashed blue) moves from reactants up the axis, indicating H–H separation, just as we expect. In addition to the trajectories, we show several actual and proposed stationary points of the surface; see caption.

The roaming region depicted in the upper half of the figure is the region defined in eqs. 2.62 and 2.61. The horizontal lines follow trivially from the bounds on Hydrogen separation in eq. 2.62. The curve bounding the right side of the region comes from the force constraint (eq. 2.61). We obtain this boundary by projecting the configurations from the Monte Carlo random walk for sampling the roaming region (described in Section 2.3.3) into the $r_{\text{H-H}}-r_{\text{H}_2-\text{CO}}$ space and drawing a bounding edge.

The figure immediately reveals some of the key features of configuration space and distinctions of the roaming trajectories. The roaming trajectory (solid black) spends a long time with the loosely associated hydrogen far from the formyl parent. This is consistent with the findings of other workers, who observe that roaming hydrogens may spend up to several picoseconds a distance of several Angstroms from the parent [17]. To get an idea of what an extraordinarily long time this is—and to what extent the relevant timescales differ in roaming—consider that the typical vibrational period of the H–H or H–C stretch is of the order of 10 femtoseconds—two orders of magnitude faster. The many wiggles in the black curve while in the roaming region are variations in H–H separation caused by H–C vibrations in the formyl!

We can also see the distinction in product H₂ vibrational excitation for the roaming and direct pathways. By observing the H–H oscillations once H₂ has dissociated from CO (along the lower right), we see how much more vibrationally excited is the roaming H₂ than the directly dissociating H₂ as evidenced in the amplitude of the oscillations. vibrationally hot H₂ is in line with energy-conservation requirements for rotationally cold CO.

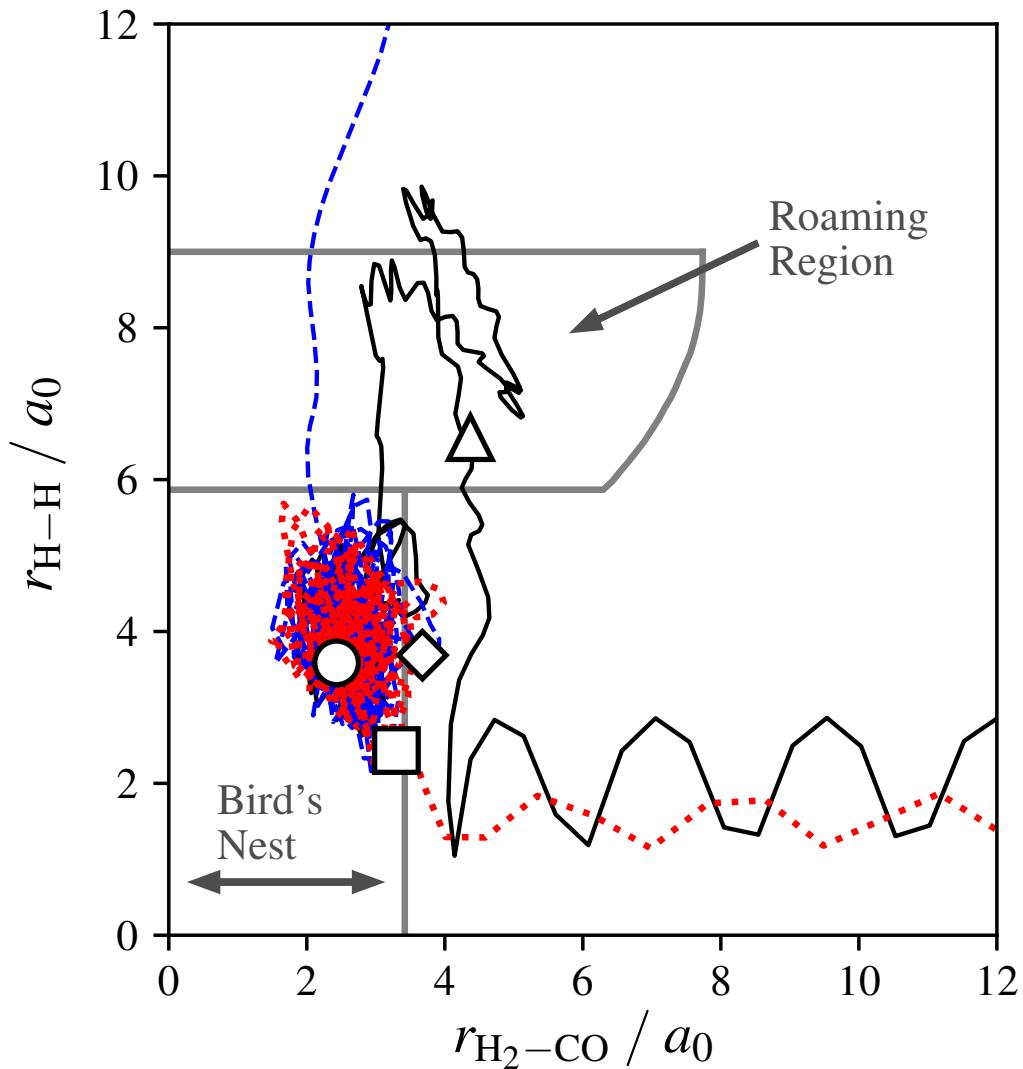


Figure 3.2: Orientation to the H_2CO surface

Vertical and horizontal axes are respectively $\text{H}-\text{H}$ and H_2-CO separation in atomic units. The circle, \circlearrowleft , is the global minimum of the surface. The square, \square , is the tight transition state to molecular products from the global minimum, TS_3 . The triangle, \triangle , is the location of the putative roaming saddle. The diamond, \diamond , is the newly-located saddle of degree two associated with the S_0/S_1 conical intersection; see Section 3.2. A typical **directly dissociating** trajectory is shown in dotted red. A typical **radical dissociation** trajectory is shown in dashed blue. A typical roaming trajectory is shown in solid black. In this and all other figures with axes of dimension length, the axes are measured in atomic units. The trajectories shown were computed with a microcanonical energy, $E_\mu = 37600 \text{ cm}^{-1}$ measured from the global minimum (as will be all E_μ).

Though the roaming trajectory appears to pass quite close to the proposed location of a roaming saddle (Δ), recall that this figure projects flat 4 degrees of freedom and the roaming trajectory may pass quite far “above” or “below” the region of configuration space where, for other potentials, a saddle is reported. To show the diversity of roaming behavior and emphasize that not all roaming trajectories pass by the proposed roaming saddle, we offer Figure 3.3, which shows six randomly selected roaming paths, 3 of which avoid the roaming saddle’s location entirely.

As an aside, this particular projection may be of general utility for visualizing roaming reactions, many of which have the same general scheme:



Where $\cdots\mathbf{C}$ indicates the roaming species. In the case of formaldehyde **A** is CO and **B** and **C** are both H. For the case of acetaldehyde (CH_3HCO), all is the same except that **C** is now the methyl group, CH_3 . The projection we use in Figures 3.2 and 3.3 is the **A**–**BC** center of mass separation along one axis and **B**–**C** center of mass separation along the other. We will make use of the general scheme in eq. 3.2 again when we visualize geodesics in Section 3.5.

3.3.1 Bird’s nest

The region in the lower right of both of Figures 3.2 and 3.3, the “Bird’s Nest”, turns out to be key to our analysis. It is so named for the jumbled (nest like?) appearance of the traces of the trajectories around the reactant well. We defined the extent of the bird’s nest to be mutually exclusive with the roaming region, which bounds the bird’s nest from above $r_{\text{H}-\text{H}} < 5.867 \text{ } a_0$. The right bound is the was selected to create the smallest region which still contained the bulk of the pre-dissociation (and pre-roaming) configurations $r_{\text{H}_2-\text{CO}} < 3.420 \text{ } a_0$. This selection was made without regard for the stationary points identified on the surface⁵, but we note that the edge of the bird’s nest as defined runs along the ridge, which connects the first order saddle corresponding to the transition state to molecular products and the second order saddle associated with the S_0/S_1 conical intersection. The ridge is visible in the projection of the potential shown in Figure 3.1. We will discuss its significance to geodesic analysis in Section 3.6.1

⁵In fact, we defined the bird’s nest *prior* to identifying the relevant stationary point.

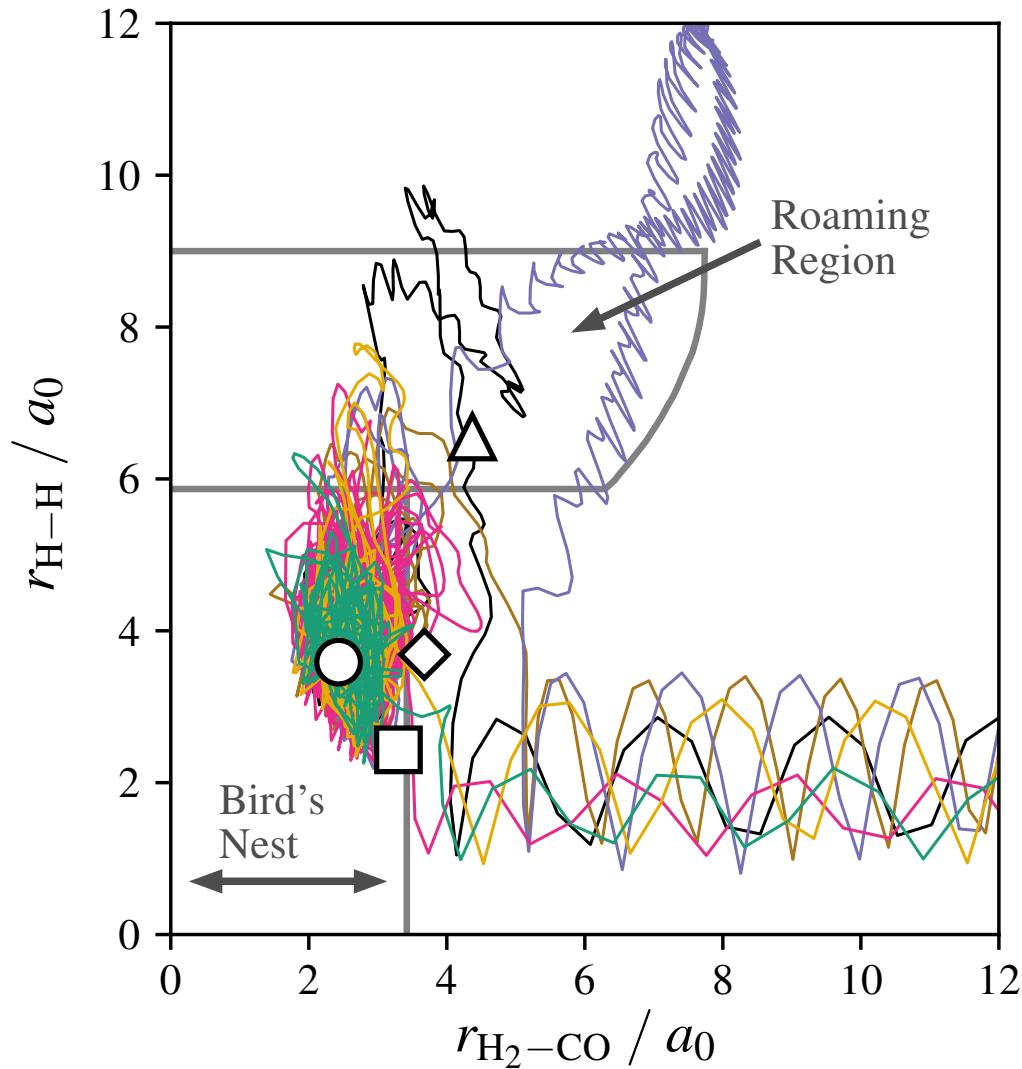


Figure 3.3: Six roaming trajectories

To drive home the point that not all roaming trajectories pass through the neighborhood of the roaming saddle, we here show six randomly selected roaming trajectories. They show a diversity of behavior, but all have in common that they enter the roaming region and then return to dissociate to molecular products. Each trajectory was computed with a microcanonical energy, $E_\mu = 37600 \text{ cm}^{-1}$.

3.3.2 Molecular dynamics branching ratios

Though our analysis is not concerned with detailed prediction of branching ratios for the H₂CO dissociation reaction, we can estimate them for the classical surface under study. Table 3.10 gives the energy-dependent results of molecular dynamics runs. The ratios are also plotted in Figures 3.4 and 3.5; the former shows only reactive trajectories and the latter shows all trajectories. Recall that these plots and figures should not be taken as representative of what formaldehyde would do in an experiment as we perform no semi-classical filtering or selection of trajectories⁶, but rather of the classical dynamics of the Bowman potential surface (see Section 2.1.1). We are concerned with the dynamics of roaming—a phenomenon which manifests purely classically.

Figure 3.4 tells the bulk of the story as it shows ratios normalized with respect to trajectories that succeeded in reacting. All trajectories are initiated at an energy above the radical dissociation threshold ($E_{\mu}^{\text{rad}} > 33240 \text{ cm}^{-1}$). Even though it is accessible, the radical channel is not a dominant component of the flux until energies of 38814 cm⁻¹ and higher, when it achieves parity with the direct dissociation channel. Of particular interest to us is the behavior of the roaming channel. Lower energies show a rocky rise in the fraction of roaming trajectories and then a decline that begins at 38814 cm⁻¹. We should note that above 41060 cm⁻¹, roaming is not observed experimentally [17]. It is likely that above this energy, the same electronic pathway that—at lower energies—allows the deposition of enormous amounts of vibrational energy into the S_0 , ground state surface is no longer accessible [17]. That we use the results above 41060 cm⁻¹ for comparison to geodesic results is sensible because we are seeking to understand the way the inherent dynamics of *this surface* shift with energy.

All of these observations come with a caveat that for lower energies, they are biased towards fast processes. We know that each trajectory has sufficient energy to dissociate, but given that we stop integration after roughly 12 ps, not all do. We do not know the relative time scales on which reactions through each channel occur, but we certainly suspect that they are different, especially at high energies (see how much slower is roaming in Figure 3.6).

⁶Our purely classical treatment does not discard states for violating zero-point energy or other “semi-quantum” concerns; see [58].

Energy / cm ⁻¹	unsuccessful	no reaction	direct	radical	roaming
34500	184	38441	1097	260	18
35000	326	36967	1825	813	69
35500	507	34637	2943	1806	107
36000	778	31231	4533	3283	175
36223	943	29295	5304	4226	232
37000	1331	22329	8549	7415	376
37600	*1	*18253	11211	10016	519
37688	1814	15750	11340	10553	543
38000	1963	13101	12747	11590	599
38688	2279	8042	15161	13761	757
38814	2322	7359	15462	14071	786
41010	2544	993	18549	17040	874
45000	2782	24	17671	18711	812

Table 3.10: Dispositions of molecular dynamics trajectories with energy

For each energy, we give the final disposition of the 40000 molecular dynamics trajectories we computed on the formaldehyde energy surface (details given in Sections 2.3.1 and 2.3.2). Unsuccessful trajectories were deemed insufficiently accurate for the time step at some point during integration and discarded. Trajectories are classified as described in Section 2.3.2. Except for the energy column, in units of wavenumbers, all figures give absolute number of trajectories.

*: Trajectories at 37600 cm⁻¹ were computed without properly detecting consistency failure and so some “unsuccessful” trajectories are erroneously categorized as “no reaction”. Assuming a failure rate equal to that of 37688 cm⁻¹, the error will skew branching ratios down by no more than 5% of the expected value.

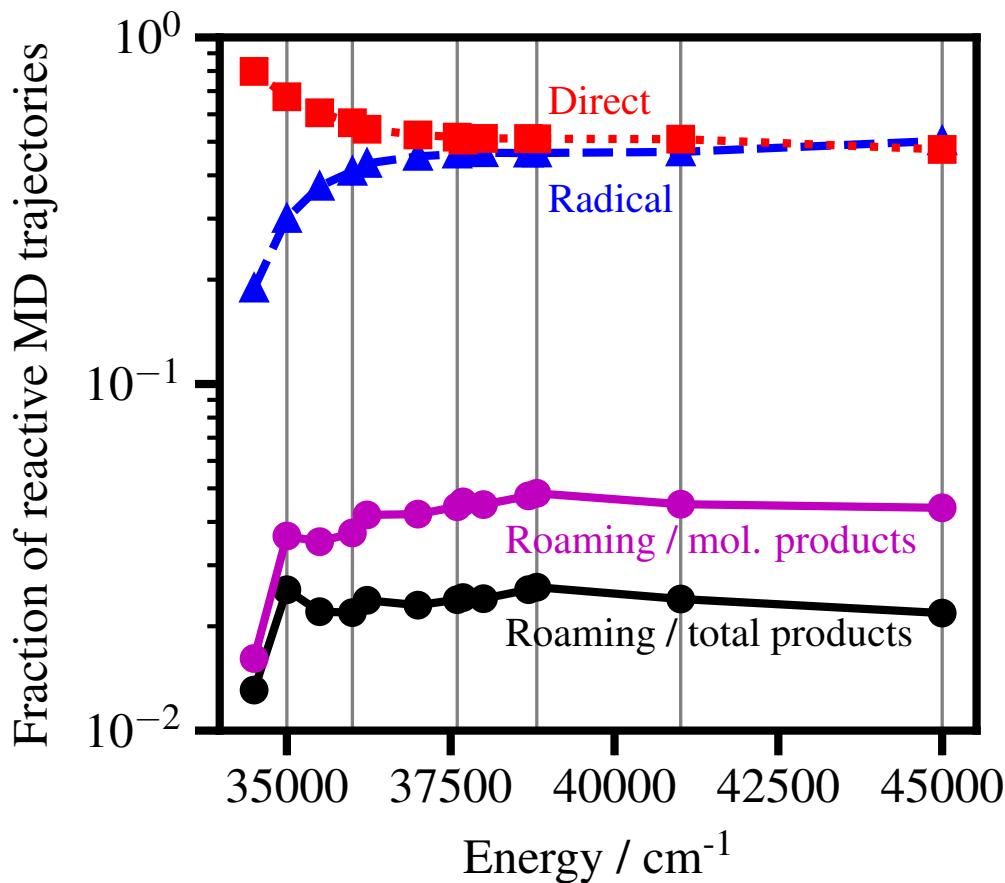


Figure 3.4: Molecular dynamics channels (reactive)

Final states of reactive molecular dynamics trajectories. Trajectories not reacting within 12.1 ps are excluded from the figure and the relative abundance of each channel is shown (see Figure 3.5 for all trajectories). The magenta curve, “Roaming / mol. products”, shows the fraction of molecular products that arose from roaming: $\text{roaming}/(\text{roaming} + \text{direct})$. The black curve gives roaming as a fraction of all products: $\text{roaming}/(\text{roaming} + \text{direct} + \text{radical})$. Error bars giving the two standard-deviation estimate for the proportions plotted would be smaller than the symbols used. Vertical lines indicate energies for which geodesics were computed.

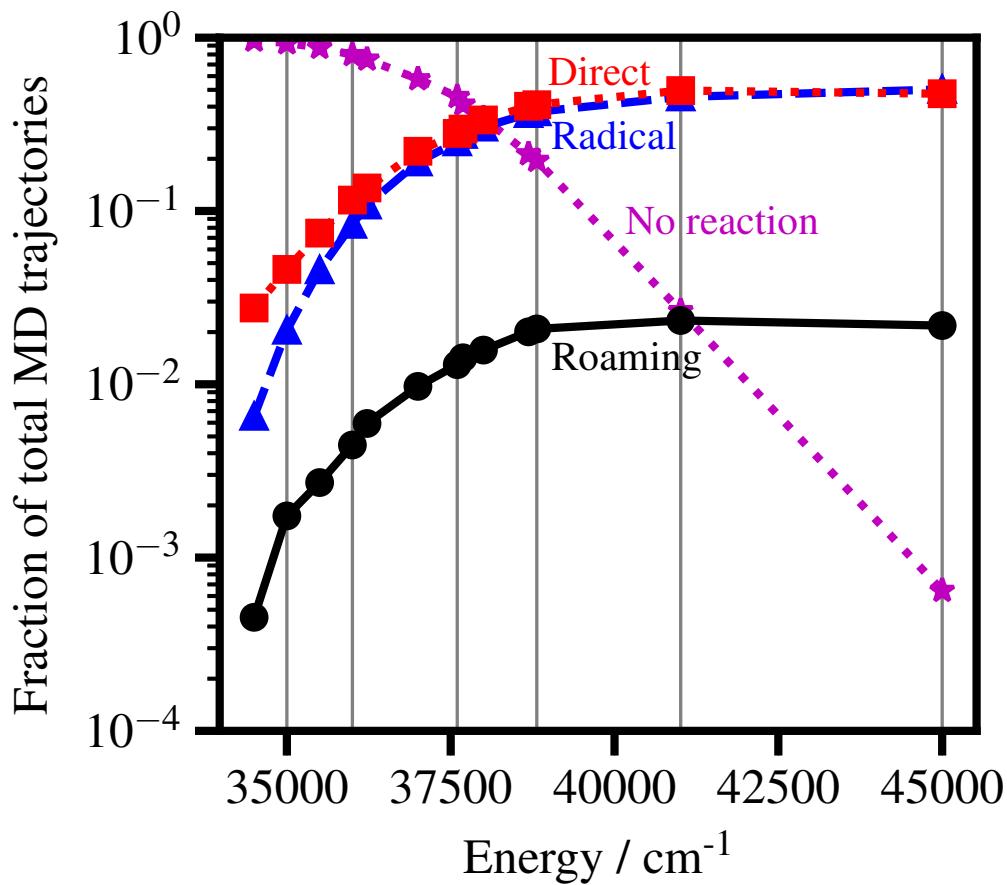


Figure 3.5: Molecular dynamics channels

Disposition of molecular dynamics trajectories as a function of excitation energy. In contrast to Figure 3.4, the figure includes trajectories that do not react (“No reaction”), but does not include those trajectories that failed for numerical reasons (see Table 3.10). Error bars giving the two standard-deviation estimate for the proportions plotted would be smaller than the symbols used for all points except the roaming fraction at 34500 cm⁻¹, which is $4.5 \cdot 10^{-4} \pm 2.1 \cdot 10^{-4}$. The vertical lines indicate the energies for which geodesics were also computed.

3.3.3 An estimate of the energy-dependent incubation period

We were interested in computing the times to reaction for each of the dissociation channels: direct, radical, and roaming. To determine this “incubation period”, we use a scheme similar to Houston and coworkers [58]: counting frames of the stored configurations in reverse until the criteria for dissociation no longer hold. We define our criteria in terms of the same space we visualized in Figure 3.2. For trajectories going to radical products, we take the last frame (counted in reverse) for which $r_{\text{H}-\text{H}} > 9 \text{ a}_0$. This corresponds to just after the last time the system exits the roaming region in terms of hydrogen separation. For trajectories going to molecular products (directly or by way of roaming) we take the last frame, again counted in reverse, for which $\|\vec{r}_{\text{CO}} - \vec{r}_{\text{H}_2}\| > 7 \text{ a}_0$. This is well past the moment that the trajectory exits the bird’s nest, but the delay is necessary to avoid mis-classifying roaming trajectories as already dissociated. From Figure 3.3, it can be seen that we would have made such an error for roaming trajectories if we had simply taken the limits of the bird’s nest to demark the moment of dissociation.

To treat the large number of trajectories that did not dissociate, we computed the time by averaging the rate (inverse reaction time). Those trajectories that did not dissociate required effectively infinite time therefore contribute a rate of 0 to the mean. In Figure 3.6 we plot the inverse of the mean of inverses, that is the time. The incubation period for each channel at each energy is computed as:

$$t_{\text{incubation}}^{-1} = \frac{1}{N} \sum_i^N t_i^{-1} \quad (3.3)$$

where the $\{t_i\}$ are the set of N reaction times as defined in the previous paragraph.

The major salient observation is that roaming is consistently a slower event than direct or radical dissociation. At present we have not been able to quantitatively connect this fact to the analysis afforded by geodesics, but it is an area for further study.

A technique we tried, but found to be unsuccessful for estimating the rate of reaction is instructive in its failure. Houston and coworkers [58] estimated the *total* rate of formaldehyde decomposition by fitting the distribution of lifetimes, t , to the distribution:

$$\rho(t) \propto \exp(-kt) \quad (3.4)$$

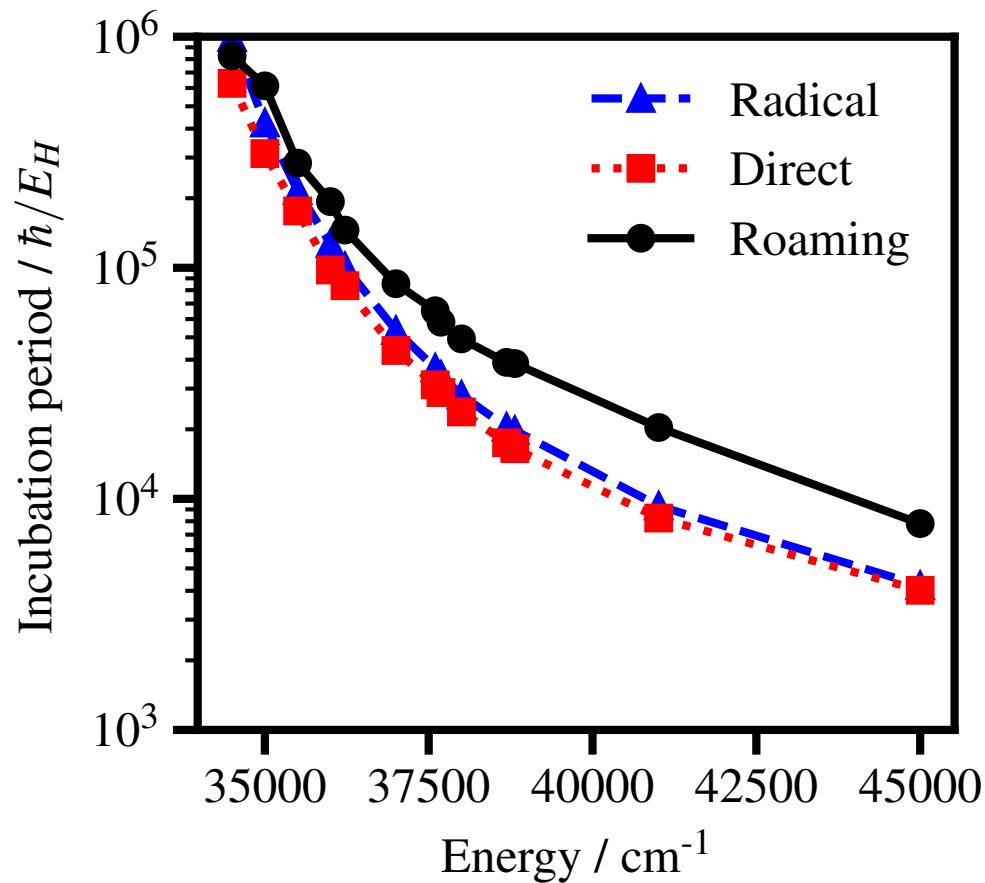


Figure 3.6: Rates of reaction

Incubation periods in atomic units for each reaction channel as a function of excitation energy. We treat non-reactive trajectories as having effectively 0 rate of reaction when contributing to the mean.

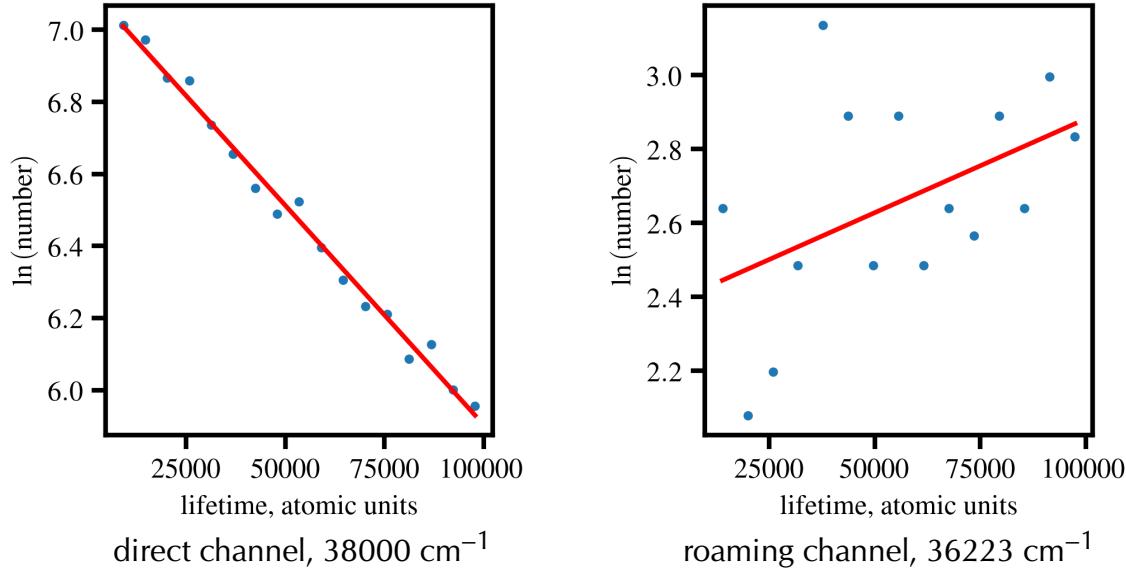


Figure 3.7: Failure of statistical distribution of reaction times

In both figures, the ordinate is natural log of the number of trajectories as a function of lifetime in atomic units (along the abscissa). If eq. 3.4 were a good description of the rates, then the slope of the line of best fit would be $-k$, the rate constant for the mechanism. The plot on the left is for the direct channel at 38000 cm^{-1} and shows the prediction succeeding quite well. The figure on the right is for the roaming channel at 36223 cm^{-1} and is showcases one of the many failure modes of the channel; this instance is particularly spectacular as it predicts a *negative* rate, which is nonsense.

Such a distribution results from the simple assumptions that [79]:

- The system is ergodic—that all states with sufficient energy will eventually react.
- All vibrational states have equal probability of reacting per unit time; *i.e.* vibrational energy redistribution is fast relative to the reaction.

While this method yielded credible fits at most energies for the direct and radical channels, it was not suitable for roaming, even at energies for which there were many samples available. The failure points to the non-trivial consequences roaming has for theories of unimolecular dissociation. See Figure 3.7.

3.3.4 Existence of high J_{CO} roamers

In formaldehyde, the spectroscopic signature that first hinted at the existence of roaming was the presence of a low-energy, “cold”, shoulder in the product CO rotational distribution [9]. In retrospect, this comes as no surprise. One would certainly imagine that, given the long excursion taken by the roaming hydrogen, its plunging reentry and abstraction of the other hydrogen would leave the product H₂ highly vibrationally excited. By necessity of energy conservation, the CO would be less rotationally excited than if the dissociation had proceeded via the tight transition state to molecular products, which features far less dramatic hydrogen-hydrogen separation. This “cold shoulder” has been used in the identification of roaming in other systems [7, 19] and it is tempting to assume that the phenomenon and the observation are inseparably coupled.

In the present case, we find just such a “cold shoulder” in the J_{CO} distribution for products of molecular dynamics. We estimate J_{CO} via the crude quantization:

$$\left\| \vec{L}_{\text{CO}} \right\| \approx \hbar J_{\text{CO}} \quad (3.5)$$

where \vec{L}_{CO} is the moment of inertia of CO about its center of mass. We then take

$$J_{\text{CO}} = \left\lfloor \left\| \vec{L}_{\text{CO}} \right\| / \hbar \right\rfloor \quad (3.6)$$

where $\lfloor \cdot \rfloor$ is the floor function. This analysis is certainly sufficient as a heuristic and for order-of-magnitude observations, but not for a detailed analysis that invoked small differences. The results of such an analysis from three representative energies are shown in Figure 3.8.

Our approximation represents a small deviation from full expression:

$$\vec{L}_{\text{CO}}^2 = J_{\text{CO}} (J_{\text{CO}} + 1) \hbar$$

Defining $\vec{L}_{\text{CO}}^* \equiv \vec{L}_{\text{CO}} / \hbar$, a better semi-classical expression would be:

$$J_{\text{CO}} = \left\lfloor \sqrt{\vec{L}_{\text{CO}}^{*2} + 1/4} - \frac{1}{2} \right\rfloor$$

Better, but only marginally better, especially for large \vec{L}_{CO}^* , which we have here. Indeed for $\vec{L}_{\text{CO}}^* = 10$, the fractional deviation is only 5% between our approximation and the exact result prior to applying the floor. A more rigorous analysis would also demand a

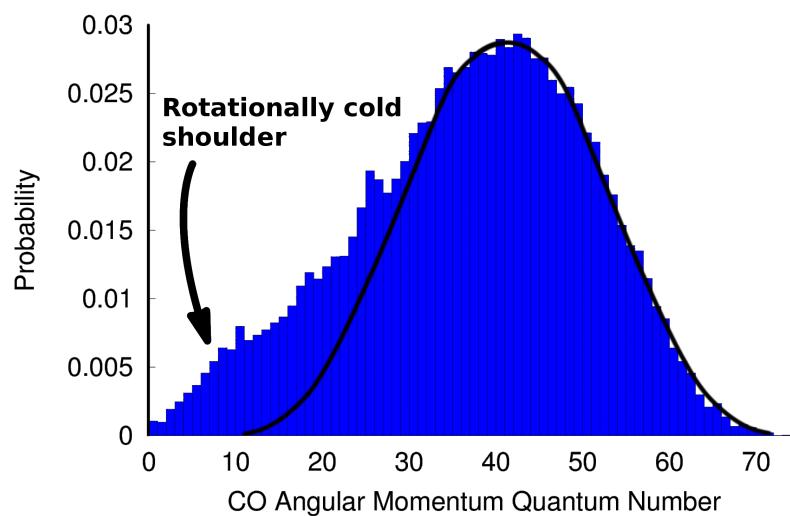


Figure 3.8: A “cold” shoulder in the J_{CO} distribution

Histogram of J_{CO} made by combining the molecular products from three energies: 37000 cm^{-1} , 37600 cm^{-1} , and 38000 cm^{-1} , corresponding to a 330 nm photolysis laser. Molecular dissociation accounted for 23414 of 119997 trajectories (39999 at each energy). The black curve is symmetric about its mean and serves as a guide to the eye. The low-energy shoulder to the distribution is often taken as an indicator of roaming. We cannot tell from the above figure, but there *do* exist roaming trajectories giving rise to product states with $J_{CO} > 30$.

Energy / cm ⁻¹	35000	36000	37000	38814	41010	45000
$J_{\text{CO}} \geq 30$	0.188	0.259	0.309	0.377	0.398	0.266
$J_{\text{CO}} \geq 60$	0	0.0058	0.0133	0.0115	0.0206	0.0185

Table 3.11: Fraction of roaming trajectories with high J_{CO}

The table above gives the fraction of roaming trajectories with $J_{\text{CO}} \geq 30$ and $J_{\text{CO}} \geq 60$ for several energies. For the given energies, at least 15% of the roaming trajectories hide in the symmetric, $J_{\text{CO}} \geq 30$ portion of the distribution comparable to Figure 3.8. Up to 2% of the roaming trajectories have $J_{\text{CO}} \geq 60$.

sampling scheme that properly treated the quantum nature of the system; here we use J_{CO} simply as a proxy for the magnitude of the rate of angular rotation.

Seeing the low-energy component of the distribution in Figure 3.8 was one of our first global indicators that the surface supplied to us functioned as it should. When we examined the trajectories corresponding to low J_{CO} , we observed that they manifestly roamed. Our early definitions of roaming even exploited this, calling all trajectories with $J_{\text{CO}} \leq 15$ “roaming” and all those with $J_{\text{CO}} \geq 30$ “directly dissociating”. The excluded middle was an acknowledgment of the inherent ambiguity in the definition of roaming though was of little consequence as we were not seeking branching ratios at that stage.

However, we discovered that this simple classification doesn’t tell the whole story. In Section 2.3.3, we give a definition of roaming that relies only on passage through and return from a special region of configuration space. The trajectories satisfying the our definition based on the “roaming region” do indeed “look” like roaming, but some exhibit decidedly higher J_{CO} . As seen in Figure 3.9, there are roaming trajectories with J_{CO} even higher than 60! See Table 3.11 for the explicit fractions. This surprising and complicating fact must be borne in mind when describing and defining roaming or we risk missing large components of the distribution of roaming dynamics. Later, we will see that roaming is indeed enabled by just such a wide distribution in features.

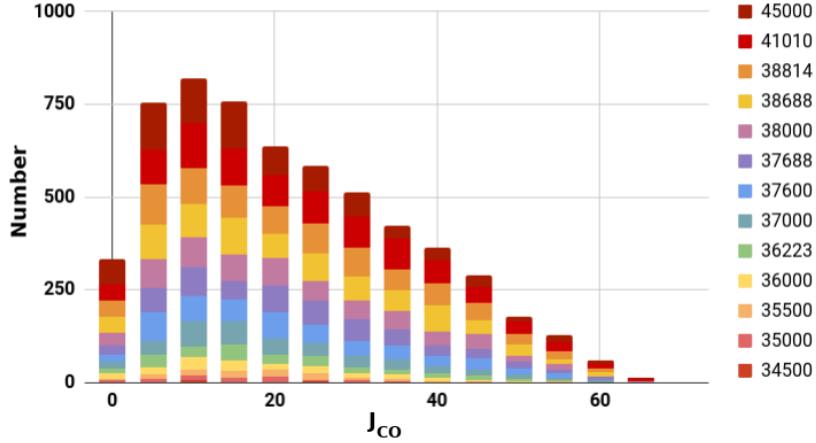


Figure 3.9: Roamers exhibiting high J_{CO}

Distribution of product J_{CO} for roaming trajectories at all microcanonical energies, E_μ , studied. Each bin along the abscissa is a range of 5 quanta, e.g. 0 corresponds to 0–4, 20 to 20–24, etc. While slow rotation of the CO fragment has been regarded as a hallmark of roaming behavior, clearly this simplistic, energy-conservation story does not capture the range over which energy is deposited.

3.4 Geodesics are consistent with estimate

In Section 2.5, we established a method to estimate geodesic total and vibrational lengths. Though we deliberately do not typically analyze geodesics over their entire length (see the discussion to come of the “bird’s nest” in Section 3.6.1), our calculations yield accurate estimates for lower-bounds on both the vibrational and total length of the *entire* path. In Figure 3.10, we show distributions of the total and H₂ vibrational length computed at $E_L = 37600 \text{ cm}^{-1}$.

From eq. 2.105, we have an estimate of total length, ℓ_{Total} :

$$\ell_{\text{Total}} \gtrsim 585 a_0 \sqrt{m_e} \quad (3.7)$$

and the H₂ vibrational length, $\ell_{\text{H}_2\text{-Vib}}$, as:

$$\ell_{\text{H}_2\text{-Vib}} \gtrsim 30 a_0 \sqrt{m_e} \quad (3.8)$$

Comparing to Figure 3.10, we see that both estimates are indeed near the lower bound of their respective distributions. Quantitatively, $\ell_{\text{Total}} = 585 a_0 \sqrt{m_e}$ corresponds to the 0.166th percentile of $\rho(\ell_{\text{Total}})$ and $\ell_{\text{H}_2\text{-Vib}} = 30 a_0 \sqrt{m_e}$ corresponds to the 0.067th

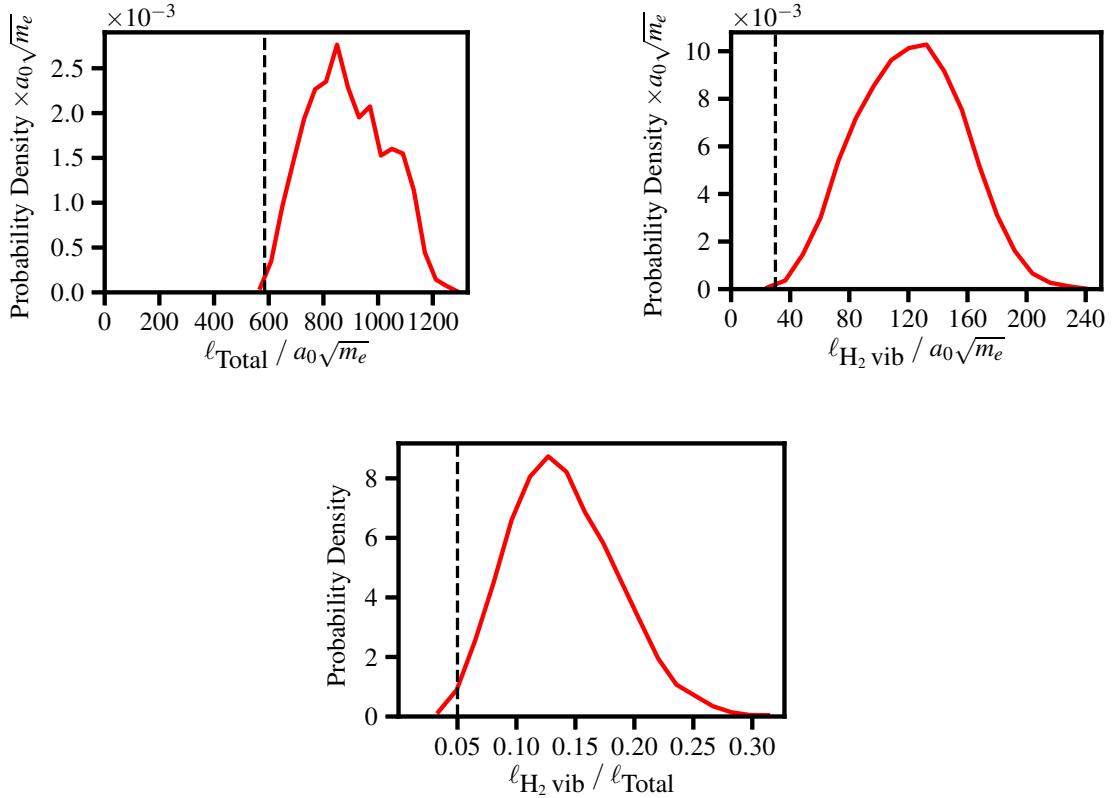


Figure 3.10: Distribution of lengths over full geodesic paths

Distributions composed of 19321 direct geodesics computed at $E_L = 37600 \text{ cm}^{-1}$. Lengths calculated as described in Sections 2.4.1 and 2.4.2 over the full extent of the paths. The lower-middle distribution is of the ratios of the total and vibrational length for each geodesic. The dashed, vertical lines give our (quite accurate) estimates for the lower bounds of each distribution. Note that our estimate for $\ell_{\text{H}_2 \text{ vib}}/\ell_{\text{Total}}$ is not intended to be a lower bound as we take the low estimates for both distributions rather than a lower bound for the numerator and an upper bound for the denominator.

percentile of $\rho(\ell_{\text{H}_2\text{-Vib}})$. These extremes makes sense because we ignored many of the motions involved; modulo our physically reasonable guesses for $\|\Delta\vec{r}_{\text{H}_2\text{-CO}}\|$ and $\|\Delta\vec{r}\|$, the estimates *should* be lower bounds! Combining the two results allows us to estimate the vibrational fraction to be:

$$\ell_{\text{H}_2\text{-Vib}}/\ell_{\text{Total}} \approx 30/585 = 0.05 \quad (3.9)$$

which is also within the tail of the distribution shown in Figure 3.10, the 0.756th percentile.

3.5 Geodesics are inherent dynamics

In Section 2.1.2 we defined the *inherent dynamics* of a system as a filtered view of its classical dynamics—revealing the essential motions stripped away of noise. A sketch of the relationship between both such dynamics is shown in Figure 3.11. We then make the case (in Section 2.2) that geodesics in the landscape ensemble are these inherent dynamics.

While the group has been computing geodesics and using them to understand the inherent dynamics of systems under study for over a decade, the geodesics themselves have resisted direct visualization except in the two-degree of freedom case [48]. The systems treated in the past simply have too many degrees of freedom to allow their rendering after the fashion of Figure 3.11. Unlike simple liquids, which have many hundreds or thousands of degrees of freedom, formaldehyde has only 6—still non-trivial, but perhaps geodesics on H₂CO are amenable to visualization. The general scheme for roaming reactions given in eq. 3.2 offers three suggestive coordinates. Recall that we have **ABC** forming **A** + **BC** by way of roaming **C**. Specifically CO is **A** and the two H are **B** and **C**. Forming coordinates from the inter-center-of-mass distances for each of the **ABC** centers gives three coordinates: **AB**, **BC**, **CA**, which we can plot simultaneously.

Figure 3.12 shows a single roaming molecular dynamics trajectory overlaid with two geodesics, one that proceeds directly from reactants to products and another that is constrained to pass through one of the points along the trajectory in the roaming region—a geodesic that roams. The degree of similarity between Figure 3.11 (a cartoon) and Figure 3.12 (real data from a physical surface) and is remarkable—the geodesics

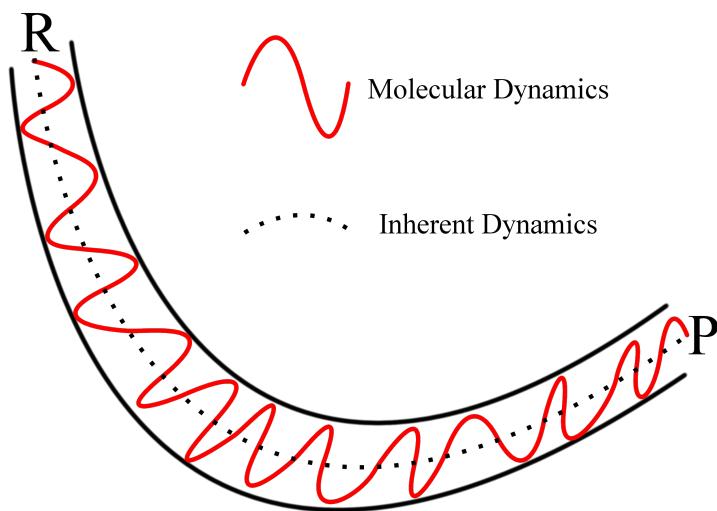


Figure 3.11: Inherent dynamics pictograph

We sketch the inherent dynamics (dotted black curve) of a process connecting (**R**eactants to (**P**roducts and show the relationship to the **classical dynamics** (solid red). The classical curve shows the oscillations of the molecular dynamics trajectories within the reactive tube of possible paths (solid black, bounding curves). The Inherent dynamics are the smooth, dotted path that washes out the non-essential features of the molecular dynamics.

clearly match our model of the inherent dynamics. The roaming geodesic, in blue, exits the bird’s nest with the trajectory and then swings towards the products. Even the direct geodesic, which is not constrained to pass through the roaming region, captures the essence of the trajectory after dissociation. This image, or one like it was the inspiration for the the naming of the “bird’s nest” (see Section 3.3.1). It is visible as the dark scribble in the center and results from pre-dissociation vibration of the molecule.

While this picture is evocative, we will not dwell on images like it. As we discuss in Section 3.6.3, there is naught to be gained by restricting ourselves to dynamically correlated⁷ endpoints as we do here. Indeed we can more widely explore the energy landscape when freed from such constraints. Figure 3.12 is useful for confirming our notion of what geodesics represent, but the power in using geodesics lies in the way they globally sample the energy landscape and comes from the inferences we can draw by studying them in aggregate.

⁷This dynamical correlation refers to the fact that each of the boundary configurations are connected by a single classical trajectory.

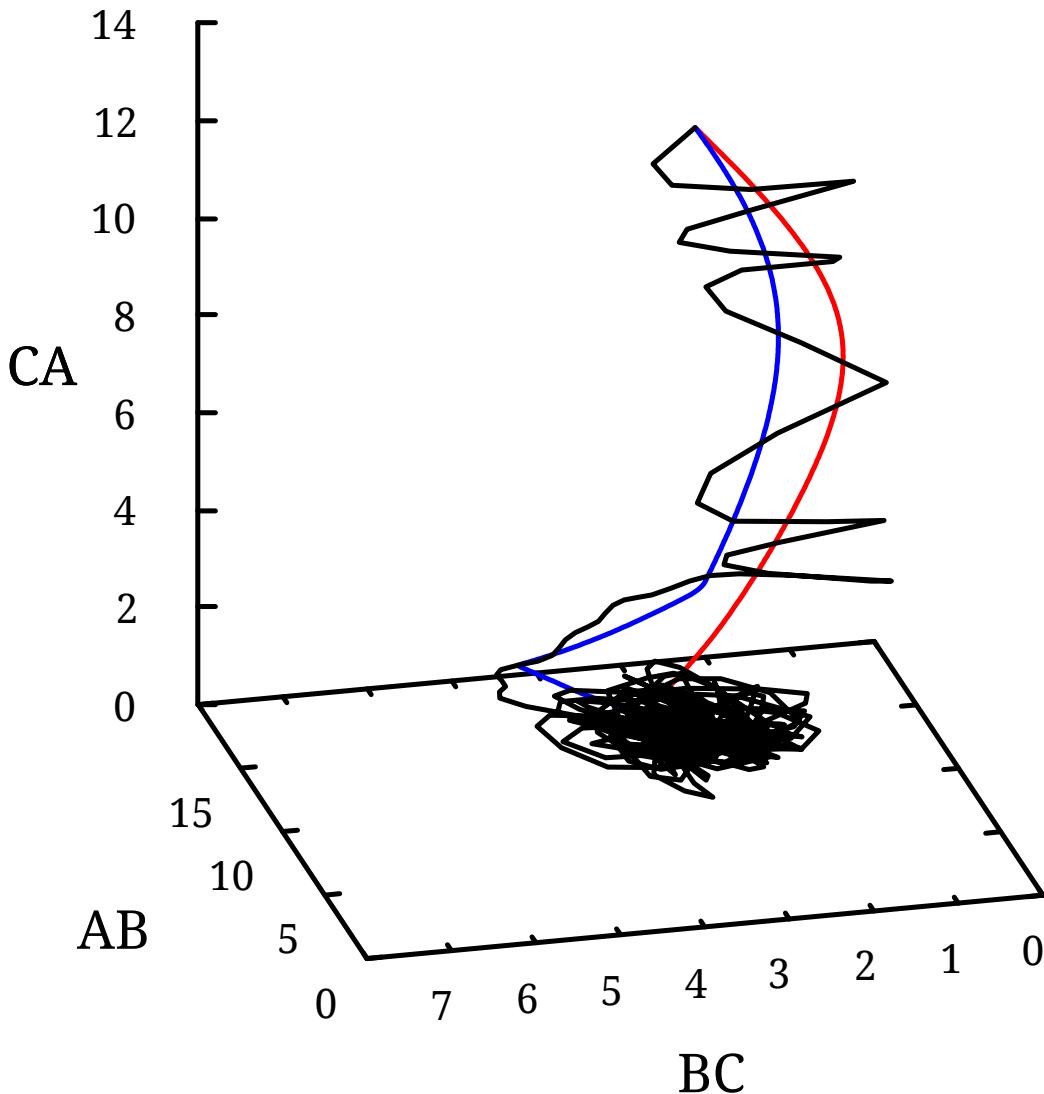


Figure 3.12: Geodesic as inherent dynamics of formaldehyde

The black curve is a classically computed molecular dynamics trajectory, which roams; $E_\mu = 37600 \text{ cm}^{-1}$. The geodesics, **direct** (in red) and **roaming** (in blue), were computed as part of our previous study [1] with $E_L = 34928.7 \text{ cm}^{-1}$. Incidentally, this is very nearly equal to the value of the potential at the point in the roaming region. The red curve is an unconstrained geodesic between the starting and ending points of the molecular dynamics trajectory. The blue curve is the geodesic between the same points, constrained to pass through a point on the trajectory in the roaming region. The axes are defined in terms of the center of mass separations between components of the general **ABC** system defined in eq. 3.2. The **BC** axis is H–H separation and the **AB** and **CA** axes are CO–H separation for each hydrogen, **B** and **C**.

3.6 Observations from geodesics

Having established that the geodesics we compute are reasonable in their magnitude and that, in at least in some sense, they match our mental model of inherent dynamics, we can interrogate them to understand the dissociation dynamics of formaldehyde. Recall that geodesics are global measures of the surface that do not depend on the fine details of minima and transition states; rather they give us information about the connectivity and structure of the large-scale features of the energy surface.

For each of the six landscape energies we attempted to compute at least 12000 geodesics of each class (see Table 3.12). On our in-house cluster (`ted`), which mostly runs Intel Xeon E5440 processors at 2.83 GHz, the average time to compute a single, optimized geodesic is approximately 1 core-hour⁸. Nodes have at least 512 MB of memory per core and computation is always processor- rather than memory-bound. The geodesics computed for the present study represent on the order of a quarter of a million core-hours of computational effort.

Table 3.12 gives the rate of success for our algorithm in computing the different classes of geodesics with energy. The broad observation is that direct and radical geodesics are generally successfully completed at all energies. Roaming geodesics, on the other hand vary in the their rate of completion: 69–76% at low energies and not approaching 100% until the upper end of our scale. The algorithm will fail if there is no path connecting the endpoints which respects the landscape criterion. The prevalence of failures, then suggests⁹ that at the lower energies under study, configuration space is disconnected—and that the dynamics are not *ergodic*. This is a non-trivial observation: surely the dynamics are not ergodic at energies so low that the system is confined to the reactive well. However, even the lower end of our energy range is nearly 2000 cm⁻¹ above the radical dissociation threshold, $E_\mu^{\text{rad}} = 33240 \text{ cm}^{-1}$. At these, relatively *high*, energies non-ergodicity is surprising.

We must draw this inference because the landscape energy is equal to the micro-canonical energy, $E_L = E_\mu$. By choosing the landscape energy, E_L , to be equal to the microcanonical energy, E_μ , we have the property that *all* configurations that result

⁸Using more modern hardware, such as the cluster at Brown's Center for Computation and Visualization improved this time by a rough factor of two.

⁹provided that our algorithm is free from more-fundamental sources of failure

Energy / cm ⁻¹	direct	radical	roaming
35000	11998	11993	8256
36000	12000	12000	8352
37600	*19321	12000	8816
38814	12000	12000	9135
41010	11998	12000	11949
45000	12000	11997	11712

Table 3.12: Success rate of computed geodesics by channel and energy

Here we give the number of successfully computed geodesics by channel and energy. *: For all channel/energy combinations except direct geodesics at 37600 cm⁻¹, exactly 12000 geodesics were attempted. For direct geodesics at 37600 cm⁻¹, 19321 were attempted; all were completed.

from classical trajectories computed at E_μ are members of the corresponding landscape ensemble at $E_L = E_\mu$. Therefore, there cannot be regions of the surface explored by molecular dynamics or by the random walk sampling the roaming region which are energetically permitted in the landscape ensemble. This disconnection is consistent with our observation, in Section 3.6.3, that the roaming geodesics for dynamically correlated boundary points have a broader distribution than those that are not. The disconnect also points back to the failure of a simple statistical theory to capture the rate of reaction for roaming events in Section 3.3.3; indeed one of the required assumptions for the statistical theory was a connected, ergodic space.

3.6.1 Implications of the bird's nest

The next property of the geodesics we will examine is their length. This is natural as geodesics are defined in terms of a minimization condition on their length and the length gives information about the relative amounts of motion involved in the different channels. Throughout this chapter, we've been hinting at the significance of the region about the reactive well, which we call the bird's nest. One of the striking observations revealed by Figure 3.2 is how different are trajectories from each channel. All go their own way and have little overlap outside of the region labeled bird's nest. That the trajectories spend much time *together* in this tangle of pathways (resembling, to the author's eyes, the nest of a bird) is key to our analysis.

By construction, the geodesics for different channels go to manifestly different regions of configuration space—indeed we demand that roaming geodesics detour before heading towards molecular products. It would then be no surprise if the lengths were different. To make a just comparison, we therefore confine our attention to a special region of overlap—the bird’s nest. We see that the system oscillates there, about the global minimum, for some time no matter the products or route taken to them. This is the incubation period during which energy is redistributed among the different degrees of freedom and the ultimate trajectory of the system is determined. The bird’s nest provides a region where we can ask the same questions of different classes of paths and make equitable comparison between them. This method does not require us to identify a reaction coordinate, but only to locate a region where all trajectories spend significant time before diverging.

To use the bird’s nest when analyzing geodesics, we identify as t_b the moment when a geodesic first exits the bird’s nest and then measure the length (or other property) of the geodesic only until departure. For example, we modify the form of the total length as given in eq. 2.67 to be:

$$\ell_{\text{Total}} = \int_0^{t_b} d\tau [2 \cdot T(\tau)]^{1/2} \quad (3.10)$$

where the limits of integration have changed to only include the portion of the geodesic in the bird’s nest. We make similar modification to our expressions for the component kinematic lengths (eqs. 2.73 and 2.88) and for the boundary fraction (eqs. 2.90 and 3.11).

Given that trajectories for each channel overlap strongly in the bird’s nest, there is no reason so expect the geodesics for different channels to have markedly different properties there. Any differences we do observe then, should be attributed to the unique dynamical features associated with each pathway.

3.6.2 Key observations at a single energy

We can use geodesics to understand the differences between the dynamical pathways available to the formaldehyde molecule as it dissociates. The distributions of total lengths of geodesics give us information about the relative distances the system has to traverse to transform from one species into another. By partitioning the kinetic energy into other, physically interesting degrees of freedom such as diatomic vibration or

rotation we can monitor the role that those motions have in the reaction as well. For example, the distribution of $\ell_{\text{H}_2 \text{ vib.}}$ along geodesics gives an indication of the amount of H₂ stretching required within the bird's nest along this minimum-motion path. We will also examine the amount of time spent encountering obstacles, the boundary fraction, which indicates how constrained are the paths taken by geodesics.

We will come to the energy-dependent properties of geodesics in Section 3.6.4, but it is instructive and will allow us to orient our thinking, to first consider a single, extremal landscape energy. In this section, we take up geodesics¹⁰ with $E_L = 35000 \text{ cm}^{-1}$, an energy soon after the onset of roaming in formaldehyde to see what they can tell us about how and why roaming occurs.

Broad roaming distribution in length at low energy

Figure 3.13 serves as our principle point of departure, showing the distributions of total kinematic lengths lengths in the bird's nest for each reaction channel. Here, even when confined exclusively to the bird's nest, we see a striking difference between the roaming paths and either the direct or radical paths. While the distributions for direct and radical paths are highly peaked, the roaming distribution is smeared across the abscissa. This broad distribution of minimal transit lengths points to a plurality of routes to the roaming region.

We are able to take a more detailed look at the distributions shown in Figure 3.13 by examining the component kinematic lengths in Figure 3.14. The internal degrees of freedom plotted here show the same motif played out in each partitioning of the kinetic energy. Direct and radical path-length distributions are sharply peaked near zero and the roaming path distributions are considerably broader with longer tails.

These component length distributions do *not* give us information about the relative product yields in formaldehyde photodissociation. Rather, the component length distributions yield information about the extent to which kinetic energy is focused in into different degrees of freedom during the incubation period prior to dissociation. For example, it is remarkable that there is more H₂ stretching for geodesics heading to

¹⁰Though the distributions we examine in this chapter are quite similar to those in our paper [1], they are computed with a single landscape and geodesics energy, $E_L = E_\mu = 35000 \text{ cm}^{-1}$. In that paper, the landscape energy was lower than the microcanonical excitation energy, $E_L = 34928.7 \text{ cm}^{-1} < E_\mu = 37000 \text{ cm}^{-1}$.

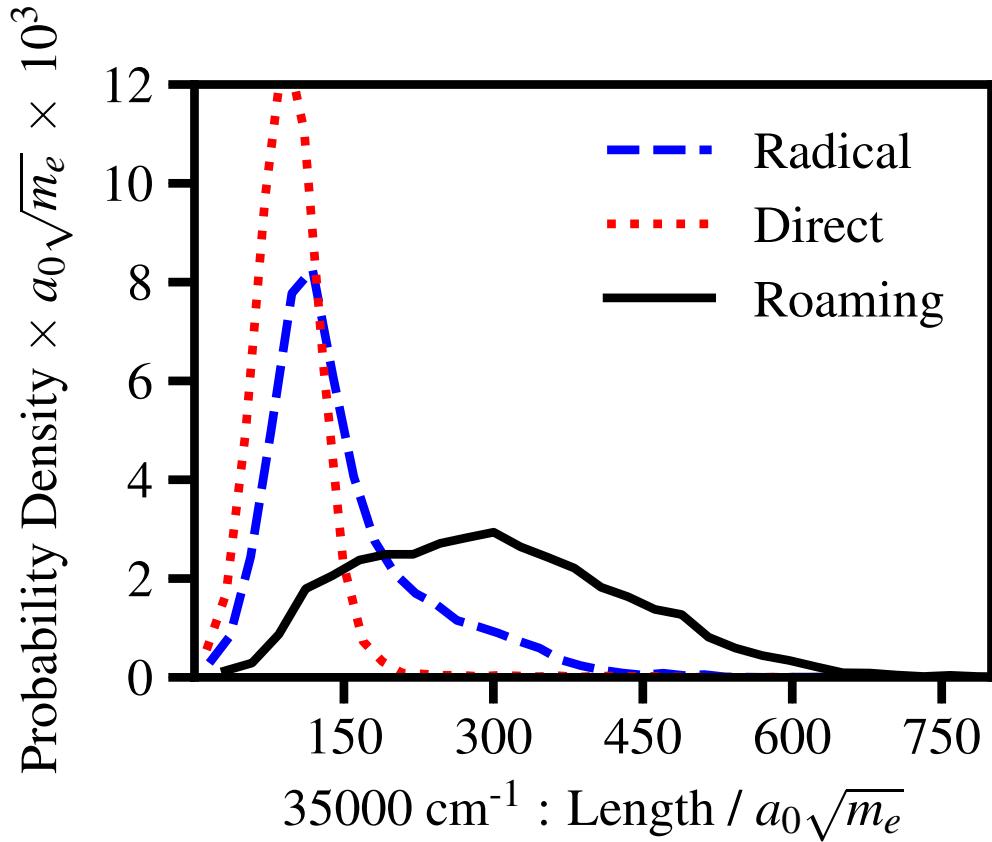


Figure 3.13: Total length distributions

Total kinematic length probability distributions for geodesics computed at $E_L = 35000 \text{ cm}^{-1}$ within the bird's nest. We have no reason to suspect that, when confined to this region, the geodesics for each channel would be different from each other. However, the distributions for geodesics directly between reactants and molecular products (**direct**) and from reactants to radical products (**radical**) are substantively more peaked than for those that go to molecular products by way of the roaming region (**roaming**). Though they appear nearly indistinguishable to those distributions in our recent paper [1] (where we used $E_L = 34928.7 \text{ cm}^{-1}$ unequal to $E_\mu = 37000 \text{ cm}^{-1}$) the distributions in the figure were calculated with landscape and microcanonical energies equal, $E_L = E_\mu = 35000 \text{ cm}^{-1}$.

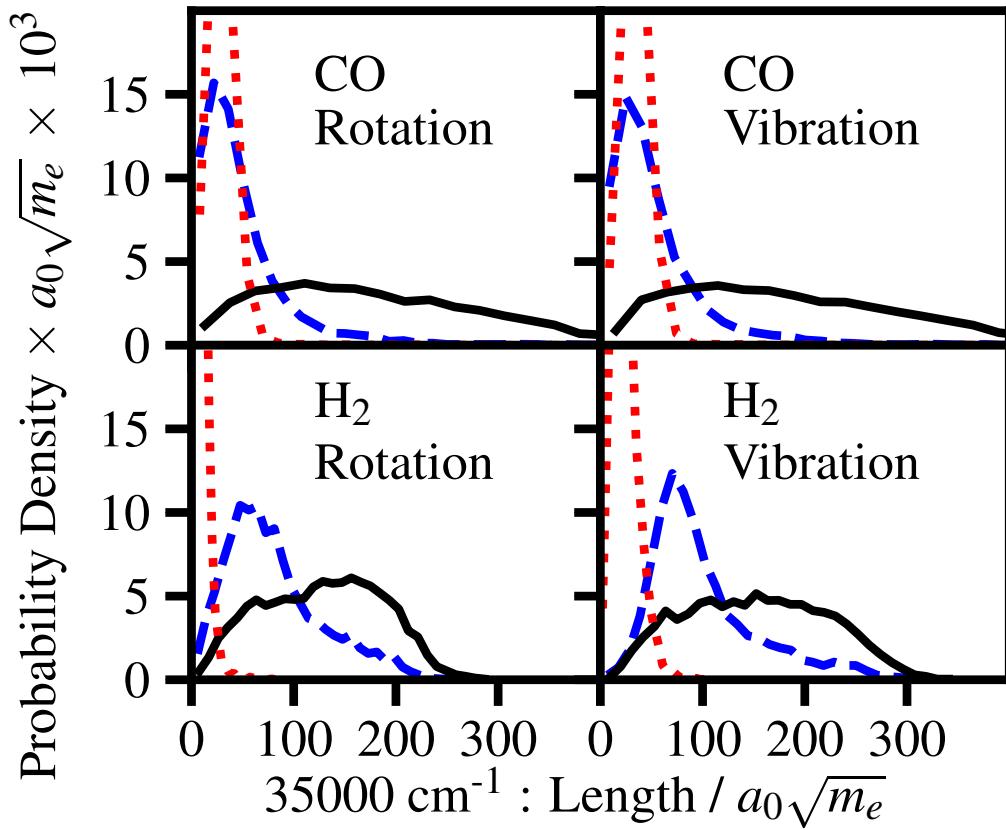


Figure 3.14: Component length distributions

Probability distributions for component kinematic lengths computed at $E_L = 35000 \text{ cm}^{-1}$. The figure provides a more detailed look at Figure 3.13. As there, **direct** geodesics are in dotted red, **radical** geodesics are in dashed blue, and **roaming** geodesics are in solid black. Again, the geodesics are only measured while in the bird's nest around the reactive well. As an aside, even though the chemically distinct species indicated by our labels (H_2 , CO) do not meaningfully exist, we can still measure motion in the corresponding degrees of freedom, e.g.: H–H stretching or change in C–O orientation. As in Figure 3.13 for total kinematic length, the roaming geodesics have a much broader range of motions available to them. Though they appear nearly indistinguishable to those distributions in our recent paper [1] (where we used $E_L = 34928.7 \text{ cm}^{-1}$ unequal to $E_\mu = 37000 \text{ cm}^{-1}$) the distributions in the figure were calculated with landscape and microcanonical energies equal, $E_L = E_\mu = 35000 \text{ cm}^{-1}$.

the roaming region than to radical products. This suggests that roaming is not simply aborted radical dissociation.

Even though geodesics cannot see into the future, they reveal sharp distinction between the bird's-nest-dynamics for the roaming channel and for direct or radical dissociation. The distributions in Figures 3.14 and 3.13 indicate that of the three channels, only roaming admits a wide array of partitionings of kinetic energy. In contrast, the direct and radical channels have narrowly defined requirements on the pre-dissociation distribution of kinetic energy.

Boundary fraction

We can understand *why* the roaming channel has this broad distribution of kinetic energies through distribution of boundary fractions, $\langle \Theta \rangle$. Recall, from eq. 2.90 in Section 2.4.4 that we define the boundary fraction as:

$$\langle \Theta \rangle = \frac{\int_0^{t_b} d\tau [2T_{\text{Total}}]^{1/2} \Theta[\underline{R}(\tau); E_L]}{\int_0^t d\tau [2T_{\text{Total}}]^{1/2}} \quad (3.11)$$

where t_b is the moment of exit from the bird's nest and where $\Theta[\underline{R}; E_L]$ is finite for the portions of the geodesic, $\underline{R}(\tau)$ encountering an obstacle:

$$\Theta[\underline{R}; E_L] = \begin{cases} 1 & E_L - V(\underline{R}) < \delta E_L \\ 0 & \text{otherwise} \end{cases}$$

We take the fractional deviation from the landscape energy, $\delta = 10^{-4}$; reducing δ to 10^{-5} produces no changes to the distribution.

In Figure 3.15 we find another stark difference between roaming and the direct or radical pathways. All three classes of pathway have a peak in the density at zero; *i.e.* for each channel, there are at least some geodesics that escape from the bird's nest via the straight-line, no-obstacles path. However, is much more pronounced for direct and radical paths than for roaming. Of the radical distribution, 56% of the density is identically 0, as is 74% for direct dissociation¹¹. In contrast, less than 14% of the roaming geodesics escape the bird's nest without encountering an obstacle. Aside from this binary observation, we see that the distribution for roaming geodesics is heavily

¹¹See Table 3.15 for these figures as a function of landscape energy.

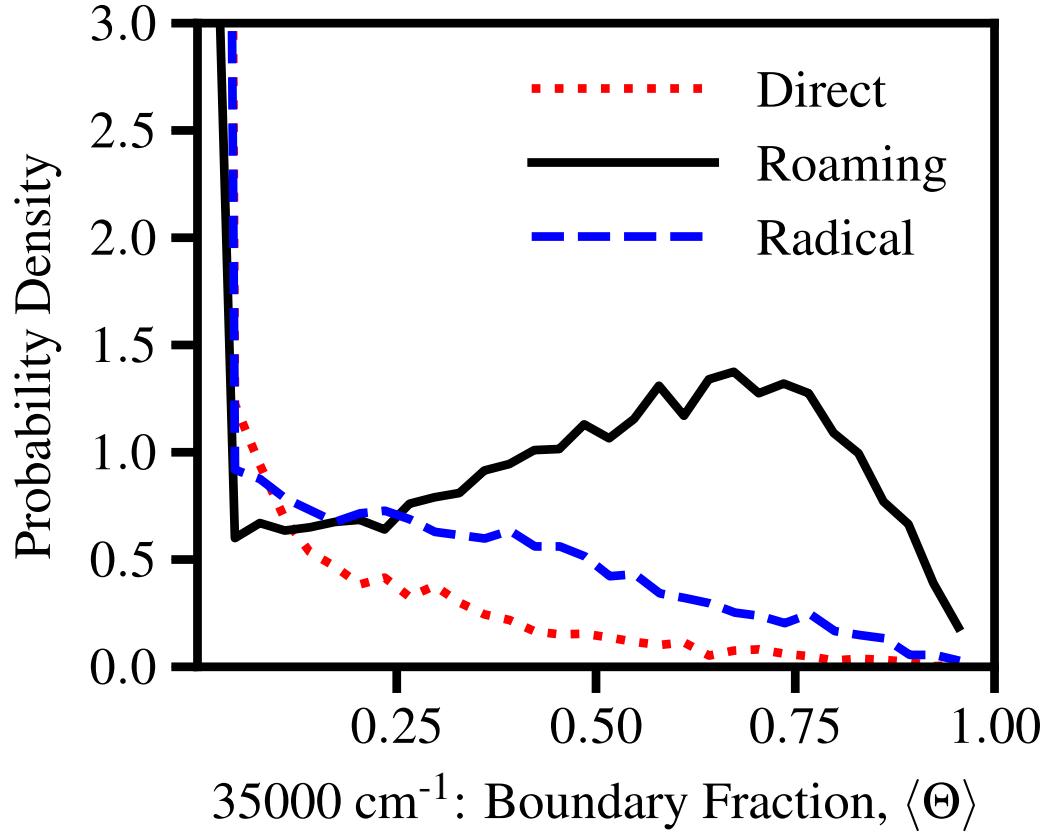


Figure 3.15: Boundary fraction

Probability distributions for the fraction of length spent tracing boundaries of the potential surface within the bird's nest for geodesics computed at $E_L = 35000 \text{ cm}^{-1}$. The distributions for all three channels include some geodesics that make no contact with obstacles before exiting the bird's nest. This delta function contribution is shown here with finite width because of our binning. In contrast to **direct** and **radical** geodesics, the majority of the density for roaming geodesics is finite, and skewed towards contacting obstacles half of the time. Indeed the mean boundary fraction for roaming geodesics is 0.45, whereas for radical and direct geodesics the figures are respectively 0.15 and 0.057 (see Table 3.16). Though they appear nearly indistinguishable to those distributions in our recent paper [1] (where we used $E_L = 34928.7 \text{ cm}^{-1}$ unequal to $E_\mu = 37000 \text{ cm}^{-1}$) the distributions in the figure were calculated with landscape and microcanonical energies equal, $E_L = E_\mu = 35000 \text{ cm}^{-1}$.

weighted towards higher boundary fraction: the most probable being $\langle \Theta \rangle = 0.70 \pm 0.05$. Winding around obstacles in this way is certainly consistent with the longer paths evidenced in Figure 3.13. It is important to note that the obstacles encountered by roaming geodesics are regions of quite high potential energy; they must be of the order of the landscape energy, which, at 35000 cm^{-1} , is already larger than the radical dissociation threshold.

Taken together, the observation of broad length distributions and obstacle-lined pathways to roaming offer a picture of the potential energy landscape where roaming occurs. By comparison to radical or direct dissociation, the paths are tortured and constrained. These obstacles give rise to a broad, unfocused distribution kinematic lengths. The distinction between paths leading to direct or radical dissociation and those leading to roaming is how much more narrowly constrained are the former than the wandering latter. The relative freedom afforded roaming paths at this energy points to an entropic advantage for the channel. We will explore this notion of an entropic advantage quantitatively in Section 3.6.5.

3.6.3 The null effect of dynamical endpoint correlation

The results in this section were computed in conjunction with our earlier paper [1] and correspond to a different set of geodesics than the rest described in this chapter. The major difference is that, unlike everywhere else, the microcanonical and landscape energies are *not* the same. Rather the landscape energy, $E_L = 34928.7 \text{ cm}^{-1}$, is considerably lower than the microcanonical energy, $E_\mu = 37600 \text{ cm}^{-1}$. In that paper [1] and for this subsection, we are nevertheless able to draw useful conclusions about the nature of geodesics on the formaldehyde energy surface. If, in this section, the numbers of molecular dynamics trajectories or computed geodesics seem unfamiliar, this is also why.

We compute geodesics as described in Section 2.3.4: from reactants to molecular and radical products, perhaps by way of the roaming region. In all cases, we use endpoints that are not dynamically correlated—that are not connected by classical trajectories¹². We do this to increase the number of geodesics we can compute from a

¹²While all the boundary- (and intermediate roaming-) points are *computed* via classical trajectories, the boundaries for a given geodesic do not necessarily come from the *same* classical trajectory.

	unpaired	paired
direct	92 ± 36 (49,060)	94 ± 36 (5,965)
		95 ± 37 (1,491)
roaming	358 ± 146 (36,608)	224 ± 117 (275)

Table 3.13: Comparison to dynamically correlated (paired) geodesics

Total lengths are reported in units of $a_0 m_e^{1/2}$ as mean \pm standard-deviation with the number of samples (in parentheses). In keeping with the rest of our analysis, lengths are computed only within the bird’s nest.

given number of trajectories (N trajectories yield $N^2/2$ endpoint pairs).

To verify that our results are unaffected by this randomization, we also compute direct and roaming geodesics between “paired” endpoints—configuration space points linked by a single molecular dynamics trajectory. For direct geodesics we select 6,000 of the 7,595 directly dissociating trajectories¹³ at $E_\mu = 37600 \text{ cm}^{-1}$, and attempt to compute geodesics between their starting and ending configurations. Our algorithm successfully locates geodesics for 5,965 of them. In the case of roaming trajectories—which are rare events—the requirement of dynamical connection drastically reduces our sampling statistics¹⁴. Of the 381 roaming trajectories¹³ at $E_\mu = 37600 \text{ cm}^{-1}$, 292 have points in the roaming region that also satisfy the landscape criterion, $E_L = 34928.7 \text{ cm}^{-1}$. For each roaming trajectory, we randomly select a single configuration in the roaming region and satisfying the landscape criterion and then pair this with the starting and ending configurations of the trajectory. The results of these calculations are summarized in Table 3.13 in the column for paired geodesics.

To determine if the breadth of the paired, direct distribution is an artifact of our statistics, we select a random subset of 1/4 of the geodesics and repeat the measurement. If the breadth is statistical, we expect the standard deviation to double. However, we find that the mean and variance are essentially identical, suggesting that the origin of the width is indeed physical.

Our primary finding is that dynamical correlation has no impact on the distribution

¹³The number of trajectories is different than that reported in Table 3.10 because this is a different set of molecular dynamics trajectories; see above.

¹⁴And for this reason do we go through all the trouble described in Section 2.3.3 to sample the roaming region via a random walk.

of total length for direct geodesics. The data suggest that there is perhaps a modest impact for the distribution of total lengths for roaming geodesics. That the dynamically correlated (paired) endpoints produce a broader distribution of roaming lengths than the unpaired endpoints may be tied to in the degree to which the space is connected (see Section 3.6). Our algorithm successfully finds roaming geodesics between uncorrelated endpoints less often than for correlated endpoints. The success rate for paired geodesics here is 94%, while we would expect closer to 70% for the same energy with unpaired endpoints (see Table 3.12). It is likely that all sub-spaces of the roaming region are not uniformly connected.

The possibility of disconnect aside, qualitatively, we see that the distribution of lengths of roaming geodesics is much broader than for direct geodesics even when computed between dynamically linked endpoints. This is the essential observation upon which our analyses rest and provided that we are consistent in our methods, we expect the impacts of dynamical correlation to be minimal.

3.6.4 Trends with increasing landscape energy

At low energies, we observe that the geodesics corresponding to roaming pathways have a broad distribution of kinematic lengths. Direct and radical geodesics seem much more narrowly constrained and have sharply peaked distributions in kinematic length. We conclude then that roaming has an entropic advantage over the other dissociative channels at low energy. In this section we explore what happens as energy increases.

Figure 3.16 gives the distributions for the total kinematic length for geodesics in the bird’s nest at the upper end of our energy scale, 45000 cm^{-1} (in analogy to Figure 3.13, which shows low-energy results). The roaming distribution at high energy could not be more changed from its low-energy form. It has narrowed and the mean kinematic length has shortened considerably. The distribution of kinematic lengths for roaming geodesics at 45000 cm^{-1} more closely resembles the radical distribution at all other sampled energies than it resembles the roaming distribution at 35000 cm^{-1} . It seems that the singular aspect of the roaming pathway has diminished at this high energy.

In Figure 3.17, we see distributions of kinematic length for each of the three channels at all of the energies under study. The distributions for energies intervening between 35000 cm^{-1} and 45000 cm^{-1} effectively interpolate between the extremal distributions.

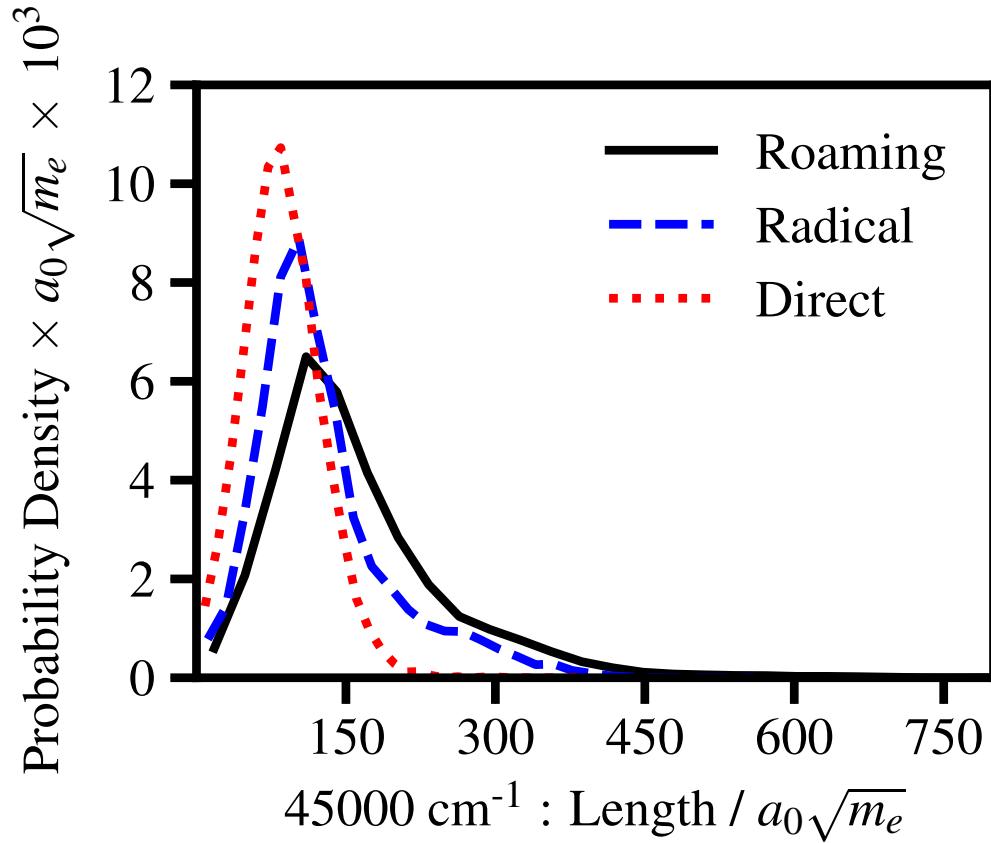


Figure 3.16: Total length distributions in the high-energy limit

Probability distributions for total kinematic lengths of geodesics calculated at $E_L = 45000 \text{ cm}^{-1}$ within the bird's nest. The figure is a high-energy analog of Figure 3.13, which was for $E_L = 35000 \text{ cm}^{-1}$. In stark contrast to the low-energy figure, here we see the distribution of roaming kinematic collapsed towards those of **radical** and **direct** dissociation.

The primary observation is that with increase in landscape energy, the roaming distributions of lengths converge towards the distribution of radical lengths. The distributions for the radical and direct pathways are largely unchanged. It appears as though the entropic advantage possessed by the roaming channel at low energy is lost as the energy increases. The information in the figure is recapitulated in Table 3.14, which gives the mean and standard deviation of the distributions by energy and channel. Though it is not easy to discern from Figure 3.17, the tabulated lengths show that the mean length of the direct geodesics does indeed decrease with rising landscape energy.

In the component length distributions (Figure 3.18), we see much the same high

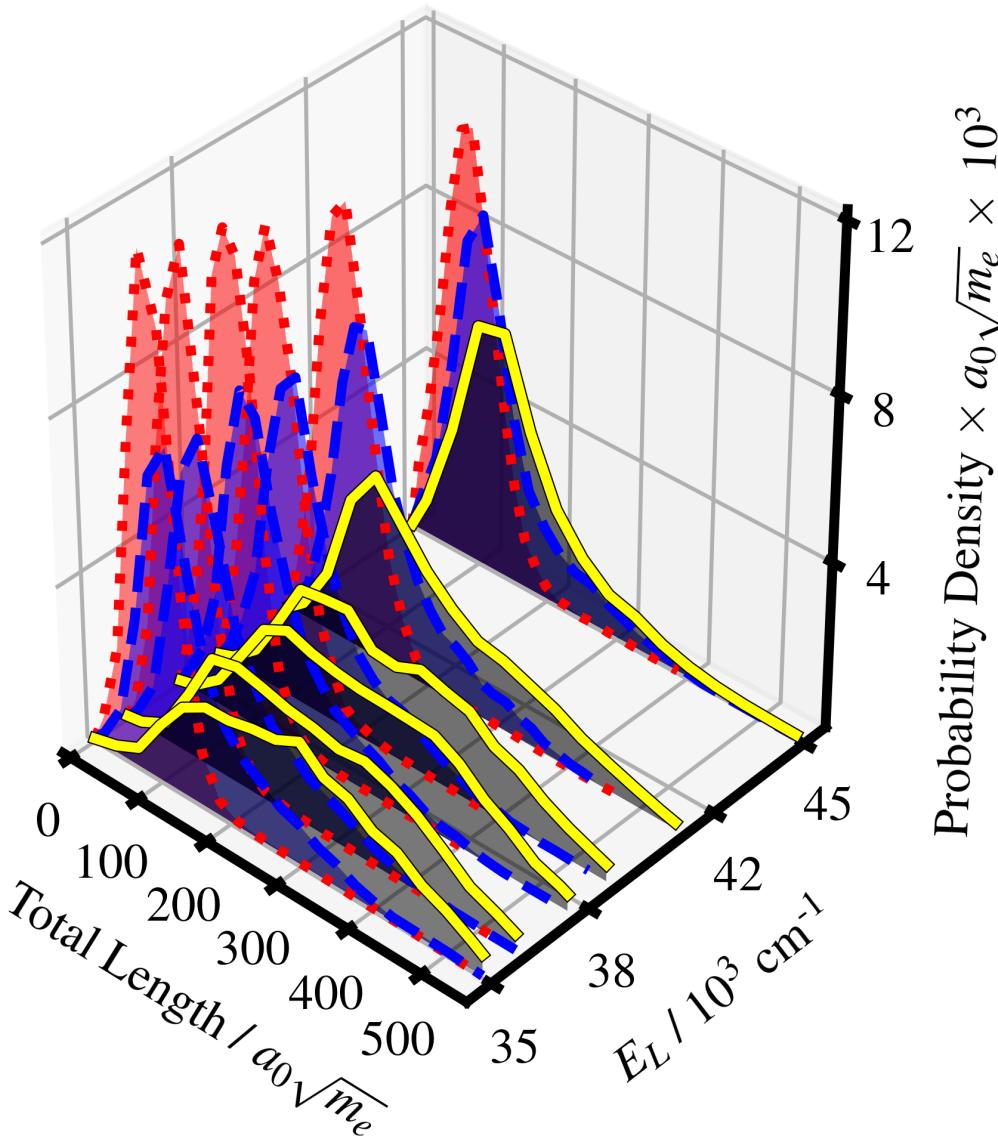


Figure 3.17: Convergence of total length distributions

Total kinematic length distributions as a function of energy for geodesics while confined to the bird's nest. The distributions of lengths of **direct** geodesics are shown in dotted red. The dashed blue curves give the **radical** distributions. Roaming distributions are the black fields with solid, yellow bounds. While the direct and radical distributions are relatively static with increasing energy, the roaming distribution undergoes a dramatic shift towards a narrower distribution with smaller mean. The collapse towards the radical length distribution signals that the roaming channel no longer has the entropic advantage it did at lower energies.

Energy / cm ⁻¹	direct	radical	roaming
35000	95.79 ± 38.47	149.2 ± 79.05	300.0 ± 132.4
36000	93.37 ± 39.35	148.9 ± 78.24	283.2 ± 131.1
37600	86.29 ± 36.49	141.8 ± 73.58	283.1 ± 133.7
38814	89.89 ± 39.21	144.1 ± 73.81	273.7 ± 135.3
41010	89.80 ± 40.18	141.2 ± 72.64	220.2 ± 124.7
45000	86.64 ± 38.83	126.9 ± 69.56	158.1 ± 91.4

Table 3.14: Mean total kinematic length for geodesics by channel and energy

We give the mean and standard deviation for all the distributions of total kinematic length given in Figures 3.17 (and 3.13 and 3.13). As there, these are lengths for the geodesics while in the bird's nest. Kinematic lengths are reported as mean ± standard-deviation in units of $a_0 m_e^{1/2}$. From the table as from the figures, we see the pronounced contraction of the distribution of roaming kinetic lengths with increasing energy.

energy trend. The roaming distributions have moved toward the radical distribution. For the H₂ degrees of freedom, the roaming and radical distributions are nearly coincident. While the overlap for the CO degrees of freedom is less pronounced, comparison with the distribution at 35000 cm⁻¹, (Figure 3.14) shows that the reorganization is still quite dramatic.

The picture coming together is that at high energies, roaming dynamics during the predissociation incubation period mirror those of radical dissociation, particularly in the terms of the motions of the hydrogen atoms. We can again look to the distribution of boundary fractions to understand the way that the energy surface influences these dynamics. Figure 3.19 shows the high energy result in detail and Figure 3.20 shows all the energies in our study. At 45000 cm⁻¹, the distribution of roaming boundary fractions shows marked depression, a trend also visible in Table 3.16. From the figures and from Table 3.15 we also see that much of the density has moved into the delta peak at $\langle \Theta \rangle = 0$.

The roaming boundary fraction density at high energy, most closely resembles the distribution for the radical boundary fraction at 35000 cm⁻¹ (Figure 3.15). The large obstacles that broadened the roaming length distributions at lower energies are gone and have lost their influence.

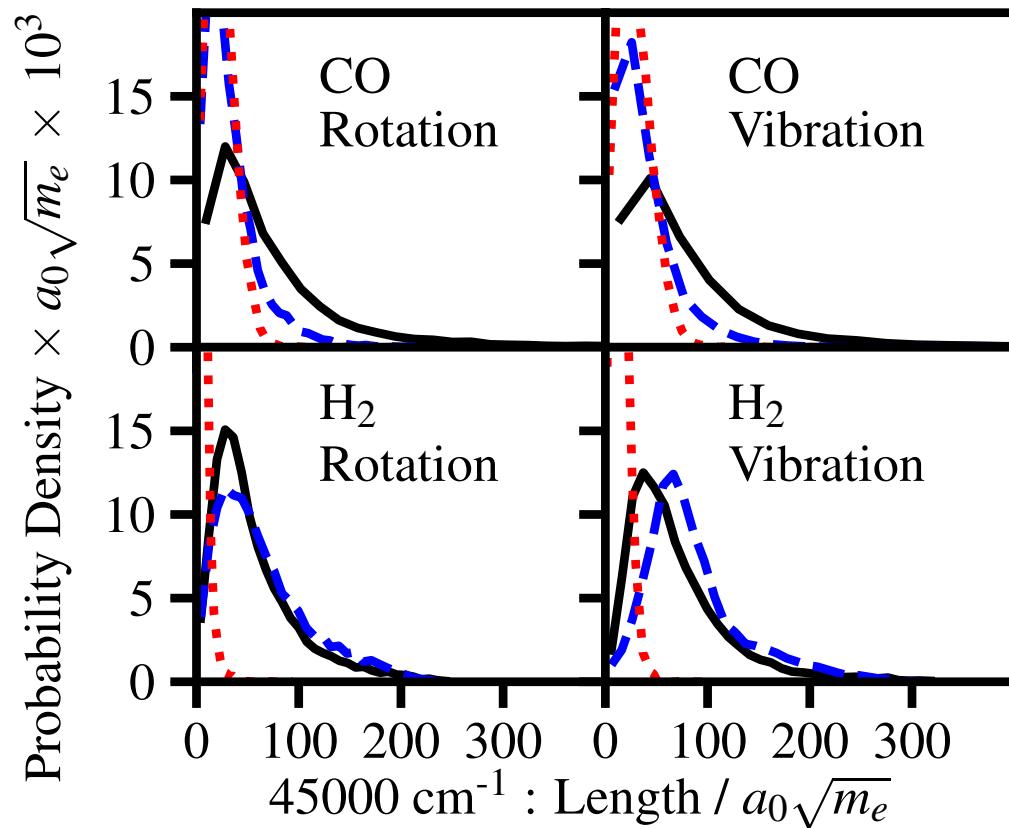


Figure 3.18: Component length distributions in the high-energy limit

Offering a more detailed view of Figure 3.16, we show the component kinematic lengths for geodesics computed at $E_L = 45000 \text{ cm}^{-1}$ in the bird's nest. Direct geodesics are in dotted red, radical geodesics are in dashed blue, and roaming geodesics are in solid black. Comparing the roaming distributions above to their low-energy analogs in Figure 3.14, we see the same dramatic change that was evident in the distribution of total kinematic lengths in Figure 3.17. Here, as in Figure 3.16, the roaming distributions now bear striking similarity (rather than difference) to the radical distribution.

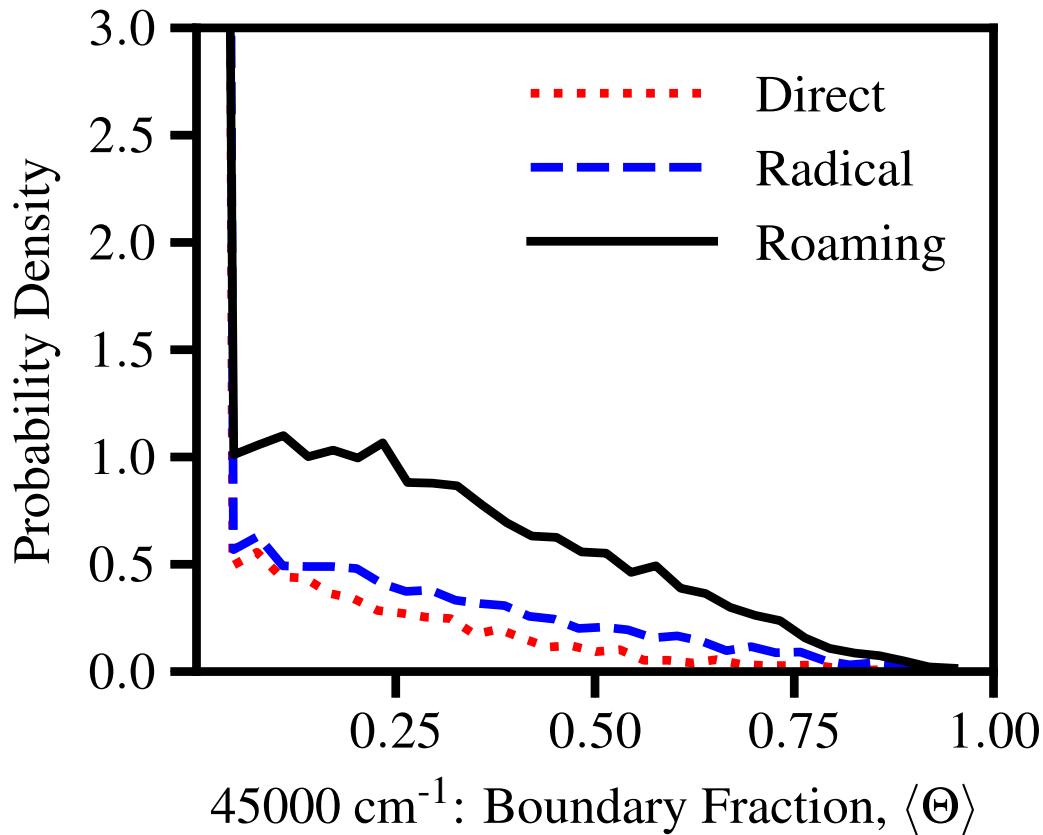


Figure 3.19: Boundary fraction in the high-energy limit

Probability distributions for the fraction of length spent tracing boundaries of the potential surface within the bird's nest for geodesics computed at $E_L = 45000 \text{ cm}^{-1}$. In this high-energy limit, the contrast evident in Figure 3.15, between the roaming distribution and those of the **direct** and **radical** channels, has faded. The fraction of each distribution contributing to the delta function at $\langle \Theta \rangle = 0$ has grown for all channels, but by the largest amount for the roaming channel. The portion of the density identically 0 for roaming geodesics has climbed to 0.48; for radical and direct geodesics the figures are respectively 0.76 and 0.84 (see Table 3.15).

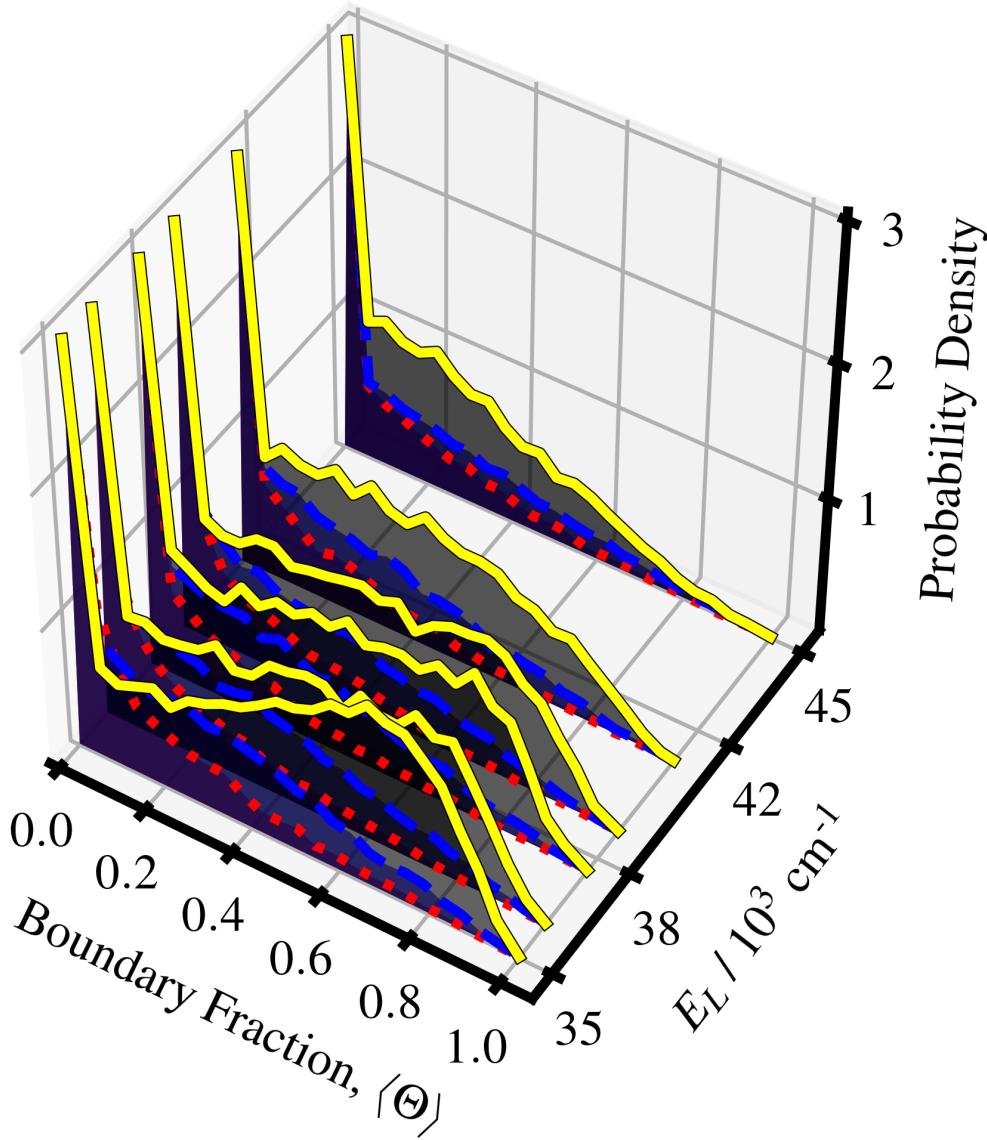


Figure 3.20: Convergence of distribution of boundary fractions

Boundary fraction distributions as a function of energy for geodesics while confined to the bird's nest. The distributions for **direct** geodesics are shown in dotted red. The dashed blue curves give the **radical** distributions. Roaming distributions are the black fields with solid, yellow bounds. Similarly to the total length distributions shown in Figure 3.17, the direct and radical distributions change little as energy increases, but the roaming distribution essentially inverts its center: over the energy range we study, the mean boundary fraction for roaming shifts from $\langle \Theta \rangle = 0.45$ at $E_L = 35000 \text{ cm}^{-1}$ to $\langle \Theta \rangle = 0.17$ at $E_L = 45000 \text{ cm}^{-1}$. At higher landscape energies, there is a paucity of large obstacles for the roaming geodesics to impact as they make their way to the roaming region. The high-energy landscape encountered by roaming paths is more similar to that of the radical and direct paths than it is different.

Energy / cm ⁻¹	direct	radical	roaming
35000	0.741	0.556	0.134
36000	0.766	0.540	0.177
37600	0.828	0.587	0.182
38814	0.799	0.593	0.210
41010	0.810	0.640	0.293
45000	0.843	0.758	0.476

Table 3.15: Portion of identically zero boundary fraction vs. energy

As shown in Figures 3.15, 3.19, and 3.20, all of the distributions of boundary fraction, $\langle \Theta \rangle$, have some non-trivial portion of the density identically 0. That is, at all energies, for all channels, *some* geodesics exit the bird's nest without encountering an obstacle. The table above gives the proportion of each distribution that makes up the delta function contribution at $\langle \Theta \rangle = 0$ by channel and energy.

Energy / cm ⁻¹	direct	radical	roaming
35000	0.0572	0.149	0.446
36000	0.0539	0.155	0.391
37600	0.0383	0.134	0.375
38814	0.0450	0.131	0.351
41010	0.0438	0.111	0.284
45000	0.0369	0.0707	0.165

Table 3.16: Mean boundary fraction vs. energy

For the boundary fractions shown in Figures 3.15, 3.19, and 3.20, we give the means of the distributions as a function of energy by channel. Evident are modest and barely perceptible shifts for the radical and direct channels as well as the transformation in the boundary fraction distribution for roaming geodesics.

Energy / cm ⁻¹	S^{direct}/k_B	S^{radical}/k_B	S^{roaming}/k_B
35000	4.6448	5.3179	5.9962
36000	4.6744	5.3312	5.9507
37600	4.6439	5.2765	5.9977
38814	4.7565	5.2986	5.9996
41010	4.7763	5.2935	5.8328
45000	4.7739	5.2168	5.4787

Table 3.17: Entropy of total length distributions vs. energy

3.6.5 Convergence of entropy of distributions with energy

We suspect that the broad, roaming distributions in kinematic length are of higher entropy than the sharply peaked distributions for the direct or radical channels. Recall that the entropy of a continuous distribution, $\rho(\ell)$, is given by (eq. 2.123):

$$S/k_B = - \int d\ell \rho(\ell) \log [\rho(\ell)] \quad (3.12)$$

Recall also, for dimensioned $\rho(\ell)$, there is always an undetermined and constant offset, which render the absolute magnitudes of entropy without meaning, though the differences do convey information. Following the method in Section 2.8 to estimate the entropy of a sampled distribution, we compute and compare the entropies of the length distributions as a function of energy. Figures 3.21 and 3.22 give the results of such analysis for total length and component distributions within the bird's nest respectively; Table 3.17 gives the numeric results for the total lengths.

The qualitative observations of the previous section are largely borne out quantitatively. Firstly, the roaming distributions for total length, have entropy of substantively larger absolute magnitude than the distributions for direct or radical dissociation. This is consistent with our notion of roaming having some kind of entropic advantage. We also see that the roaming channel has the greatest shift in entropy. Over the 10000 cm⁻¹ that span our energies of study, roaming entropy drops by roughly 0.5 entropy units. Over the same energy window, the direct and radical channels have much smaller change: on the order of 0.1 units. The magnitude of the change for the roaming channel is substantial. We also see the roaming entropy converging towards that of the radical distribution for total length. In the component distributions for the CO and H₂ degrees

of freedom, we see much the same story with higher roaming entropy, and entropies for the roaming channel converging towards that of the radical channel.

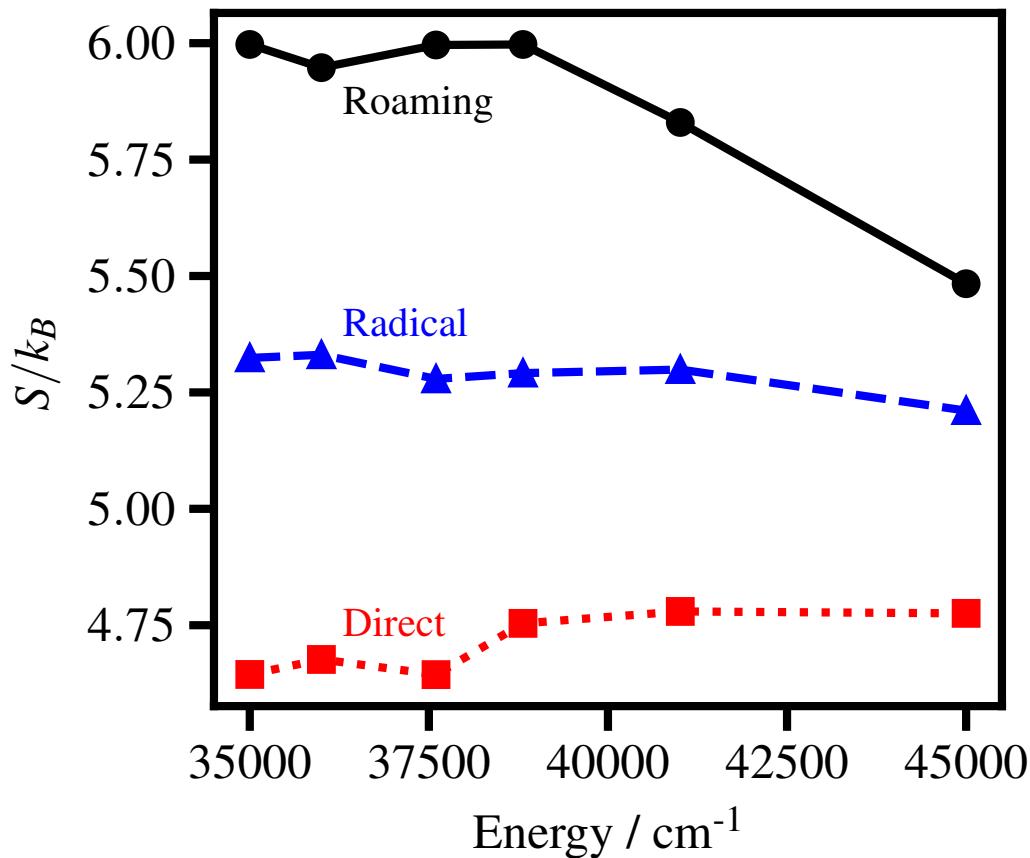


Figure 3.21: Entropy of total length distributions

Absolute magnitude of entropy of total kinematic length distributions of geodesics in the bird's nest. Here we see that the entropy of the roaming distributions is greater than for either of the other channels. We also see that change in magnitude is much larger for roaming than for the direct pathway or the radical paths. As the entropy of the roaming distribution decreases with energy, we see that the entropy of the roaming distribution does indeed approach that of the radical distribution.

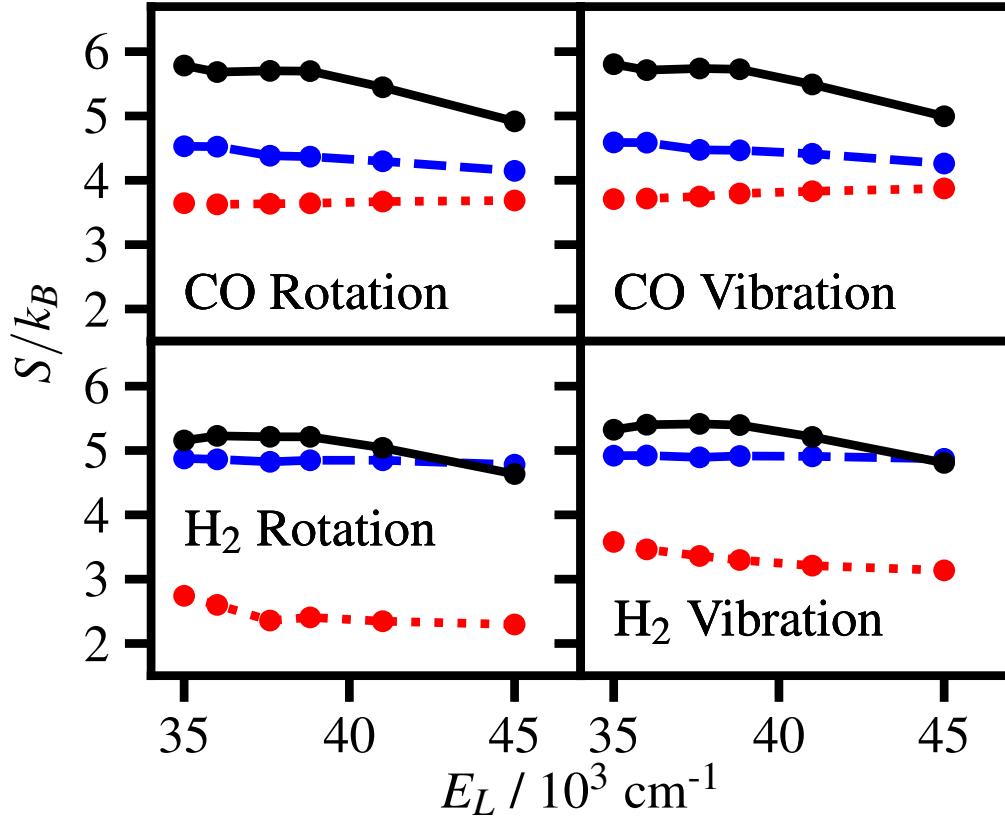


Figure 3.22: Entropy of component length distributions

Entropies for distributions of component lengths for the three channels while in the bird's nest: roaming in solid black, radical in dashed blue, and direct in dotted red. Distributions are formed for the lengths, ℓ_α as described in Sections 2.4.2 and 2.4.3. Again, we use the component kinematic lengths to provide more detail to the picture from the total kinematic length. As in Figure 3.21, we see that the absolute magnitude of the entropy of the roaming distributions is higher than for geodesics for either the direct or radical channels. We also see that in the CO as well as H₂ degrees of freedom, entropy of the roaming distributions approaching that of the radical channel.

The distinction between the CO and H₂ degrees of freedom is particularly of note—such a divergence is usually not present. The CO entropies tell much the same story as the entropy of the total distributions. However, in the H₂ components, the direct pathway is widely separated from the radical and roaming channels. We also see decline in the H₂ entropies for the direct pathway at low energies before leveling off; roaming entropies, in contrast, are roughly constant at low energies before declining at higher energies.

3.7 Roaming as an entropically mediated phenomenon

The foregoing observations, based on the statistical properties of geodesics on the formaldehyde energy surface lead us to conclude that, at least in formaldehyde, roaming is an entropically mediated phenomenon.

At lower energies, below 38814 cm^{-1} , dynamical effects dominate—the competition between the radical and direct channels is evident in this energy range in Figure 3.4. Once the dynamical effects have stabilized, statistical effects become dominant. At energies of 38814 cm^{-1} and higher, the branching ratios for the direct and radical channels are roughly constant, but roaming begins a decline, roughly halving in frequency up to 45000 cm^{-1} . Just as the incidence of roaming drops from 38814 cm^{-1} , so too does the entropy of the distribution of lengths of roaming geodesics decline; see Figure 3.21.

The entropy of distributions of geodesic lengths in the bird's nest predict the direction of change in the branching ratio to roaming. This is a remarkable result. Even though the geodesics have no knowledge of the future course the system will take, their distribution through the bird's nest reveals the relative change in frequency of roaming.

CHAPTER 4

CONCLUDING REMARKS

At the outset, our goal in this work was to contribute to the understanding of the story of roaming in formaldehyde. We take the perspective that such understanding means apprehending the dynamical peculiarities that give rise to roaming and comes from characterizing the inherent dynamics of the formaldehyde system. The geodesic formalism enables us to globally probe the potential energy landscape of formaldehyde. Through this window, we see that roaming dynamics are of greater entropy—are far less constrained—than the pathways giving rise to direct or radical dissociation. Our observation may shine some light on why roaming seems to be prevalent in many systems. From the landscape perspective, all that is required for roaming-like (broadly distributed) geodesics is a sufficiently flat, asymptotic region of the potential in a system of high enough dimensionality that arriving there is an entropic (navigational) challenge. It is then not at all surprising that roaming is a recurrent phenomenon.

4.1 New methods, new observations

We have provided evidence that roaming in formaldehyde is an entropically mediated mechanism. The roaming system has a far greater range of dynamical options than when dissociating via either the direct pathway to molecular products or by the radical channel. This entropic advantage correlates with the incidence of roaming in formaldehyde. Indeed, as the entropy difference between distributions of geodesic lengths for the radical and roaming channels decreases with energy, the frequency of roaming in the

system also decreases.

In our work, we also expand the range of dynamical systems to be treated with geodesics to include small molecules. We derive highly accurate lower bounds on the lengths of these small-molecule geodesics. We also visualize geodesic paths such that their correspondence to the inherent dynamics is clear. Our study is the first time geodesics have been used in a system with heterogeneous masses. We confront and resolve the attendant problems associated with center of mass conservation and present two other general methods for doing so.

4.2 Future directions

Clearly the first angle of attack for extending our work involves quantitatively connecting the entropic relationships we have established to other observables. It may be that the mutual information between the distributions of roaming, radical, and direct geodesic lengths can be tied to the relative incubation periods or branching ratios for each channel. Such a connection would be of interest, not just to those theorists and experimentalists desiring to calculate these properties, but also to those wishing to understand the ways that geodesics carry information about the dynamics of small-molecule systems. The theory of large deviations [80] or even ideas of surprisal analysis [81] may be fruitful sources of direction.

Another important avenue of inquiry is repeating some of this work in other roaming systems. Acetaldehyde, CH_3HCO , is an obvious target because of its structural similarity to formaldehyde and because a global potential surface has already been developed [33]. As an aside, such a potential is not required as our methods are amenable to “on-the-fly” calculations of potentials and gradients from first principles.

While the notion of roaming as aborted radical dissociation is not useful for understanding the dynamical influences at play, it does make a useful point about the energetics involved: roaming fragments have insufficient translational energy to escape the parent. Surely, during the incubation period prior to dissociation, energy is distributed throughout different degrees of freedom of formaldehyde. This internal energy redistribution is part of what controls whether or not the system will roam. By centering our analysis in the bird’s nest, we have focused our attention squarely on

this critical period. One might wonder what geodesics can say about the organization of energy flow during this time. Even though geodesics are *static* objects and exist in configuration-space, where there is no notion of time or speed, it may be that changes in their *direction* are correlated with energy transfer. Some hope for this thought is provided by Miller’s work on the Reaction Path Hamiltonian [67], where the curvature of an intrinsic reaction coordinate appears as a term in the kinetic energy.

In recent years, there has been some hand-wringing [82, 83] over the notion of “post transition state bifurcations”—instances where multiple, chemically distinct products arise after traversing what can be identified only as a single transition state. The authors are surely sensible in pointing out that dynamical factors and not just potential barriers and basins are influencing the outcomes of these reactions. However, they may be premature in suggesting that the apparent futility of studying the structure of the surface leaves us no tools other than trajectory simulations. As we have seen, even though individual geodesics cannot forecast the future, by globally characterizing the H₂CO surface and analyzing their properties *while confined to the bird’s nest* we are still able to discriminate between the reactive channels, which are yet-to-be-chosen. The entire history and future of a classical trajectory are specified by a single point in phase space and knowledge of the potential. Because geodesics offer a global perspective on that potential, they may be well positioned to shed light on what happens after a transition state by looking before it.

Bibliography

- [1] D. V. Cofer-Shabica, R. M. Stratt. What is special about how roaming chemical reactions traverse their potential surfaces? Differences in geodesic paths between roaming and non-roaming events. *The Journal of Chemical Physics* **2017**, *146*, 214303.
- [2] J. M. Budarz, M. P. Minitti, D. V. Cofer-Shabica, B. Stankus, A. Kirrander, J. B. Hastings, P. M. Weber. Observation of femtosecond molecular dynamics via pump-probe gas phase x-ray scattering. *Journal of Physics B: Atomic Molecular and Optical Physics* **2016**, *49*.
- [3] D. V. Cofer-Shabica, R. M. Stratt, *The geometries of potential energy landscapes imply dynamical signatures for roaming reactions*, American Chemical Society, 250th National Meeting, Boston, MA, 2015, PHYS 554.
- [4] M. P. Minitti, J. M. Budarz, A. Kirrander, J. Robinson, T. J. Lane, D. Ratner, K. Saita, T. Norhey, B. Stankus, V. Cofer-Shabica, J. Hastings, P. M. Weber. Toward structural femtosecond chemical dynamics: Imaging chemistry in space and time. *Faraday Discussions* **2014**, *171*, 81–91.
- [5] D. V. Cofer-Shabica, *Wandering Molecules*, Brown University, Research Matters, Providence, RI, **2014**. Invited Talk. <https://www.youtube.com/watch?v=X3xyMP9Eaco>.
- [6] D. Townsend, S. A. Lahankar, S. K. Lee, S. D. Chambreau, A. G. Suits, X. Zhang, J. Rheinecker, L. B. Harding, J. M. Bowman. The Roaming Atom: Straying from the Reaction Path in Formaldehyde Decomposition. *Science* **2004**, *306*, 1158–1161.

- [7] J. M. Bowman, B. C. Shepler. Roaming Radicals. *Annu. Rev. Phys. Chem.* **2011**, *62*, 531–553.
- [8] C. Wang, R. M. Stratt. Global perspectives on the energy landscapes of liquids, supercooled liquids, and glassy systems: Geodesic pathways through the potential energy landscape. *The Journal of Chemical Physics* **2007**, *127*, 224504.
- [9] R. D. van Zee, M. F. Foltz, C. B. Moore. Evidence for a second molecular channel in the fragmentation of formaldehyde. *J. Chem. Phys.* **1993**, *99*, 1664–1673.
- [10] E. Wigner. The transition state method. *Trans. Faraday Soc.* **1938**, *34*, 29–41.
- [11] W. M. Gelbart, M. L. Elert, D. F. Heller. Photodissociation of the formaldehyde molecule: does it or doesn't it? *Chemical Reviews* **1980**, *80*, 403–416.
- [12] K.-M. Weitzel. On the distinction between tight and loose transition states in unimolecular dissociations. *Chemical Physics Letters* **1991**, *186*, 490 – 494.
- [13] X. Li, J. M. Millam, H. B. Schlegel. Ab initio molecular dynamics studies of the photodissociation of formaldehyde, $\text{H}_2\text{CO} \longrightarrow \text{H}_2 + \text{CO}$: Direct classical trajectory calculations by MP2 and density functional theory. *The Journal of Chemical Physics* **2000**, *113*, 10062–10067.
- [14] J. Rheinecker, X. Zhang, J. Bowman. Quasiclassical trajectory studies of the dynamics of H_2CO on a global ab initio-based potential energy surface. *Molecular Physics* **2005**, *103*, 1067–1074.
- [15] P. Zhang, S. Maeda, K. Morokuma, B. J. Braams. Photochemical reactions of the low-lying excited states of formaldehyde: T1/S0 intersystem crossings, characteristics of the S1 and T1 potential energy surfaces, and a global T1 potential energy surface. *J. Chem. Phys.* **2009**, *130*, 114304.
- [16] S. A. Lahankar, S. D. Chambreau, X. Zhang, J. M. Bowman, A. G. Suits. Energy dependence of the roaming atom pathway in formaldehyde decomposition. *The Journal of Chemical Physics* **2007**, *126*, 044314.

- [17] S. A. Lahankar, V. Goncharov, F. Suits, J. D. Farnum, J. M. Bowman, A. G. Suits. Further aspects of the roaming mechanism in formaldehyde dissociation. *Chem. Phys.* **2008**, *347*, 288–299.
- [18] A. C. Terentis, S. E. Waugh, G. F. Metha, S. H. Kable. HCO (N,Ka,Kc,J) distributions from near-threshold photolysis of H₂CO (J,Ka,Kc). *The Journal of Chemical Physics* **1998**, *108*, 3187–3198.
- [19] J. M. Bowman. Roaming. *Mol. Phys.* **2014**, *112*, 2516–2528.
- [20] P. L. Houston, S. H. Kable. Photodissociation of acetaldehyde as a second example of the roaming mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 16079–16082.
- [21] B. R. Heazlewood, M. J. T. Jordan, S. H. Kable, T. M. Selby, D. L. Osborn, B. C. Shepler, B. J. Braams, J. M. Bowman. Roaming is the dominant mechanism for molecular products in acetaldehyde photodissociation. *Proceedings of the National Academy of Sciences USA* **2008**, *105*, 12719–12724.
- [22] M. P. Grubb, M. L. Warter, A. G. Suits, S. W. North. Evidence of Roaming Dynamics and Multiple Channels for Molecular Elimination in NO₃ Photolysis. *J. Phys. Chem. Lett.* **2010**, *1*, 2455–2458.
- [23] M. P. Grubb, M. L. Warter, H. Xiao, S. Maeda, K. Morokuma, S. W. North. No Straight Path: Roaming in Both Ground- and Excited-State Photolytic Channels of NO₃ → NO + O₂. *Science* **2012**, *335*, 1075–1078.
- [24] B. Fu, J. M. Bowman, H. Xiao, S. Maeda, K. Morokuma. Quasiclassical Trajectory Studies of the Photodissociation Dynamics of NO₃ from the D-0 and D-1 Potential Energy Surfaces. *J. Chem. Theory Comput.* **2013**, *9*, 893–900.
- [25] D. L. Osborn. Exploring Multiple Reaction Paths to a Single Product Channel. *Adv. Chem. Phys.* **2008**, *138*, 213–265.
- [26] L. B. Harding, S. J. Klippenstein. Roaming Radical Pathways for the Decomposition of Alkanes. *J. Phys. Chem. Lett.* **2010**, *1*, 3016–3020.

- [27] A. Li, J. Li, H. Guo. Quantum Manifestation of Roaming in $\text{H} + \text{MgH} \longrightarrow \text{Mg} + \text{H}_2$: The Birth of Roaming Resonances. *J. Phys. Chem. A* **2013**, *117*, 5052–5060.
- [28] A. S. Mereshchenko, E. V. Butaeva, V. A. Borin, A. Eyzips, A. N. Tarnovsky. Roaming-mediated ultrafast isomerization of geminal tri-bromides in the gas and liquid phases. *Nature Chemistry* **2015**, *7*, 562–568.
- [29] B. Carpenter. Intramolecular Dynamics for the Organic Chemist. *Accts. Chem. Res.* **1992**, *25*, 520–528.
- [30] S. J. Klippenstein, Y. Georgievskii, L. B. Harding. Statistical Theory for the Kinetics and Dynamics of Roaming Reactions. *J. Phys. Chem. A* **2011**, *115*, 14370–14381.
- [31] L. B. Harding, S. J. Klippenstein, A. W. Jasper. Ab initio methods for reactive potential surfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4055–4070.
- [32] L. B. Harding, Y. Georgievskii, S. J. Klippenstein. Roaming Radical Kinetics in the Decomposition of Acetaldehyde. *J. Phys. Chem. A* **2010**, *114*, 765–777.
- [33] B. C. Shepler, B. J. Braams, J. M. Bowman. “Roaming” Dynamics in CH_3CHO Photodissociation Revealed on a Global Potential Energy Surface. *J. Phys. Chem. A* **2008**, *112*, 9344–9351.
- [34] D. U. Andrews, S. H. Kable, M. J. T. Jordan. A Phase Space Theory for Roaming Reactions. *J. Phys. Chem. A* **2013**, *117*, 7631–7642.
- [35] B. C. Shepler, Y. Han, J. M. Bowman. Are Roaming and Conventional Saddle Points for H_2CO and CH_3CHO Dissociation to Molecular Products Isolated from Each Other? *J. Phys. Chem. Lett.* **2011**, *2*, 834–838.
- [36] L. B. Harding, S. J. Klippenstein, A. W. Jasper. Separability of Tight and Roaming Pathways to Molecular Decomposition. *J. Phys. Chem. A* **2012**, *116*, 6967–6982.
- [37] K. Fukui. Formulation of the reaction coordinate. *J. Phys. Chem.* **1970**, *74*, 4161–4163.
- [38] H. Xiao, S. Maeda, K. Morokuma. Global ab Initio Potential Energy Surfaces for Low-Lying Doublet States of NO_3 . *J. Chem. Theory Comput.* **2012**, *8*, 2600–2605.

- [39] Z. Homayoon, J. M. Bowman. Quasiclassical Trajectory Study of CH₃NO₂ Decomposition via Roaming Mediated isomerization Using a Global Potential Energy Surface. *J. Phys. Chem. A* **2013**, *117*, 11665–11672.
- [40] H. Waalkens, S. Wiggins. Geometrical models of the phase space structures governing reaction dynamics. *Regular and Chaotic Dynamics* **2010**, *15*, 1–39.
- [41] F. A. Mauguière, P. Collins, G. S. Ezra, S. C. Farantos, S. Wiggins. Multiple transition states and roaming in ion-molecule reactions: A phase space perspective. *Chem. Phys. Lett.* **2014**, *592*, 282–287.
- [42] F. A. Mauguière, P. Collins, G. S. Ezra, S. C. Farantos, S. Wiggins. Roaming dynamics in ion-molecule reactions: Phase space reaction pathways and geometrical interpretation. *J. Chem. Phys.* **2014**, *140*.
- [43] F. A. Mauguière, P. Collins, Z. C. Kramer, B. K. Carpenter, G. S. Ezra, S. C. Farantos, S. Wiggins. Phase Space Structures Explain Hydrogen Atom Roaming in Formaldehyde Decomposition. *The Journal of Physical Chemistry Letters* **2015**, *6*, 4123–4128.
- [44] F. A. Mauguière, P. Collins, Z. C. Kramer, B. K. Carpenter, G. S. Ezra, S. C. Farantos, S. Wiggins. Roaming: A Phase Space Perspective. *Annual Review of Physical Chemistry* **2017**, *68*, 499–524.
- [45] K. Müller, L. D. Brown. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica chimica acta* **1979**, *53*, 75–93.
- [46] C. Wang, R. M. Stratt. Global perspectives on the energy landscapes of liquids, supercooled liquids, and glassy systems: The potential energy landscape ensemble. *The Journal of Chemical Physics* **2007**, *127*, 224503.
- [47] W. E, E. Vanden-Eijnden. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391–420.
- [48] C. N. Nguyen, J. I. Isaacson, K. B. Shimmyo, A. Chen, R. M. Stratt. How dominant is the most efficient pathway through the potential energy landscape of a slowly diffusing disordered system? *The Journal of Chemical Physics* **2012**, *136*, 184504.

- [49] C. N. Nguyen, R. M. Stratt. Preferential solvation dynamics in liquids: How geodesic pathways through the potential energy landscape reveal mechanistic details about solute relaxation in liquids. *The Journal of Chemical Physics* **2010**, *133*, 124503.
- [50] D. Jacobson, R. M. Stratt. The inherent dynamics of a molecular liquid: Geodesic pathways through the potential energy landscape of a liquid of linear molecules. *The Journal of Chemical Physics* **2014**, *140*, 174503.
- [51] Q. Ma, R. M. Stratt. Potential energy landscape and inherent dynamics of a hard-sphere fluid. *Phys. Rev. E* **2014**, *90*, 042314.
- [52] L. Frechette, R. M. Stratt. The inherent dynamics of isotropic- and nematic-phase liquid crystals. *The Journal of Chemical Physics* **2016**, *144*, 234505.
- [53] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, of Cambridge Molecular Science, Cambridge University Press, **2003**.
- [54] X. Zhang, S. Zou, L. B. Harding, J. M. Bowman. A Global ab Initio Potential Energy Surface for Formaldehyde. *The Journal of Physical Chemistry A* **2004**, *108*, 8980–8986.
- [55] X. Zhang, J. L. Rheinecker, J. M. Bowman. Quasiclassical trajectory study of formaldehyde unimolecular dissociation: $\text{H}_2\text{CO} \rightarrow \text{H}_2 + \text{CO}$, $\text{H} + \text{HCO}$. *The Journal of Chemical Physics* **2005**, *122*, 114313.
- [56] H. W. Kuhn, A. W. Tucker, *Nonlinear Programming in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, California, pp. 481–492.
- [57] C. Eckart. Some Studies Concerning Rotating Axes and Polyatomic Molecules. *Phys. Rev.* **1935**, *47*, 552–558.
- [58] P. L. Houston, R. Conte, J. M. Bowman. Roaming Under the Microscope: Trajectory Study of Formaldehyde Dissociation. *The Journal of Physical Chemistry A* **2016**, *120*, 5103–5114.

- [59] G. H. Peslherbe, H. Wang, W. L. Hase, *Monte Carlo Sampling for Classical Trajectory Simulations*, Vol. 105 of *Advances in Chemical Physics*, John Wiley & Sons, Inc., **1999**, pp. 171–201.
- [60] H. Goldstein, *Classical Mechanics*, of Addison-Wesley series in physics, Addison-Wesley Publishing Company, Reading, MA, USA, **1980**.
- [61] M. Abramowitz, I. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, of Applied mathematics series, Dover Publications, **1964**.
- [62] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi, *GNU Scientific Library Reference Manual 3rd ed.*, Network Theory, Ltd., **2009**, <http://www.gnu.org/software/gsl>.
- [63] U. M. Ascher, L. R. Petzold, *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*, SIAM: Society for Industrial and Applied Mathematics, **1998**.
- [64] J. W. Brady, J. D. Doll, D. L. Thompson. Cluster dynamics: A classical trajectory study of $A^*n \rightarrow An-1 + A$. *The Journal of Chemical Physics* **1981**, *74*, 1026–1028.
- [65] M. Araujo, B. Lasorne, M. J. Bearpark, M. A. Robb. The photochemistry of formaldehyde: Internal conversion and molecular dissociation in a single step? *Journal of Physical Chemistry A* **2008**, *112*, 7489–7491.
- [66] C. Cerjan, W. Miller. On finding transition-states. *Journal of Chemical Physics* **1981**, *75*, 2800–2806.
- [67] W. H. Miller, N. C. Handy, J. E. Adams. Reaction path Hamiltonian for polyatomic molecules. *The Journal of Chemical Physics* **1980**, *72*, 99–112.
- [68] J. Nelder, R. Mead. A simplex-method for function minimization. *Computer Journal* **1965**, *7*, 308–313.
- [69] S. van der Walt, S. C. Colbert, G. Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **2011**, *13*, 22–30.

- [70] S. Peyton Jones, et al.. The Haskell 98 Language and Libraries: The Revised Report. *Journal of Functional Programming* **2003**, *13*, 0–255.
- [71] M. Matsumoto, T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation* **1998**, *8*, 3–30.
- [72] O. Tange. GNU Parallel - The Command-Line Power Tool. ;login: The USENIX Magazine **2011**, *36*, 42–47.
- [73] D. V. Cofer-Shabica, *Source code from ‘Potential landscape perspectives on roaming: Insights on formaldehyde from geodesics paths’*, C & Haskell source code, Brown University, **2018**, doi:10.7301/Z0FT8JJ6.
- [74] D. Eberly, *Derivative Approximation by Finite Differences*, Tech. Rep., Geometric Tools, LLC, **2001**, <https://www.geometrictools.com/Documentation/FiniteDifferences.pdf>.
- [75] T. Cover, J. Thomas, *Elements of Information Theory*, Wiley, **2006**.
- [76] H. Theil. The entropy of the maximum entropy distribution. *Economics Letters* **1980**, *5*, 145 – 148.
- [77] D. R. Yarkony. Nuclear dynamics near conical intersections in the adiabatic representation: I. The effects of local topography on interstate transitions. *The Journal of Chemical Physics* **2001**, *114*, 2601–2613.
- [78] D. G. Truhlar, B. C. Garrett. Variational transition-state theory. *Accounts of Chemical Research* **1980**, *13*, 440–448.
- [79] W. L. Hase, *Dynamics of Unimolecular Reactions*, Vol. 2 of *Modern Theoretical Chemistry*, Plenum Press, **1976**, chapter 3.
- [80] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports* **2009**, *478*, 1–69.
- [81] R. D. Levine, R. B. Bernstein, *Thermodynamic Approach to Collision Processes*, Vol. 2 of *Modern Theoretical Chemistry*, Plenum Press, **1976**, chapter 7.

- [82] D. H. Ess, S. E. Wheeler, R. G. Iafe, L. Xu, N. Çelebi-Ölçüm, K. N. Houk. Bifurcations on Potential Energy Surfaces of Organic Reactions. *Angewandte Chemie International Edition* **2008**, *47*, 7592–7601.
- [83] Y. J. Hong, D. J. Tantillo. Biosynthetic consequences of multiple sequential post-transition-state bifurcations. *Nature Chemistry* **2014**, *6*, 104–111.
- [84] F. W. Wiegel, *Introduction to Path-integral Methods in Physics and Polymer Science*, World Scientific Publishing Co Pte Ltd, Singapore, **1986**.
- [85] J. G. Simmonds, *A Brief on tensor analysis*, of *Undergraduate texts in mathematics*, Springer-Verlag, New York, NY, USA, **1982**.
- [86] G. Golub, C. Van Loan, *Matrix Computations*, of *Matrix Computations*, Johns Hopkins University Press, **2012**.
- [87] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**, *57*, 97–109.
- [88] Information Technology Laboratory, *Secure Hash Standard (SHS)*, Tech. Rep. FIPS PUB 180-4, National institute for Standards and Technology, Gaithersburg, MD 20899-8900, **2012**, <http://csrc.nist.gov/publications/fips/fips180-4/fips-180-4.pdf>.

APPENDIX A

DERIVATION OF PATH-INTEGRAL FORM OF THE EQUATION FOR FREE DIFFUSION

We present a derivation and discussion of Wang and Stratt's path-integral expression for the diffusion propagator [8], which suggests geodesics on energy landscapes may be a way of getting at the inherent dynamics of a chemical system. The propagator is given as:

$$G(\underline{R}_0 \rightarrow \underline{R}, t) = \int_{\underline{R}_0, 0}^{\underline{R}, t} \mathcal{D}[\underline{R}(\tau)] \exp \left[-\frac{1}{4D} \int_0^t \left(\frac{d\underline{R}}{d\tau} \right)^2 d\tau \right], \quad (\text{A.1})$$

which specifies the probability of transitioning from state \underline{R}_0 to \underline{R} in time t for a system under free diffusion (no forces) as an integral over the space of all paths, $\underline{R}(\tau)$. D is the diffusion constant and \underline{R} is the configuration-space vector specifying the state of the system. This appendix contains two parts: a derivation of the above expression and then a review of the calculus of variations to derive the unconditionally-minimized, straight-line portion of the geodesics.

A.1 Derivation

This section closely follows the first chapter of Weigel's excellent introduction to path integrals [84]. In the absence of external forces¹, diffusion is a process characterized by the following dynamical relation:

$$\frac{d}{dt}G(\underline{R}, t) = D\nabla^2 G(\underline{R}, t), \quad (\text{A.2})$$

where $G(\underline{R}, t)$ is the probability of finding a system in configuration \underline{R} at time t and D is the diffusion constant, which governs the intrinsic rate of diffusion.

With the initial condition $G(\underline{R}, t = 0) = \delta(\underline{R} - \underline{R}_0)$, one can find the solution for a Cartesian system of α degrees of freedom to be:

$$G(\underline{R}_0 \rightarrow \underline{R}, t) = (4\pi Dt)^{-\alpha/2} \exp\left[-\frac{(\underline{R} - \underline{R}_0)^2}{4Dt}\right], \quad (\text{A.3})$$

as can be verified by substitution.

Suppose we are interested in a particular path² between \underline{R}_0 and \underline{R} . We could represent such a path by discretizing it over $(N + 1)$ bits of time, each of duration ε , such that: $(N + 1)\varepsilon = t$. For notational ease, we define,

$$\begin{aligned} t_i &= \varepsilon \cdot i \\ \underline{R}_i &= \underline{R}(t_i), \end{aligned}$$

giving the discrete time and system configuration along the path. We also have $t_{N+1} = t$. This procedure is illustrated in figure A.1.

To compute the probability of tracing such a path, we rely on the independence of the probabilities of taking each such step. The probability of the path is then the probability of propagating from \underline{R}_0 to \underline{R}_1 in time ε multiplied by the probability of

¹ Studying free diffusion allows us to treat complicated systems from another perspective. We could solve the diffusion equation in the presence of a complicated potential, but this is neither easy nor transferable. Free diffusion is much more straight-forward. However, we still don't get a free lunch: the potential will impose stringent boundary conditions in a yet unspecified manner.

²This may seem an odd question. Why are we creating more work if we have solved already the diffusion equation? The answer lies in the formal expression to which we arrive if we study the solution as a function of *paths*.

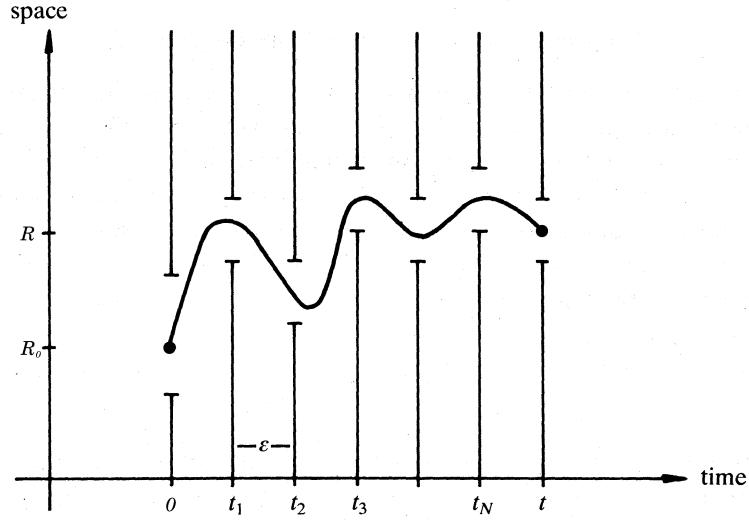


Figure A.1: Discretizing a Path

Note that while this figure traces the time evolution of a single spatial coordinate, paths of arbitrary dimension are amenable to this decomposition. *Figure modified from [84].*

propagating from \underline{R}_1 to \underline{R}_2 in time ε and so on. By inserting eq. A.3 for each step, we have:

$$\begin{aligned} G(\underline{R}_0 \rightarrow \underline{R}, t; \{\underline{R}_1, \underline{R}_2, \dots, \underline{R}_N\}) &= \prod_{i=0}^N G(\underline{R}_i \rightarrow \underline{R}_{i+1}, \varepsilon) \\ &= [(4\pi D\varepsilon)^{-\alpha/2}]^{N+1} \exp \left[-\sum_{i=0}^N \frac{(\underline{R}_{i+1} - \underline{R}_i)^2}{4D\varepsilon} \right], \quad (\text{A.4}) \end{aligned}$$

which gives the probability of diffusing from \underline{R}_0 to \underline{R} in a time t via the N ordered points $\{\underline{R}_1, \underline{R}_2, \dots, \underline{R}_N\}$. We could recover our original expression (eq. A.3) by integrating³ eq. A.4 over the domain of each \underline{R}_j :

$$G(\underline{R}_0 \rightarrow \underline{R}, t) = \int d\underline{R}_1 \int d\underline{R}_2 \dots \int d\underline{R}_N G(\underline{R}_0 \rightarrow \underline{R}, t; \{\underline{R}_1, \underline{R}_2, \dots, \underline{R}_N\}) \quad (\text{A.5})$$

Suppose we increased N until the path appeared smooth; in the continuous limit,

³ Recall the Gaussian integral:

$$\int_{-\infty}^{+\infty} dx e^{-ax^2} = \sqrt{\frac{\pi}{a}}$$

$N \rightarrow \infty$ and $\varepsilon \rightarrow 0$. Focusing for the moment on the exponential, we can write⁴:

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ N \rightarrow \infty}} \exp \left[- \sum_{i=0}^N \frac{(\tilde{R}_{i+1} - \tilde{R}_i)^2}{4D\varepsilon} \right] = \quad (\text{A.6})$$

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ N \rightarrow \infty}} \exp \left[- \frac{1}{4D} \sum_{i=0}^N \left(\frac{\tilde{R}_{i+1} - \tilde{R}_i}{\varepsilon} \right)^2 \varepsilon \right] = \quad (\text{A.7})$$

$$\exp \left[- \frac{1}{4D} \int_0^t \left(\frac{d\tilde{R}}{d\tau} \right)^2 d\tau \right], \quad (\text{A.8})$$

where $\tilde{R}(\tau)$ is a continuous function on $[0, t]$ specifying the path. Now we can write:

$$G(\tilde{R}_0 \rightarrow \tilde{R}, t) = \lim_{\substack{\varepsilon \rightarrow 0 \\ N \rightarrow \infty}} \left[(4\pi D\varepsilon)^{-\alpha/2} \right]^{N+1} \int d\tilde{R}_1 \int d\tilde{R}_2 \dots \int d\tilde{R}_N \exp \left[- \frac{1}{4D} \int_0^t \left(\frac{d\tilde{R}}{d\tau} \right)^2 d\tau \right] \quad (\text{A.9})$$

To express this more compactly, we define the following operator⁵:

$$\int_{\tilde{R}_0, 0}^{\tilde{R}, t} \mathcal{D}[\tilde{R}(\tau)] \equiv \lim_{\substack{\varepsilon \rightarrow 0 \\ N \rightarrow \infty}} (4\pi D\varepsilon)^{-\alpha(N+1)/2} \int d\tilde{R}_1 \int d\tilde{R}_2 \dots \int d\tilde{R}_N, \quad (\text{A.10})$$

which explicitly specifies the boundary values: \tilde{R}_0 at $t = 0$ and \tilde{R} at time t . This is a path integral. With this notation in hand, we arrive at our original equation:

$$G(\tilde{R}_0 \rightarrow \tilde{R}, t) = \int_{\tilde{R}_0, 0}^{\tilde{R}, t} \mathcal{D}[\tilde{R}(\tau)] \exp \left[- \frac{1}{4D} \int_0^t \left(\frac{d\tilde{R}}{d\tau} \right)^2 d\tau \right], \quad (\text{A.11})$$

⁴ Using the definition of a derivative:

$$\frac{df}{dt} = \lim_{\varepsilon \rightarrow 0} \frac{f(t + \varepsilon) - f(t)}{\varepsilon}.$$

And the approximation for an integral:

$$\lim_{N \rightarrow \infty} \sum_{i=0}^N f(a + i\Delta\tau) \Delta\tau \approx \int_a^b d\tau f(\tau),$$

where $\Delta\tau = \frac{b-a}{N}$.

⁵ *Nota bene*, this “operator” is really just notational shorthand for the procedure we followed to arrive here. Take care that any subsequent manipulations respect with the discretization—they must!

which expresses the diffusion propagator as an integral over the space of all possible *paths* connecting the boundary values.

From eq. A.11, we can observe that when diffusion is slow, D is small and therefore the paths that will dominate *must* minimize the integral within the exponential⁶. Those paths will be the *shortest* paths. The key observation is the following: the extremization of any quantity of the form:

$$I = \int d\tau f(\tau) \quad (\text{A.12})$$

can also be effected by extremizing⁷:

$$I' = \int d\tau H[f(\tau)] \quad (\text{A.13})$$

where $H[x]$ is a strictly positive function with non-negative derivative, defined on the range of $f(\tau)$. Thus, minimizing

$$\int_0^t d\tau \left(\frac{d\mathcal{R}}{d\tau} \right)^2 \quad (\text{A.14})$$

can be achieved by taking $H[x] = \sqrt{x}$ and minimizing

$$\int_0^t d\tau \sqrt{\left(\frac{d\mathcal{R}}{d\tau} \right)^2} \quad (\text{A.15})$$

Bringing the differential into the root, eq. A.15 reduces to

$$\int_0^t \sqrt{d\mathcal{R} \cdot d\mathcal{R}} = \int_0^t \|d\mathcal{R}\| \quad (\text{A.16})$$

which is clearly the length of the path.

A.2 Unconditional minimization

In this section, we follow Goldstein [60] to find the form of the (shortest) paths, which unconditionally minimize the action. Again, we will work in Cartesian coordinates for

⁶That is, the dominant paths will obey the classical mechanical principle of least action:

$$\delta S[\mathcal{R}(\tau)] = 0, \quad S[\mathcal{R}(\tau)] = \int d\tau (2T).$$

From here it is a simple extension to systems with unequal masses and/or non-Cartesian bases.

⁷Consider the chain rule with the variational derivative.

an otherwise unconstrained system. See Jacobson and Stratt [50] for work dealing with free rotations and Frechette and Stratt [52] for a case where a constraint is applied to the free steps.

Using calculus to find the extrema of a function, for instance $h(x)$, is a familiar operation. By equating the derivative of $h(x)$ with 0, we locate values of x for which $h(x)$ is unchanging or stationary; these locations often correspond to local extrema of the function. The problem treated by the calculus of variations is analogous. We seek to extremize an object that takes a *function* (in our case, a path) as its argument. That is, we find some path such that the object is stationary with respect to variations in that function.

From eq. A.11, we seek to minimize an integral over kinetic energy, which in general is a function of coordinates and their derivatives. Working in one dimension⁸, we seek to find the path $R(\tau)$ that extremizes:

$$J = \int_0^t d\tau f(R, \dot{R}) \quad (\text{A.17})$$

That is, we seek $R(\tau)$ such that J is unchanged for infinitesimal variations in R . We can represent the paths neighboring the correct (extremal) path, $R(\tau)$ by adding an arbitrary function, $\eta(\tau)$ in proportion to a small variable parameter, α :

$$R(\tau, \alpha) = R(\tau) + \alpha\eta(\tau) \quad (\text{A.18})$$

As we are interested in R such that it connects $R(0)$ to $R(t)$, we require that:

$$\eta(0) = \eta(t) = 0 \quad (\text{A.19})$$

J in eq. A.17 is now a function of α as well:

$$J(\alpha) = \int_0^t d\tau f(R(\tau, \alpha), \dot{R}(\tau, \alpha)) \quad (\text{A.20})$$

We can encode the requirement for stationary J in the following familiar form⁹:

$$0 = \left. \frac{dJ}{d\alpha} \right|_{\alpha=0} \quad (\text{A.21})$$

⁸This procedure extends to systems with an arbitrary number of degrees of freedom.

⁹We require that the derivative be evaluated at $\alpha = 0$ because this corresponds to the unperturbed path we seek.

We can bring the α derivative inside the integral in eq. A.20 and applying the chain rule, we have:

$$\frac{dJ}{d\alpha} = \int_0^t d\tau \left(\frac{\partial f}{\partial R} \frac{\partial R}{\partial \alpha} + \frac{\partial f}{\partial \dot{R}} \frac{\partial \dot{R}}{\partial \alpha} \right)$$

From eq. A.18, we have that $\frac{\partial R}{\partial \alpha} = \eta$ and $\frac{\partial \dot{R}}{\partial \alpha} = \dot{\eta}$, the arbitrary function of τ and its time derivative. This leaves:

$$\frac{dJ}{d\alpha} = \int_0^t d\tau \left(\frac{\partial f}{\partial R} \eta + \frac{\partial f}{\partial \dot{R}} \dot{\eta} \right) \quad (\text{A.22})$$

Consider the second term in the integral:

$$\int_0^t d\tau \frac{\partial f}{\partial \dot{R}} \dot{\eta}$$

This expression is susceptible by integration by parts¹⁰ Re-arranging leaves us with:

$$\int_0^t d\tau \frac{\partial f}{\partial \dot{R}} \dot{\eta} = \frac{\partial f}{\partial \dot{R}} \eta \Big|_0^t - \int_0^t d\tau \frac{d}{d\tau} \left(\frac{\partial f}{\partial \dot{R}} \right) \eta$$

Because of the bounds on η , $\eta(0) = \eta(t) = 0$, the surface terms vanish! Inserting back into eq. A.22 and factoring out η gives:

$$\frac{dJ}{d\alpha} \Big|_{\alpha=0} = \int_0^t d\tau \left[\frac{\partial f}{\partial R} - \frac{d}{d\tau} \left(\frac{\partial f}{\partial \dot{R}} \right) \right] \eta = 0 \quad (\text{A.23})$$

where evaluating the derivative at $\alpha = 0$ gives R and \dot{R} as functions of τ alone. Since η is an arbitrary function, the integral is stationary *only if*

$$\frac{\partial f}{\partial R} - \frac{d}{d\tau} \frac{\partial f}{\partial \dot{R}} = 0 \quad (\text{A.24})$$

which gives the conditions on R to minimize J ¹¹.

¹⁰ Recall the general form,

$$\int_a^b u dv = uv \Big|_a^b - \int_a^b v du$$

And identify:

$$\begin{aligned} u &= \frac{\partial f}{\partial \dot{R}} & du &= \frac{d}{d\tau} \left(\frac{\partial f}{\partial \dot{R}} \right) d\tau \\ dv &= \dot{\eta} d\tau & v &= \eta \end{aligned}$$

¹¹ Notice that inserting the Lagrangian for f gives $S = J$, the action, and yields Lagrange's equations of motion.

Minimizing The Force-Free Action

From eq. A.11, we have $f = \left(\frac{d\tilde{R}}{d\tau}\right)^2$ in eq. A.24. Proceeding, we find that the path which minimizes eq. A.11 is specified by the differential equation:

$$\ddot{\tilde{R}} = 0 \quad (\text{A.25})$$

which, when integrated and solved for its boundary values, yields the equation for a straight line (and the geodesic!):

$$\tilde{R}(\tau) = \tilde{R}(0) \left(1 - \frac{\tau}{t}\right) + \tilde{R}(t) \left(\frac{\tau}{t}\right) \quad (\text{A.26})$$

this amounts to linearly interpolating between $\tilde{R}(0)$ and $\tilde{R}(t)$ in time t for each of the $3N$ coordinates. To actually compute such paths we discretize¹² eq. A.26 as follows:

$$\tilde{R}^{i+1} = \tilde{R}^i + \delta_s \frac{\tilde{R}_f - \tilde{R}^i}{|\tilde{R}_f - \tilde{R}^i|} \quad (\text{A.27})$$

Where $\tilde{R}_f = \tilde{R}(t)$ and δ_s is a small step-size. *Small* here is relative to features of the landscape. In practice, δ_s should be as large as it can be such that the lengths of computed geodesics are still converged to a maximum; see discussion in Section 2.2.2.

¹²Consider $\tilde{R}(\tau + \Delta) - \tilde{R}(\tau)$.

APPENDIX B

TWO MORE WAYS OF PRESERVING THE CENTER OF MASS DURING ESCAPE STEPS

Vanilla Newton-Raphson

When the procedure for free steps described in Section 2.2 (eq. 2.9) moves the system such that $V(\underline{R}(\tau)) > E_L$, we need a way back to the boundary. That is, we seek the nearest solution to $V(\underline{R}(\tau)) = E_L$. A classic strategy for solving this problem is a Newton-Raphson root search.

Here, we expand the function of interest to first order and run downhill:

$$V(\underline{R}) \approx V(\underline{R}_0) + \nabla V \Big|_{\underline{R}_0} \cdot (\underline{R} - \underline{R}_0) = E_L \quad (\text{B.1})$$

Where \underline{R}_0 was the first offending step in eq. A.27. We make the guess that heading downhill via the steepest path will be optimal; this suggests a solution of the form:

$$(\underline{R} - \underline{R}_0) = C \nabla V \Big|_{\underline{R}_0} \equiv \Delta \underline{R} \quad (\text{B.2})$$

Which immediately yields the constant¹:

$$C = \frac{E_L - V(\underline{R}_0)}{\left| \nabla V|_{\underline{R}_0} \right|^2}$$

This procedure is then iterated until the $V(\underline{R}) = E_L$ boundary is found. It may look like eq. B.2 renders our problem solved, but as we prove in Section 2.2.1, this naïve root search fails to preserve the center of mass and spuriously lengthens geodesics computed with it. In this appendix, we present two wildly different methods of coping with the problem. The first is the essence of simplicity, though limited in its generality. By comparison, the second technique is a swatting a fly with a laser cannon, but demonstrates a useful and general principle.

B.1 Swatting a fly without moving: Follow the inverse-mass-weighted gradient

Never let it be said that a little physical intuition won't go a long way. It turns out that the techniques discussed in Section 2.2.1 and to be discussed in Section B.2 solve a much harder problem than is necessary—they will both project our center-of-mass perturbing motion from *any* configuration-space displacement. They are included because of their generality and because the result in Section 2.2.1 is what we actually implemented in our investigations of geodesics on formaldehyde.

Unfortunately physical insight doesn't always occur prior to pages of vector algebra or hundreds of thousands of core-hours of computer time². However, the result may still be useful in the future and so we provide a pedagogical exposition here; if we're not careful, the lead-up will take longer than the proof.

Our task is to follow a gradient downhill without perturbing the center of mass. That sounds awfully similar to the way that classical mechanics works; surely the general purpose projection operators are doing more than is required. If we re-write Newton's

¹The solution should be familiar from Section 2.2, eq. 2.11, from which we have $\zeta = -C$.

²which is not to say that we could have avoided all those hours if the insight had struck earlier. See the notes on performance at the end of Appendix E on page 164; this section does not address the bottle-neck.

second law for Cartesian coordinates in terms of the influence of a potential [60], we have:

$$\underline{\underline{F}} = \mathbf{M} \cdot \underline{\underline{a}} \quad (\text{B.3})$$

$$= -\nabla \underline{\underline{V}} \quad (\text{B.4})$$

Where $\underline{\underline{F}}$ is the force experienced by the system due to the potential V , $\underline{\underline{a}}$ is the system's acceleration and \mathbf{M} is the diagonal $N \times N$ mass matrix defined in eq. 2.69:

$$(\mathbf{M})_{i\mu,j\nu} = \delta_{i,j}\delta_{\mu,\nu}m_i \quad i,j = 1 \dots N \quad ; \quad \mu,\nu = x,y,z \quad (\text{B.5})$$

In the absence of external fields, classical mechanics *never* changes the center of mass³. Suppose a dynamical system is momentarily at rest; in what direction does it first move? Instantaneously, it moves in the direction of its *acceleration* (later this becomes complicated by the fact that it is already moving). That acceleration is:

$$\underline{\underline{a}} = -\mathbf{M}^{-1} \nabla \underline{\underline{V}} \quad (\text{B.6})$$

where \mathbf{M}^{-1} is simply the diagonal inverse-mass matrix given by:

$$(\mathbf{M}^{-1})_{i\mu,j\nu} = \delta_{i,j}\delta_{\mu,\nu} \frac{1}{m_i} \quad (\text{B.7})$$

Equation B.6 suggests that we can escape center-of-mass perturbation simply following the inverse-mass-weighted gradient.

Could this really be? In Section 2.2.1, we arrived at eq. 2.35 right before showing that the “vanilla” Newton-Raphson root search displaced the center of mass:

$$\vec{0} \stackrel{?}{=} \vec{\mathbf{M}} \cdot \nabla \underline{\underline{V}}|_{\mathbb{R}^n} \quad (\text{B.8})$$

where $\vec{\mathbf{M}}$ is the $3 \times 3N$ dimensional, block-diagonal matrix composed of N ordered, 3×3 blocks with m_i/M_{total} along the diagonals.

$$(\vec{\mathbf{M}})_i = \frac{m_i}{M_{\text{total}}} \overset{\leftrightarrow}{1} \quad , \quad i = 1 \dots N \quad (\text{B.9})$$

³This is a restatement of Newton’s first law.

If, instead of following the gradient, we followed the inverse-mass-weighted gradient suggested by eq. B.6, we would require that:

$$\vec{0} = \vec{M} \cdot \mathbf{M}^{-1} \cdot \nabla V \Big|_{\tilde{\mathcal{R}}^n} \quad (\text{B.10})$$

$$= \vec{1} \cdot \nabla V \Big|_{\tilde{\mathcal{R}}^n} \quad (\text{B.11})$$

where $\vec{1}$ is the $3 \times 3N$ dimensional, block-diagonal matrix composed of N ordered, 3×3 identity blocks⁴.

$$\left(\vec{1}\right)_i = \overset{\leftrightarrow}{1} \quad , \quad i = 1 \dots N \quad (\text{B.12})$$

Equation B.11 is a sum along each coordinate, $\{x, y, z\}$, and can be rewritten as:

$$0 \stackrel{?}{=} \sum_i^N \left(\nabla V \Big|_{\tilde{\mathcal{R}}^n} \right)_\mu ; \quad \mu = x, y, z \quad (\text{B.13})$$

Which can be cast in a more familiar form as the requirement that the sum of forces in an isolated system vanish:

$$0 = \sum_i^N = \vec{F}_i \quad (\text{B.14})$$

This is proven in many classical mechanics texts [60] and follows from $\vec{F}_{12} = -\vec{F}_{21}$. Namely that the force exerted on one object by another is equal and opposite to the force exerted by the second on the first—Newton’s third law. A Newton-Raphson root search based around eq. B.6 will therefore preserve the center of mass of the system.

Solving eq. B.1 with the suggested guess, $\Delta \tilde{\mathcal{R}} = C \cdot \mathbf{M}^{-1} \cdot \nabla V \Big|_{\tilde{\mathcal{R}}_0}$ yields:

$$C = \frac{E_L - V(\tilde{\mathcal{R}})}{\nabla V \Big|_{\tilde{\mathcal{R}}_0} \cdot \mathbf{M}^{-1} \cdot \nabla V \Big|_{\tilde{\mathcal{R}}_0}} \quad (\text{B.15})$$

And so our center-of-mass respecting analogue of the steepest-descent Newton-Raphson search takes steps like:

$$\Delta \tilde{\mathcal{R}} = \frac{E_L - V(\tilde{\mathcal{R}})}{\nabla V \Big|_{\tilde{\mathcal{R}}_0} \cdot \mathbf{M}^{-1} \cdot \nabla V \Big|_{\tilde{\mathcal{R}}_0}} \cdot \mathbf{M}^{-1} \cdot \nabla V \Big|_{\tilde{\mathcal{R}}_0} \quad (\text{B.16})$$

⁴We saw this matrix earlier in eq. 2.43 under the name \vec{C}^\top .

B.2 Laser-cannon: A general projection

Suppose we modify eq. B.1 to respect a constraint? After all eq. B.2 was based on a guess about following the gradient. If, instead, we chose to follow a path constrained to be *parallel* to some other direction, \hat{n} , eq. B.2 would have the following form:

$$(\underline{R} - \underline{R}_0) = C\hat{n} \quad (\text{B.17})$$

and inserting into eq. B.1 would yield:

$$V(\underline{\underline{R}}) \approx V(\underline{\underline{R}}_0) + C \nabla V \Big|_{\underline{\underline{R}}_0} \cdot \hat{n} = E_L \quad (\text{B.18})$$

Solving for C yields:

$$C = \frac{E_L - V(\underline{\underline{R}}_0)}{\nabla V \Big|_{\underline{\underline{R}}_0} \cdot \hat{n}}$$

While this is well an good, it doesn't quite solve our problem. We need to be able to impose a constraint that is *orthogonal* to some directions, those of center of mass motion. What if we follow the path of steepest descent subject to the constraint that it is orthogonal to some configuration-space vector which would move the center of mass, $\hat{\xi}$?

To construct such a step we take a brief detour to discuss dyads. The following treatment uses the notation of real-space vectors, but is applicable to vectors of arbitrary dimension (such as our configuration space vectors). This treatment is reprised and expanded in [85].

A dyad (or dyadic) is a 2nd order tensor formed from the direct product of 2 vectors. The dyadic $\vec{u}\vec{v}$ is defined such that the action of this tensor on a vector \vec{w} is:

$$\vec{u}\vec{v}(\vec{w}) = \vec{u}(\vec{v} \cdot \vec{w}) \quad (\text{B.19})$$

The Cartesian components are $(\vec{u}\vec{v})_{\alpha,\gamma} = u_\alpha v_\gamma$ or

$$\vec{u}\vec{v} = \begin{pmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \end{pmatrix}$$

Given a unit vector, \hat{u} , the definition in eq. B.19, makes it easy to construct an operator which projects another vector onto \hat{u} . For example:

$$\hat{u}\hat{u}(\vec{w}) = \hat{u}(\hat{u} \cdot \vec{w}) = \text{Proj}_{\hat{u}}\vec{w} \quad (\text{B.20})$$

gives the projection of \vec{w} on \hat{u} . See Figure B.1 for a geometric interpretation.

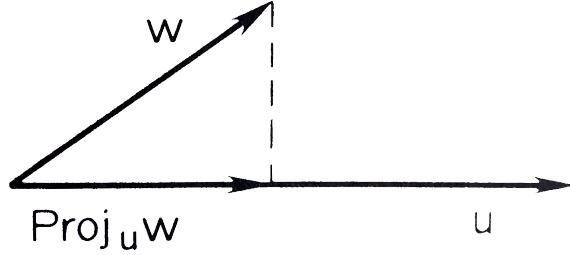


Figure B.1: A projection

Geometric realization of the projection of \vec{w} onto \hat{u} . Modified from [85].

A useful property of projection operators is their *idempotency*. From the geometric significance, it must follow that repeat application of any projection, is equivalent to single application. We show that this holds for $\hat{u}\hat{u}$ as follows:

$$(\hat{u}\hat{u})^2 = \hat{u}\hat{u} \hat{u}\hat{u} \quad (\text{B.21})$$

$$= \hat{u}(\hat{u} \cdot \hat{u})\hat{u} \quad (\text{B.22})$$

$$= \hat{u}(1)\hat{u} \quad (\text{B.23})$$

$$= \hat{u}\hat{u} \quad (\text{B.24})$$

By induction, we can reason that $(\hat{u}\hat{u})^n = \hat{u}\hat{u}$ for all $n \geq 1$.

In addition to projection on \hat{u} , we can construct the projection off of (or orthogonal to) \hat{u} . Such a projection is given by:

$$(1 - \hat{u}\hat{u}) \quad (\text{B.25})$$

The orthogonality of the two projections is easily shown:

$$(1 - \hat{u}\hat{u})\hat{u}\hat{u} \quad (\text{B.26})$$

$$= \hat{u}\hat{u} - (\hat{u}\hat{u})^2 \quad (\text{B.27})$$

$$= \hat{u}\hat{u} - \hat{u}\hat{u} \quad (\text{B.28})$$

$$= 0 \quad (\text{B.29})$$

Armed with such a projection operator, the analogy to eq. B.17 for an *orthogonal* downhill direction is clear. We now take a step down and project out all motion along $\hat{\xi}$:

$$(\underline{R} - \underline{R}_0) = C \left(\mathbb{1} - \hat{\xi} \hat{\xi} \right) \cdot \nabla V|_{\underline{R}_0} \quad (\text{B.30})$$

This, we again insert into eq. B.1, yielding:

$$V(\underline{R}) \approx V(\underline{R}_0) + C \nabla V|_{\underline{R}_0} \cdot \left(\mathbb{1} - \hat{\xi} \hat{\xi} \right) \cdot \nabla V|_{\underline{R}_0} = E_L \quad (\text{B.31})$$

Solving for C yields:

$$C = \frac{E_L - V(\underline{R}_0)}{\nabla V|_{\underline{R}_0} \cdot \left(\mathbb{1} - \hat{\xi} \hat{\xi} \right) \cdot \nabla V|_{\underline{R}_0}}$$

To enforce a stationary center of mass, we can introduce a coordinate transformation to separate center of mass motion from internal coordinates:

$$\vec{R}_{\text{c.m.}} \equiv \frac{\sum_{j=1}^N m_j \vec{R}_j}{\sum_{j=1}^N m_j} \equiv \frac{1}{M} \sum_{j=1}^N m_j \vec{r}_j \quad (\text{B.32})$$

where m_j is the mass of the j th particle. The internal coordinates are then:

$$\vec{r}_j \equiv \vec{R}_j - \vec{R}_{\text{c.m.}} \quad (\text{B.33})$$

Note that our system of $3N$ degrees of freedom is now specified with $3N + 3$ coordinates, which have the following 3 constraints:

$$\sum_{j=1}^N m_j \vec{R}_j = 0 \quad (\text{B.34})$$

one for each spatial dimension. We encode these constraints as vector operations in configuration space as:

$$\left(\xi_\mu \right)_{j\nu} \equiv \delta_{\mu,\nu} m_j \quad (\text{B.35})$$

$$\implies \xi_\mu \cdot \underline{R} = 0 \quad \forall \mu \in \{x, y, z\} \quad (\text{B.36})$$

What does ξ_μ look like? By way of example:

$$\xi_x = \begin{pmatrix} m_1 \\ 0 \\ 0 \\ m_2 \\ 0 \\ 0 \\ \vdots \\ m_N \\ 0 \\ 0 \end{pmatrix}$$

The magnitude of ξ_μ is the root of the sum of the squares of the masses:

$$|\xi_\mu| = \sqrt{\sum_{j=1}^N m_j^2} \quad (\text{B.37})$$

and so the unit vector pointing along ξ_μ is:

$$(\hat{\xi}_\mu)_{j\nu} = \frac{\delta_{\mu,\nu} m_j}{\sqrt{\sum_{j=1}^N m_j^2}} \quad (\text{B.38})$$

Constructing the corresponding dyad, we have:

$$(\hat{\xi}_\mu \hat{\xi}_\mu)_{j\gamma, k\alpha} = \frac{m_j m_k}{\sum_{j=1}^N m_j^2} \delta_{\mu,\gamma} \delta_{\mu,\alpha} \quad (\text{B.39})$$

and the projection operator which eliminates center of mass motion along the μ direction is then:

$$(\mathbb{1} - \hat{\xi}_\mu \hat{\xi}_\mu)_{j\gamma, k\alpha} = \delta_{\gamma,\alpha} \delta_{j,k} - \delta_{\mu,\gamma} \delta_{\mu,\alpha} \frac{m_j m_k}{\sum_{j=1}^N m_j^2} \quad (\text{B.40})$$

$$= \delta_{\gamma,\alpha} \left(\delta_{j,k} - \delta_{\mu,\gamma} \delta_{\mu,\alpha} \frac{m_j m_k}{\sum_{j=1}^N m_j^2} \right) \quad (\text{B.41})$$

Where we have made use of:

$$(\mathbb{1})_{j\gamma, k\alpha} = \delta_{\gamma, \alpha} \delta_{j, k}$$

and note that in eq. B.40

$$\delta_{\gamma, \alpha} \delta_{\mu, \gamma} \delta_{\mu, \alpha} = \delta_{\mu, \gamma} \delta_{\mu, \alpha}$$

as the term is finite only for $\alpha = \gamma = \mu$ with μ in the range of α and γ .

To eliminate center of mass motion along each Cartesian direction, we form the product of all of them:

$$(\mathbb{1} - \mathbb{P}) \equiv \prod_{\mu} (\mathbb{1} - \hat{\xi}_{\mu} \hat{\xi}_{\mu}) \quad (\text{B.42})$$

This product reduces to a sum using the following observation:

$$(\hat{\xi}_{\mu} \hat{\xi}_{\mu}) (\hat{\xi}_{\nu} \hat{\xi}_{\nu}) = (\hat{\xi}_{\mu} \hat{\xi}_{\mu}) \delta_{\mu, \nu}$$

which is true because $\hat{\xi}_{\mu}$ and $\hat{\xi}_{\nu}$ are orthogonal. This leaves:

$$(\mathbb{1} - \mathbb{P}) = \left(\mathbb{1} - \sum_{\mu} \hat{\xi}_{\mu} \hat{\xi}_{\mu} \right) \quad (\text{B.43})$$

Inserting eq. B.39 and using the same identities we did for eq. B.40, we arrive at:

$$(\mathbb{1} - \mathbb{P})_{j\gamma, k\alpha} = \delta_{\gamma, \alpha} \left(\delta_{j, k} - \frac{m_j m_k}{\sum_{j=1}^N m_j^2} \right) \quad (\text{B.44})$$

With $\mathfrak{M}_2 \equiv \sum_{j=1}^N m_j^2$ the projection defined by eq. B.44 has the following symmetric structure in 3 spatial dimensions:

$$(\mathbb{1} - \mathbb{P}) = \begin{pmatrix} 1 - \frac{m_1^2}{\mathfrak{M}_2} & -\frac{m_1 m_2}{\mathfrak{M}_2} & & & \\ -\frac{m_1 m_2}{\mathfrak{M}_2} & 1 - \frac{m_1^2}{\mathfrak{M}_2} & -\frac{m_1 m_2}{\mathfrak{M}_2} & & \dots \\ & -\frac{m_1 m_2}{\mathfrak{M}_2} & 1 - \frac{m_2^2}{\mathfrak{M}_2} & -\frac{m_1 m_2}{\mathfrak{M}_2} & \\ & & -\frac{m_1 m_2}{\mathfrak{M}_2} & 1 - \frac{m_2^2}{\mathfrak{M}_2} & \\ \vdots & & & & \ddots \end{pmatrix} \quad (\text{B.45})$$

Returning to the Newton-Raphson problem, we proceed in analogy to eq. B.30 and posit that the appropriate solution is to follow the gradient while projecting out center of mass motion:

$$(\underline{R} - \underline{R}_0) = C (\mathbb{1} - \mathbb{P}) \cdot \nabla V|_{\underline{R}_0} \quad (\text{B.46})$$

This, we again insert into eq. B.1, yielding:

$$V(\underline{R}) \approx V(\underline{R}_0) + C \nabla V|_{\underline{R}_0} \cdot (\mathbb{1} - \mathbb{P}) \cdot \nabla V|_{\underline{R}_0} = E_L \quad (\text{B.47})$$

Solving for C yields:

$$C = \frac{E_L - V(\underline{R}_0)}{\nabla V|_{\underline{R}_0} \cdot (\mathbb{1} - \mathbb{P}) \cdot \nabla V|_{\underline{R}_0}}$$

Which we insert back into eq. B.46 to find:

$$(\underline{R} - \underline{R}_0) = -\frac{V(\underline{R}_0) - E_L}{|\nabla' V|_{\underline{R}_0}} \nabla' V|_{\underline{R}_0} \quad (\text{B.48})$$

where we have defined $\nabla' \equiv (\mathbb{1} - \mathbb{P}) \cdot \nabla$ as the center-of-mass invariant gradient operator and made use of the idempotent property of projections.

Applying eq. B.48 iteratively leads to the relation (compare to eq. 2.50):

$$\underline{R}_{n+1} = \underline{R}_n - \frac{V(\underline{R}_n) - E_L}{|\nabla' V|_{\underline{R}_n}} \nabla' V|_{\underline{R}_n} \quad (\text{B.49})$$

where \underline{R}_0 is the first $\underline{R}^{(t)}$ with $V(\underline{R}^{(t)}) > E_L$ from eq. A.27. This procedure will escape the forbidden region while preserving the center of mass.

APPENDIX C

MICROCANONICAL SAMPLING FOR THE ANHARMONIC OSCILLATOR

This section follows and is expanded from the work of Peslherbe, Wang, and Hase [59].

C.1 Normal Mode Hamiltonian

In this section we give a description of the normal mode Hamiltonian and demonstrate how to microcanonically sample its degrees of freedom.

For a system of n (separable) harmonic oscillators, the Hamiltonian of the system is given as:

$$H = \sum_{i=1}^n E_i = \frac{1}{2} \sum_{i=1}^n P_i^2 + \omega_i^2 Q_i^2 \quad (\text{C.1})$$

where the ω_i are the frequencies of the oscillators and the normal coordinates, $\{Q_i, P_i\}$, are related to the Cartesian coordinates and momenta, x_i and p_i , by the transformations:

$$\begin{aligned} P_i &= \frac{p_i}{\sqrt{m_i}} \\ Q_i &= \sqrt{m_i}(x_i - x_{0i}) \end{aligned} \quad (\text{C.2})$$

where \vec{x}_0 is the equilibrium position of the oscillator.

For such a system, the distribution function to microcanonically sample phase space is an exactly obtainable object. We show its development in the next section.

Assignment of Energies to Normal Modes

Equivalent to choosing N values for Q_i and P_i , we can assign to the N modes, N energies and N phases. In this section we treat the energies. Supposing that we wish to inject a total microcanonical energy E_0 into the N normal modes, the probability that mode 1 has energy E_1 is proportional to the number of ways the energy $(E_0 - E_1)$ can be distributed to the remaining $(N - 1)$ modes. This is given by the density of states for the oscillators. We then have:

$$P(E_1) = \frac{(E_0 - E_1)^{N-1-1}}{\int_0^{E_0} (E_0 - E_1)^{N-1-1} dE_1} \quad (\text{C.3})$$

We can extrapolate to the general expression by noticing that the number of states is proportional to the remaining energy raised to the power of one less than the number of remaining oscillators. With the denominator ensuring normalization we have:

$$P(E_i) = \frac{(E_0 - E_i - \sum_{j=1}^{i-1} E_j)^{N-1-i}}{\int_0^{E_0 - \sum_{j=1}^{i-1} E_j} (E_0 - E_i - \sum_{j=1}^{i-1} E_j)^{N-1-i} dE_i} \quad (\text{C.4})$$

Sampling equation C.4 directly would be a quite difficult. However the cumulative distribution function allows sampling an arbitrary probability distribution as follows:

$$C(x) = \int_{-\infty}^x P(x') dx' = R_i \quad (\text{C.5})$$

Where $R_i \in [0, 1]$ is a uniformly distributed random variable. For this energy distribution specifically:

$$C(E_i) = \int_0^{E_i} P(E'_i) dE'_i = R_i \quad (\text{C.6})$$

Integrating eqn. C.4 and solving for E_i yields:

$$E_i = \left(E_0 - \sum_{j=1}^{i-1} E_j \right) \left(1 - R_i^{1/(N-i)} \right) \quad (\text{C.7})$$

Equation C.7 can be used do assign the first $(N - 1)$ modes. The N th mode is assigned $E_N = E_0 - \sum_{j=1}^{N-1} E_j$ with probability 1.

Assignment of Random Phases to Normal Modes

Having assigned the energies of the normal modes, we assign to each normal mode, a random phase. We uniformly sample the angle, by choosing a new random variable, $R_i \in [0, 1)$, for each mode and taking:

$$Q_i^0 = +\frac{\sqrt{2E_i}}{\omega_i} \cos(2\pi R_i) \quad (C.8)$$

$$P_i^0 = -\sqrt{2E_i} \sin(2\pi R_i) \quad (C.9)$$

Where Q_i^0 and P_i^0 are the initial trial coordinate and momentum for the i th normal mode. In this way we have specified the $2N$ coordinates of phase space with $2N$ random variables.

We can verify that this yields the target energy by simple substitution into the normal-mode Hamiltonian, eqn. C.1:

$$H = \frac{1}{2} \sum_{i=1}^n P_i^2 + \omega_i^2 Q_i^2 \quad (C.10)$$

$$= \frac{1}{2} \sum_{i=1}^n \left[-\sqrt{2E_i} \sin(2\pi R_i) \right]^2 + \omega_i^2 \left[\frac{\sqrt{2E_i}}{\omega_i} \cos(2\pi R_i) \right]^2 \quad (C.11)$$

$$= \frac{1}{2} \sum_{i=1}^n 2E_i \left[\sin^2(2\pi R_i) + \cos^2(2\pi R_i) \right] \quad (C.12)$$

$$= \sum_{i=1}^n E_i \quad (C.13)$$

$$= H \quad (C.14)$$

C.2 Construction of Normal Mode Coordinates

In this section we construct a transformation to normal coordinates for an arbitrary potential. We proceed by approximating the Hamiltonian in the vicinity of a potential minimum as a collection of harmonic oscillators. Then we diagonalize the approximated Hamiltonian, finding the normal modes and their frequencies. Though the normal modes will be an approximation for an anharmonic potential, the coordinate transformation is exact.

In the neighborhood of a potential minimum¹, \vec{x}_0 , the Hamiltonian for an arbitrary potential with N degrees of freedom can be brought into the form of the normal mode Hamiltonian, eqn. C.1. We make a Taylor series expansion of the potential and truncate all terms of greater than quadratic order:

$$V(\vec{x}) \approx V|_{\vec{x}_0} + \frac{1}{2} \sum_{i \neq j}^N \left. \frac{\partial^2 V}{\partial x_i \partial x_j} \right|_{\vec{x}_0} (x_i - x_{0i})(x_j - x_{0j}) \quad (\text{C.15})$$

Here we use the fact that the potential has a minimum at \vec{x}_0 to guarantee $\left. \frac{\partial V}{\partial x_i} \right|_{\vec{x}_0} = 0 \forall i$, which eliminates the first order terms.

Suitable shifting of the zero of the potential eliminates the zeroth order (constant) term and application of equation C.2 transforms to mass weighted coordinates. To eliminate the cross terms, we diagonalize the hessian matrix, \mathbf{H} , given by:

$$H_{ij} = \left. \frac{\partial^2 V}{\partial Q_i \partial Q_j} \right|_0 = \left[\frac{\partial^2 V}{\partial x_i \partial x_j} \frac{\partial x_i}{\partial Q_i} \frac{\partial x_j}{\partial Q_j} \right]_{\vec{x}_0} = \left. \frac{\partial^2 V}{\partial x_i \partial x_j} \right|_{\vec{x}_0} (m_i m_j)^{-1/2} \quad (\text{C.16})$$

We discuss the evaluation of such an object in Section 2.7.3.

Diagonalization [86, Chapter 8.2] yields a matrix, $\Omega_{ij} = \delta_{ij}\omega_i^2$, containing the frequencies of the normal modes and the eigenvectors (modes) associated with them. The matrix \mathbf{L} , whose columns are the eigenvectors, allows transformation from normal coordinates back to Cartesian coordinates as in equation C.18.

$$\mathbf{L} \equiv [\vec{e}_1, \vec{e}_2, \vec{e}_3, \dots] \quad (\text{C.17})$$

where \vec{e}_i is the column eigenvector associated with the frequency ω_i . In general (for systems with $N > 2$) there will be 6 normal modes with frequencies approaching 0 corresponding to net translation in three dimensions and net rotation about three axes (see also Section 2.7.3).

With these new coordinates and frequencies in hand, we follow the procedure of section C.1 in the assignment of energies and phases.

¹More generally, this procedure can be applied anywhere $\nabla V = 0$, whether a minimum, a maximum, or a saddle point.

Transformation Back to Cartesian Coordinates

Once the normal mode phases and energies have been assigned (see sections C.1 and C.1), it is a simple matter to transform back to Cartesian coordinates.

$$\vec{x} = \vec{x}_0 + \mathbf{M}^{\frac{1}{2}} \mathbf{L} \vec{Q} \quad (\text{C.18})$$

$$\vec{p} = \mathbf{M}^{\frac{1}{2}} \mathbf{L} \vec{P} \quad (\text{C.19})$$

where $M_{ij}^{\frac{1}{2}} = \delta_{ij} m_i^{\frac{1}{2}}$ is the matrix giving the masses associated with each coordinate and \vec{Q} and \vec{P} are vectors giving the previously assigned normal coordinates.

C.3 Corrections for Anharmonicity and Selection of Total Angular Momentum

The normal modes only represent an approximation for our true Hamiltonian and therefore there arises a spurious angular momentum and a deviation from the target energy, E_0 . To correct for this, we iteratively apply the following transformation until convergence² is reached.

1. Form the spurious angular momentum (this is simply the total angular momentum of the molecule),

$$\vec{j}_s = \sum_{i=1}^N \vec{r}_i \times m_i \vec{v}_i \quad (\text{C.20})$$

Where the sum is over atomic centers and \vec{r}_i , \vec{v}_i , and m_i give the Cartesian position, velocity, and mass of the i th atom. To correct for this term we add $\vec{j} = -\vec{j}_s$ to the molecule³ by computing the vector angular velocity:

$$\boldsymbol{\omega} = \mathbf{I}^{-1} \vec{j} \quad (\text{C.21})$$

where \mathbf{I}^{-1} is the inverse of the moment of inertia tensor and adding this angular velocity to each center.

$$\vec{v}'_i = \vec{v}_i + \boldsymbol{\omega} \times \vec{r}_i \quad (\text{C.22})$$

²We place the following constraints on total energy, angular momentum, and center of mass translation: $|\frac{E-E_0}{E_0}| < \delta E$ where $E = H(\vec{x}, \vec{p})$, $|\vec{j}| < \delta j$, and $|\vec{r}_{cm}| < \delta r_{cm}$; all tolerances are of the order of 10^{-8} .

³This procedure could be extended to select arbitrary angular momentum by adding the term $\vec{j} = \vec{j}_0 - \vec{j}_s$ where \vec{j}_0 is the desired angular momentum.

2. To correct the total energy, let $E = H(\vec{x}, \vec{p})$ and scale the coordinates and momenta (velocities would be fine too) as:

$$\vec{x}' = \vec{x}_0 + (\vec{x} - \vec{x}_0) \left(\frac{E_0}{E} \right)^{(1/2)} \quad (C.23)$$

$$\vec{p}' = \vec{p} \left(\frac{E_0}{E} \right)^{(1/2)} \quad (C.24)$$

I have observed that this procedure can yield a fixed point (an infinitely repeated sequence) for certain input parameters.

Explanation of Anharmonic Correction

The algorithm presented in the previous section may appear opaque, but this need not be the case.

Step 1 simply treats the molecule as a rigid body. Equation C.20 is the expression for the total angular momentum and equation C.21 is an inversion of another standard expression for the angular momentum of a rigid body

$$\vec{j} = \mathbf{I}\omega \quad (C.25)$$

Where \mathbf{I} is the moment of inertia tensor [60].

As for step 2, it is exact to the extent that the Hamiltonian is harmonic. Consider a system with a single harmonic degree of freedom (eqn. C.1 for $N = 1$) and target total energy E_0 . For convenience, we choose a shifted coordinate system such that the minimum of the harmonic oscillator is at the origin. This has the effect of setting $\vec{x}_0 = \vec{0}$ in eqn. C.23.

Our present coordinates, Q, P yield an energy E :

$$H(Q, P) = E \neq E_0 \quad (C.26)$$

and we desire that the transformed coordinates, Q', P' yield the energy E_0 :

$$H(Q', P') \stackrel{?}{=} E_0 \quad (C.27)$$

Proceeding by substitution of the transformed coordinates into the Hamiltonian, we have:

$$H(Q', P') = \frac{1}{2} (P'^2 + \omega^2 Q'^2) \quad (\text{C.28})$$

$$= \frac{1}{2} \left[P^2 \left(\frac{E_0}{E} \right) + \omega^2 Q^2 \left(\frac{E_0}{E} \right) \right] \quad (\text{C.29})$$

$$= \left(\frac{E_0}{E} \right) \cdot \frac{1}{2} [P^2 + \omega^2 Q^2] \quad (\text{C.30})$$

$$= \left(\frac{E_0}{E} \right) \cdot E \quad (\text{C.31})$$

$$= E_0 \quad (\text{C.32})$$

Which is as intended.

APPENDIX D

MONTE CARLO SAMPLING THE LANDSCAPE ENSEMBLE

The first part of this appendix closely follows [64], whose methods we imitate. We seek to generate a set of points which samples the roaming region (see Section 2.3.3) of the formaldehyde potential energy surface in the landscape ensemble (Section 2.1.2). Such sets at the appropriate landscape energies are the buckets from which the intermediate, roaming points, are drawn when constructing roaming geodesics (Section 2.3.4).

Metropolis-Hastings sampling [87] is a Monte Carlo procedure for evaluating ensemble averages of the form:

$$\langle f \rangle = \frac{\int d\mathbf{x}^N P(\underline{x}) f(\underline{x})}{\int d\mathbf{x}^N P(\underline{x})} \quad (\text{D.1})$$

where f is a property of the N -dimensional space determined by \underline{x} and $P(\underline{x})$ is the probability of a given state, \underline{x} . The Metropolis algorithm is of particular use when $P(\underline{x})$ is difficult to sample directly or to normalize. Its output is a set of states, $\{\underline{x}_i\}$, such that the average in eq. D.1 can be replaced by:

$$\langle f \rangle = \frac{1}{N} \sum_i f(\underline{x}_i) \quad (\text{D.2})$$

The set, $\{\underline{x}_i\}$, is generated by a biased random walk in the space which compares the relative probabilities, P , for the current and succeeding states. In this way the probability of being in a particular state at a particular step only depends explicitly on the prior step; such a sequence is called a “Markov chain”.

D.1 General requirements

Given an initial state, \underline{x}_0 , a (potentially un-normalized) probability distribution, $P(\underline{x})$, and a propagator for generating *trial* moves, $F : \underline{x}_n \rightarrow \underline{x}'_{n+1}$, then the sequence of the random walk is given recursively as follows:

$$\begin{aligned} \underline{x}_0 &= \underline{x}_0 \quad (\text{choose or otherwise given}) \\ \underline{x}_{n+1} &= \begin{cases} F(\underline{x}_n) & , \text{with probability: } \min \left[1, \frac{P(\underline{x}'_{n+1})}{P(\underline{x}_n)} \right] \\ \underline{x}_n & , \text{otherwise} \end{cases} \end{aligned} \quad (\text{D.3})$$

The probabilistic component of eq. D.3 can be handled by comparing a uniform, random variable on $[0, 1)$, ξ , to the target probability, p , and accepting if $\xi < p$. If this is confusing, think of the limiting cases: $p = 0$ and $p = 1$.

F , the propagator can take many forms, but here, we use a center-of-mass preserving variant of the one used in Brady *et al.* [64]. To generate trial moves for the random walk, we displace a single random atom, j , along a random coordinate, by a random amount, $\vec{\Delta}_j = \hat{\mu}_j(\xi - 0.5)\delta_x$. Where ξ is a uniformly distributed, random variable on $[0, 1)$, δ_x is a scaling factor and $\hat{\mu}_j \in \{\hat{x}, \hat{y}, \hat{z}\}$ is along a randomly selected Cartesian axis of the j th atom. After each such step, an additional move to preserve the center of mass is made by moving *all* atoms by the same, appropriate amount. Thus the total move for each atom i is:

$$\Delta \vec{r}_i = \hat{\mu}_j(\xi - 0.5)\delta_x \left(\delta_{i,j} - \frac{m_j}{M} \right) \quad ; \quad i = 1 \dots N \quad (\text{D.4})$$

where $M = \sum_{i=1}^N m_i$ is the total mass of the system. We encode this move through the equation:

$$F(\underline{x}) = \underline{x} + \Delta \underline{R} \quad (\text{D.5})$$

where $\Delta \underline{R} = (\Delta \vec{r}_1, \Delta \vec{r}_2 \dots \Delta \vec{r}_N)$ is the configuration-space vector of all single-atom displacements defined in eq. D.4.

There is a bit of lore surrounding Markov chains, which indicates that an acceptance ratio (i.e., the fraction of accepted trial steps) of 50% is “best”. Conversations with Jimmie Doll (one of the authors of [64]) indicated that this is very rough and that a better guide is perhaps “bigger than 10% and less than 90%”. One can adjust the

scaling factor, δ_x , to achieve the desired acceptance ratio during a trial phase. However, once sampling for the set in eq. D.2, δ_x must remain fixed or the chain will fail to satisfy detailed balance and the intended distribution will not be sampled. We describe a naïve, but effective implementation of a scheme to pick δ_x in section a later section on page 154.

D.2 Implementation specifics

We now give the specifics of our implementation of the foregoing procedure to sample points from the roaming region of the formaldehyde potential energy surface.

A Probability-like Function

In the first section we stated that we could use un-normalized probability distributions. This is because the distribution, P only appears in eq. D.3 in ratio with itself. This is quite convenient because it obviates the need to compute what would be—in this case and many others—a complicated, many-dimensional configurational integral.

Our definition of the roaming region is explained in detail in Section 2.3.3, but recall that the structure of the region is defined in terms of requirements that

1. The restoring force on the roaming hydrogen is of small magnitude:

$$F_{H^{(0)}} < F_{\max} \quad (\text{D.6})$$

2. The hydrogens are separated by more than a minimum and less than a maximum distance:

$$d_{\min} < |\vec{r}_{HH^{(0)}}| < d_{\max} \quad (\text{D.7})$$

Hydrogen-hydrogen separation is given by $\vec{r}_{HH^{(0)}}$ and we define the restoring force as:

$$F_{H^{(0)}} = -\hat{r}_{H^{(0)}-CM} \cdot \vec{\nabla}_{H^{(0)}} V(\mathcal{R}) \quad (\text{D.8})$$

where

$$\vec{\nabla}_{H^{(0)}} = \left(\frac{\partial}{\partial x_{H^{(0)}}}, \frac{\partial}{\partial y_{H^{(0)}}}, \frac{\partial}{\partial z_{H^{(0)}}} \right) \quad (\text{D.9})$$

and

$$\hat{r}_{H^{(0)}-CM} = (\vec{r}_{CM} - \vec{r}_{H^{(0)}}) / \|\vec{r}_{CM} - \vec{r}_{H^{(0)}}\| \quad (D.10)$$

These criteria can be formally encoded in a function as follows:

$$P_{\text{roaming}}(\mathcal{R}; F_{\max}, d_{\min}, d_{\max}) \propto \Theta(F_{\max} - F_{H^{(0)}}) \quad (D.11)$$

$$\cdot \Theta(\|\vec{r}_{HH^{(0)}}\| - d_{\min}) \quad (D.12)$$

$$\cdot \Theta(d_{\max} - \|\vec{r}_{HH^{(0)}}\|) \quad (D.13)$$

where Θ is the step function:

$$\Theta(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (D.14)$$

To sample the potential energy landscape ensemble ([46] and Section 2.1.2), we can use a similar scheme:

$$P_{\text{pele}}(\mathcal{R}; E_L) \propto \Theta(E_L - V(\mathcal{R})) \quad (D.15)$$

where E_L is the landscape energy.

Combining equation D.15 with D.11:

$$P(\mathcal{R}) = P_{\text{pele}}(\mathcal{R}; E_L) \times P_{\text{roaming}}(\mathcal{R}; F_{\max}, d_{\min}, d_{\max}) \quad (D.16)$$

allows us to sample the portion of the roaming region that is also in the landscape ensemble for a given E_L .

It should be noted that since our combined function has binary outputs, the probabilities in eq. D.3 will always be 0 or 1. This is a special case of the more general class of problems than the Metropolis-Hastings algorithm is capable of handling—within the allowed region, we seek uniform sampling rather than a variable density.

Put another way, because we have no notion of “sort of roaming” or “sort of in the potential energy landscape ensemble”, the random walk will never do something “sort of bad” *i.e.*, will never take “uphill” steps. When Brady and co-workers[64] implemented this scheme, it was to sample the energy shell of the microcanonical ensemble. While there is no notion of “sort of the right energy” in the microcanonical ensemble, they used a pre-limit form of the delta function for their probability, which allowed their walker to meander back and forth across the energy shell.

Picking a Scaling Factor

For the linear propagator described in section D.1, we would like to optimize the scaling factor, δ_x , such that the acceptance ratio falls within the desired bounds.

The binary probability function could introduce some added complications, but we avoid them under the following assumption:

- The set to be sampled is dense; that is, points in the set are arbitrarily close to others in the set

This condition allows us to assume that for sufficiently small steps, moves will always fall within the acceptance region. This is important because it allows us to take the $\delta_x \rightarrow 0$ limit as yielding an acceptance ratio of 1. In the case of a binary probability function, the other limit, $\delta \rightarrow \infty$, implies an acceptance ratio equal to the value $\langle P(\vec{x}) \rangle$. In the case of the roaming region of formaldehyde, this average is 0 because the volume of the roaming region is finite¹.

It would be convenient to assume that the set is connected; that is, it is possible to translate between any two points in the set while remaining in the set. This would actually be a stronger requirement than the first assumption and would allow us to assume that the entire set is reachable from any x_0 . But, in the potential energy landscape ensemble (or microcanonical ensemble), we have no guarantees that such an assumption will hold².

The general scheme for selecting δ_x is this:

1. Guess an initial value for δ_x
2. Compute a Markov chain of length k while keeping track of the acceptance ratio
3. If the ratio is larger than desired, scale by $(1 + \alpha)$ to increase the length. Likewise, if the ratio is smaller than desired, scale by $(1 - \alpha)$ to decrease the step.
4. Go-to 2 and repeat M times.

¹or at least grows more slowly than the volume of the allowed regions of the landscape ensemble with E_L .

²Indeed, we strongly suspect that it does not; see Section 3.6.

If α is too large, the results will be unstable as δ_x oscillates in and out of the acceptable range. Too small and convergence will take a very long time. Similar observations can be made about k . We find acceptable results with $\alpha = 0.05$ and $k = 1000$ at $E_L = 37600 \text{ cm}^{-1}$. Using this plan for picking δ_x we achieved an acceptance ratio of 89.5% with a scaling factor of $\delta_x = 0.27562 \text{ } a_0$ after 1.1×10^6 total steps. This step size was then used for sampling the roaming region at all energies. All parameters used to select δ_x are given in the table below.

Parameter	Value
δ_x (guess)	$0.01 \text{ } a_0$
k	1000
target acceptance ratio	10% to 90%
α	0.05
M	1100
δ_x (final)	$0.27562 \text{ } a_0$
acceptance ratio	89.5%

The Formaldehyde Potential

Given that we have no reason to expect that the roaming region is self-connected, we should initiate our random walk from as many x_0 as possible. To seed the random walk we took configurations satisfying both the landscape and roaming criteria from molecular dynamics trajectories (see Section 2.3.2) that went to molecular products (roaming trajectories). These criteria yielded 200–9000 points in the roaming region for each landscape energy, E_L (see Table D.1 for details). Using each of the points as a different value for x_0 , we generated Markov chains of length 10^5 , recording values every 100 moves. These points were then shuffled and used as intermediate roaming points in the geodesic path-finding algorithm; see Section 2.3.4.

E_L / cm^{-1}	acceptance ratio	seed configurations
34500	0.827227	248
*35000	0.849454	2071
35500	0.865258	2588
*36000	0.875344	3379
36223	0.878148	4040
37000	0.889417	6025
*37600	0.895281	7627
37688	0.895886	7149
38000	0.898983	7168
38688	0.903714	8884
*38814	0.903394	8723
*41010	0.904021	7319
*45000	0.942785	6432

Table D.1: Energy-specific acceptance ratios and seed configurations

We give acceptance ratios and the number of seed configurations for all energies for which we sampled the roaming region. For all walks, $\delta_x = 0.27562 a_0$ was fixed as determined at $E_L = 37600 \text{ cm}^{-1}$. For each seed configuration, we computed walks of 10^5 steps, recording a configuration every 100 moves. An asterisk (*) indicates energies for which geodesics were actually computed.

APPENDIX E

INTEGRATING EXTERNALLY-SUPPLIED FORTRAN CODE

Joel Bowman’s group at Emory University provided us with the global potential energy surface (PES) for formaldehyde [54]. Briefly, the function was formed by splicing together several fits to two sets of high level (CCSD(T)/aug-cc-PVTZ and MR-CI/aug-cc-PVTZ) *ab initio* calculations. Bowman transmitted the function as a set of “zipped” FORTRAN source code files. When decompressed, the files represent about 5000 lines of code. For future checks of the integrity of these files, their sha256 checksums [88] are given here:

```
% sha256sum *.f
f708238435ffaa0ec30088327440d4ac6fcfdb3fe55d9e67ff08fc6e50525437 bonds.f
82f074e9c3d350e76c8ca152425b4c37258914bb6e1f12fd9014a9437f259cf0 getpot.f
3eea2fd4a3a5c52df4e803460700debfd00e59638653eb846c63839c03133f22 h2co-ccsdt.f
181db4ab7fe3aec03eb0ebaa6a3c397755d1b2ba6aceea31db744b5a9e10c39b h2co-mrci.f
d4d01e238d90375e4e43fc0756e5a552c4cd15d66db3821693bd8b624d0c63a8 main.f
%
```

As the provided code was written in FORTRAN and the project implemented in C, some effort had to be expended to interpolate between the two. Primarily, we want to be able to call the subroutine Getpot, contained in the file `getpot.f`.

This function and relevant declarations are given below. The declarations are taken from the file `main.f`, which calls `Getpot`.

```
REAL(8), DIMENSION(36,3) :: xx
REAL(8), DIMENSION(36) :: xmass
REAL(8) :: v,v1,v2,s
INTEGER :: natom

subroutine Getpot (natom,xmass,xx,v,v1,v2,s)
```

This snippet declares two double precision arrays of dimension 36×3 and 36×1 called `xx` and `xmass` respectively as well as 4 double precision floating point values and an integer. The subroutine `Getpot` takes these variables as parameters. A call to `getpot` reads the contents of `xx`, `xmass`, and `natom` then sets the values of `v`, `v1`, `v2`, and `s`. The variable `v` contains the value of the potential at the position specified by `xx`; the other variables can be used internally.

The following comment in the file `h2co-mrci.f`:

```
c      note h1 and h2 are not equivalent, it is assumed that
c      h2 is further away from the other atoms than h1
```

suggest that the two hydrogens used internally are not equivalent. Indeed, one sees differences on the order of a few percent when permuting the hydrogens if one is far removed from the formyl. Our best-guess interpretation of the comment is that the hydrogen most removed from the total center of mass should be input to Bowman's function as '`h2`'. We respect this stricture by including a check on the condition in our wrapper around Bowman's code. When it was not satisfied, we swap the hydrogens in the *external* function call while preserving our internal representation. In this way, the hydrogens are indeed identical to our wrapped function.

The two arrays are of length 36 because the code was written with reuse in mind. The `natom` parameter is used to specify the portion of the arrays used. The code accepts positions in Cartesian coordinates (in the order H₁, O, C, H₂), which are then converted to internal, rotationally and translationally invariant coordinates. The 4 atom system has 12 Cartesian coordinates and $12 - 6 = 6$ internal coordinates. The internal coordinate

system in the Bowman-group code is the set of all inter-atomic distances¹.

The sources were compiled to object files using the command `gfortran -O3 -c bonds.f getpot.f h2co-ccsdt.f h2co-mrci.f`. The flag `-O3` indicates the highest level of compiler optimizations and the flag `-c` indicates that the files should be compiled, but not linked.

Once compiled, the object files can be inspected with the command `nm`, e.g.:

```
% nm --defined-only getpot.o
0000000000000000 T getpot_
0000000000000010 d rech.1042
0000000000000008 d rehh.1049
0000000000000000 d reoh.1050
%
```

Here, we can see that the function included in the file `getpot.o` is `getpot_` (indicated by `T`). Note the added underscore; the naming convention is compiler-dependent². This is the name of the function we call from C code.

To call this function in a C source file, we declare an external, void function using the name found above:

```
extern void EXTGETPOT(const int * natom, const double xmass[36]
                      ,double xx[3][36], double * v, double * v1
                      ,double * v2, double * s);
```

Several things are noteworthy. First we constant, `EXTGETPOT`, so that the name of the FORTRAN function can be specified at compile time (for `gfortran`, from the GNU compiler collection, the appropriate call is `getpot_`). Second, all variables in FORTRAN are passed by reference. Therefore C variables that are not already pointers (arrays are fine) should be passed as pointers. Finally, FORTRAN's memory layout for multidimensional arrays is the transpose of C's. For this reason, `xx` is declared as 3×36 instead of 36×3 .

With an external function declared, it can be used in code! The following function takes a `gsl_vector*` describing the position of all centers and returns the potential

¹Were there more than 4 centers, this set would over-determine the system in 3D. $\frac{N(N-1)}{2} > 3N - 6$ for $N > 4$.

²The `xl` compilers used on IBM Blue Gene systems, for instance do not add the underscore.

energy of the configuration via a call to EXTGETPOT (perhaps `getpot_`) while respecting difference between `h1` and `h2`. Comments in the Bowman group's code suggest that the return value is in wavenumbers, but it is not (we verified this by comparison to *their* published values [54]). The value is returned in atomic units or Hartrees (1 Ha = $E_H = 2.19474 \times 10^5$ cm $^{-1}$; see also Section 2.1.1).

```

double potential_formaldehyde(const gsl_vector * r){
    assert(r->size == 12);

    static const int natom = 4;
    static const double xmass[36]={
        1837.15 // H
        ,29156.95 // O
        ,21874.66 // C
        ,1837.15 // H
    }; //will initialize the first 4 compnents and leave the others

    //double xx[3][36];
    static double v = 0, v1 = 0, v2 = 0, s = 0;
    static double xx[3][36]; //transpose of fortran's 36x3

    for (size_t coor = 0; coor < 3; coor++){
        for (size_t atom = 0; atom < 4; atom++){
            xx[coor][atom]=gsl_vector_get(r,(atom*3)+coor);
        }
    }

    if (HOCHgetCMDist(r,formaldehydeOffset.H1) >
        HOCHgetCMDist(r,formaldehydeOffset.H2)){
        double temp;
        const int H1 = 0;
        const int H2 = 3;
    }
}

```

```

    for (size_t coor = 0; coor < 3; coor++){
        temp = xx[coor][H1];
        xx[coor][H1] = xx[coor][H2];
        xx[coor][H2] = temp;
    }
}

EXTGETPOT(&natom, xmass, xx, &v, &v1, &v2, &s);

return v;
}

```

The following constants and helpers are also defined in a separate header file.

```

#ifndef EXTGETPOT
#define EXTGETPOT getpot_
#endif

const int formaldehydeNumCenters = 4;
const size_t formaldehydeNumDimensions = 12;
const double formaldehydeMassTotal = 54705.91;

const struct offset_t formaldehydeOffset = {
    .H1 = 0,
    .O = 3,
    .C = 6,
    .H2 = 9
};

const struct mass_t formaldehydeMass = {
    .H1 = 1837.15,

```

```

.O  = 29156.95,
.C  = 21874.66,
.H2 = 1837.15
};

const double formaldehydeMassArray[4] = { 1837.15,
                                         29156.95,
                                         21874.66,
                                         1837.15};
const size_t formaldehydeOffsetArray[4] = {0,3,6,9};

```

Using the constants defined above, we compute the Hydrogen to center-of-mass separation in the following function:

```

/*
gives the distance of the offset-specified atom from the total
center of mass.
*/
double HOCHgetCMDist(const gsl_vector *r, size_t offset){
    const size_t space = 3;

    double cm_array[space];
    gsl_vector_view cm = gsl_vector_view_array(cm_array, space);
    gsl_vector_set_zero(&cm.vector);

    for (int i = 0; i < formaldehydeNumCenters; i++){
        gsl_vector_const_view v = gsl_vector_const_subvector(
            r,formaldehydeOffsetArray[i],space);
        gsl_blas_daxpy(
            formaldehydeMassArray[i] / formaldehydeMassTotal,
            &v.vector, &cm.vector);
    }
}

```

```

gsl_vector_const_view atom = gsl_vector_const_subvector(
    r, offset, space);
gsl_vector_sub(&cm.vector, &atom.vector);

return gsl_blas_dnrm2(&cm.vector);
}

```

The C file is compiled to object code in the usually way:

```

% gcc -std=c99 -Werror -pedantic -Wall -Wmissing-prototypes \
      -Wstrict-prototypes -Wconversion -Wshadow -Wpointer-arith -Wcast-qual \
      -Wcast-align -Wwrite-strings -Wnested-externs -fshort-enums -fno-common \
      -g -O2 -DHAVE_INLINE -I/share/apps/include -c potential_formaldehyde.c
%

```

A number of errors are checked by enabling the compiler warnings and is recommended by the GNU Scientific Library (GSL) project [62]. Note the `-I` flag, which specifies the location of the GSL headers; these may be in an alternate location. As of this writing, the version installed on ted is 1.14; version 1.99 is installed under `/home/vale/toor/`.

With C and FORTRAN object files in hand, the program can be linked:

```

% gcc bonds.o getpot.o h2co-ccsdt.o h2co-mrci.o \
      potential_formaldehyde.o # ... other C objects follow \
      -lgsl -lm -lgfortran -lgslcblas -L/share/apps/lib \
%

```

The last line of flags link the GSL, C math, and GCC FORTRAN libraries respectively. The flag `-lgslcblas` specifies using the GSL Basic Linear Algebra Subprograms (BLAS) library. Another, optimized library could also be used. We attempted using the TACC implementation, GotoBLAS2³, using the flags `-lgoto2 -L/share/apps/GotoBLAS2`, but performance was substantively harmed⁴. The location of the GSL library must be

³See <https://www.tacc.utexas.edu/tacc-projects/gotoblas2>.

⁴Note to future (student) readers, it might be worth trying the ATLAS implementation too.

specified with an `-L` flag.

Two quick notes about performance: Upon profiling the code we used to compute geodesics, we found that approximately 95% of the program’s effort was spent in the externally supplied FORTRAN code. This does not imply that the Bowman code is inefficient, but just that fitting the formaldehyde potential is computationally demanding. One result of this observation is that the program to find geodesics could have been written in just about any reasonably performant language and we would have retained the same order-of-magnitude performance as long as it was able to efficiently call the Bowman code. Also, modern compilers are good things to use as compiler technology has improved and can generate faster code. The version of `gcc` on `ted` is ancient—4.4 came out nearly a decade ago. By installing newer versions (e.g. 6.2) we were able to achieve an *order of magnitude speed-up* in pure C code written for analysis.

E.1 Form of switching functions

The switching functions employed by Bowman’s group in constructing the fit are of the form given below:

$$w(x) = \begin{cases} 1 & x \leq 0 \\ 1 - 10x^3 + 15x^4 - 6x^5 & 0 < x < 1 \\ 0 & 1 \leq x \end{cases} \quad (\text{E.1})$$

There is a typo in their paper [54] that mistakenly shows the cubic term as a quadratic. We are sure this is a mistake because their source code uses the cubic term and because, while the published form does not have the properties they describe (reviewers?), the version in source does.