COMPLEX FLUID DYNAMICS: FROM LAMINAR TO GEOPHYSICAL FLOWS

BY

JOHN J. VEYSEY II

B.S. in Physics, Yale University, 1997
M.S. in Physics, University of Illinois at Urbana Champaign, 2000

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2006

Urbana, Illinois

COMPLEX FLUID DYNAMICS: FROM LAMINAR TO GEOPHYSICAL FLOWS

John J. Veysey II, Ph.D.
Department of Physics
University of Illinois at Urbana-Champaign, 2006
Nigel Goldenfeld, Advisor

In this dissertation, we present analytical, experimental, and numerical work relating to two problems in fluid flow.

We first treat the oldest, hardest, and arguably one of the most important problems in *boundary layer* theory: determining the drag on a sphere and an infinite cylinder moving at a fixed speed in a highly viscous fluid. We necessarily extensively review previous experimental and theoretical work, and then apply techniques based on the perturbative renormalization group to produce optimal "coarse-grained" approximations. These approximations lead to a new prediction for the drag coefficient, one which both reproduces and surpasses the results of matched asymptotics.

We next present investigations into the biocomplex system at Yellowstone National Park. This research, which includes field work, experimental measurements, and numerical modelling, has shown that microbes do not play an important role in the formation of large scale carbonate terraces. Although we found that microbial communities are highly correlated with depositional facies, we demonstrated that this is a consequence of tight correlations between aqueous chemistry and the underlying travertine. We show how these correlations can be used to reconstruct ancient depositional environments.

We also present highly successful minimal models which explain the formation of depositional facies and large scale travertine motifs, including the characteristic terrace architecture found at carbonate hot springs. Our numerical model predicts dynamical phenomena which are seen at real hot springs. It also produces patterns which have the same static statistical properties as those seen at real hot springs, and indicates that

hot spring terraces are indeed scale invariant; we show that the distribution of pond areas in both simulated landscapes and real hot springs obeys the same power law. By analogy with studies of vicinal surfaces, we identify another universal statistical characterization, the *terrace width distribution*. We present evidence from both simulated landscapes and experimental data that this distribution is a universal property of carbonate terraces, applying regardless of the details of local chemistry and biology.

There are three things you don't want to see made: Sausage, Legislation, and Interdisciplinary Research.

Acknowledgments

My seven years at Urbana-Champaign have been rich, both professionally and personally. For this, I owe an enormous debt to my advisor, Prof. Nigel Goldenfeld. He has taught me how to be a scientist and led me to and through a richer and more rewarding research project than I could have ever expected. He caught my attention in his Statistical Mechanics class when he suggested there might be more to physics than reductionism, and I have never looked back. His patience, enthusiasm, and willingness to apply the ideas of physics to other disciplines have left a lasting impact on my approach to problems, and fundamentally altered my view of what it means to be a physicist. His enthusiasm for *interesting* problems, regardless of category, is inspiring, and his zealous fairness has shown me how to work in scientific collaborations. His idealism, particularly his refusal to accept military funding, has set an example which I hope I can live up to.

I appreciate his willingness to let me digress into chemistry, microbiology, geology, and even journalism during my time at UIUC.

I am grateful to the other members of Nigel's group. Héctor García Martín, my officemate for five years, is a unique and wonderful man who repeatedly offered the best advice I received in graduate school: "It could be worse; it could be boring." I also thank him for mind-broadening political discussions and dragging me to salsa dancing lessons. Pak Yuen Chan's help and collaborations with Yellowstone modelling have been invaluable. Nich Guttenberg's irreplaceable assistance with computer visualization has sped up my research by months or years, and has made me look a lot

smarter than I really am. I also thank Tae Kim and Kalin Vetsigian for interesting, informative, and educational discussions while they were here.

I could not have joined the Yellowstone project without Bruce Fouke. His enthusiasm, openness, and willingness to consider any idea has made the project a success. Few people would have insisted that I come join him in the field, and he did it not just once, but three times. He showed me how geology should be done, and I am indebted to him for teaching me all that I know about the subject. I have also found him to be an enthusiastic supporter of and participant in the writing of joint papers.

I also thank other members of the Fouke group, in no particular order: George Bonheyo for his willingness to explain microbiology to ignorant physicists; Michael Kandianis for teaching me chemistry (and for some of the most wonderfully intense discussions!); Tom Schickel, an all around swell guy who worked with me on papers, weighed all those washers (repeatedly), and is one of the two best people with whom I've ever had the pleasure of working night shifts on a hot spring[1]; Roy Johnson for his organizational skills, research focus, well grounded attitude, and excellent questions; Kelly Zimmerman for her help on field trips; and Rob Sanford for sharing his knowledge.

I am indebted to the Yellowstone Park Rangers for their knowledge, assistance, and hot spring preservation. Brian Suderman, in particular, has my gratitude for his willingness to hike out to the hot springs once a week to take pictures for our time-lapse movie. He's been at this for a year and a half now, and his dedication leaves me in his debt.

Lisa Asplen has helped sustain me during the last two years of my life, particularly during the writing of this dissertation. Her support, companionship, and copy editing has helped me survive this process. She also supplied the spiffy quote.

Many other people have also helped make my time in Urbana-Champaign the best years of my life (so far!). Thanks to Annie McCloskey and Erika Ryser for being true

[1]Thanks, Nigel, for being the other one!

friends. Thanks to Mary Upton, my friend and dance partner. The Friday lunch crowd, both old and new. Thanks to my fellow grad students, Dyutiman, Michelle, Ziggy, Jeff, Aaron, Soren, Bojan, Evan, Dom, and Jim among many others. To my roommates, Eun-Ah Kim, Michael Lawler, and Ken Esler. Thanks to all those who helped me become a dancer, especially Yuri, Natalie, Becky, and Smitha. And thanks to Andy, Dustin, Carmel, and Juan. Without all these people, my years here would have been much poorer, or perhaps I'd have just graduated sooner.

Last, but by no means least, I thank my family, the foundation on which all my work is built. My parents, John and Pam Veysey, have done nothing but help, encourage, and support me throughout my life, and I would not be here (literally and metaphorically!) if not for them. Their unconditional love and support helped make this dissertation possible. Hi Mom!

I gratefully acknowledge the support of the National Science Foundation through Grant No. NSF-DMR-99-70690 and through research awards from the NSF Biocomplexity in the Environment Program (EAR 0221743). The conclusions presented herein are those of the researchers and not necessarily those of the National Science Foundation.

Contents

1 Introduction **1**
- 1.1 Biocomplexity 4
  - 1.1.1 YNP Project Goals 5
  - 1.1.2 My Contributions 6
  - 1.1.3 Summary of Main Findings 6
- 1.2 The Renormalization Group and Minimal Models 7
  - 1.2.1 Low Reynolds Number Flow 9
- 1.3 Dissertation Outline 10
- 1.4 Publications 11

2 Introduction to Low R Flow **13**
- 2.1 Mathematical Formulation of the Problem 15
  - 2.1.1 Flow Past A Cylinder 17
  - 2.1.2 Flow Past A Sphere 18
  - 2.1.3 Calculating the Drag Coefficient 19

3 History of Low R Flow Studies **25**
- 3.1 Experiments and Numerical Calculations 25
  - 3.1.1 Measuring the Drag on A Sphere 26
  - 3.1.2 Measuring the Drag on a Cylinder 31
- 3.2 Theoretical History 34
  - 3.2.1 Stokes and Paradoxes 35
  - 3.2.2 Oseen's Equation 37
  - 3.2.3 Matched Asymptotics 58
  - 3.2.4 Other Theories 67
  - 3.2.5 Terminology 68
- 3.3 Uniformly valid approximations 70
  - 3.3.1 The correct way to calculate C_D 71

4 The Renormalization Group Applied to Low R Flow **77**
- 4.1 Introduction to the Renormalization Group 77
  - 4.1.1 Detailed Analysis of the "Terrible Problem" 79
- 4.2 Flow Past A Cylinder 95
  - 4.2.1 Rescaling 95
  - 4.2.2 Naïve Perturbation Analysis 96
  - 4.2.3 $\mathcal{O}(R^0)$ Solution 97

4.2.4 $\mathcal{O}(R^1)$ Solution . . . 97
4.2.5 Renormalization . . . 103
4.2.6 Calculating the Drag Coefficient . . . 108
4.3 Flow Past a Sphere . . . 109
4.3.1 Rescaling . . . 109
4.3.2 Naïve Perturbation Analysis . . . 111
4.3.3 $\mathcal{O}(R^0)$ Solution . . . 111
4.3.4 $\mathcal{O}(R^1)$ Solution . . . 112
4.3.5 Secular Behavior . . . 113
4.3.6 Renormalization . . . 114
4.3.7 Meeting the Boundary Conditions . . . 114
4.3.8 Calculating the Drag Coefficient . . . 118
4.4 Conclusions . . . 120

5 Biocomplexity at Yellowstone National Park . . . 123
5.1 Introduction . . . 123
5.2 Why Yellowstone? . . . 124
5.3 Geologic setting . . . 127
5.4 Carbonate Chemistry Primer . . . 130
5.5 Introduction to the facies model . . . 134
5.6 Many Important Length Scales . . . 136
5.7 Characterization of spring water . . . 137
5.8 Definition of the primary flow path . . . 140
5.9 Conclusions . . . 144

6 Relating Modern Depositional Environments to Travertine . . . 146
6.1 Introduction . . . 147
6.2 Brief introduction to microbial work at YNP . . . 150
6.2.1 The 16s rRNA gene method . . . 152
6.2.2 What is an OTU? . . . 153
6.3 Microbial Sampling at YNP . . . 153
6.4 Results in Microbial Ecology . . . 154
6.4.1 Statistical analyses . . . 154
6.4.2 Results in YNP microbial ecology . . . 156
6.4.3 Discussion . . . 160
6.5 Spring Water Chemistry . . . 162
6.5.1 Variability and fluctuations in hot spring water . . . 162
6.5.2 Relating the rock record to the depositional environment . . . 166
6.6 Reconstructing Modern and ancient Environments . . . 168
6.7 Conclusions . . . 169
6.8 Appendix A: Details of flux measurement . . . 170

7 Processes which control $CaCO_3$ precipitation . . . 173
7.1 Introduction . . . 173
7.2 Experimental Methodology at AT-3 . . . 174
7.2.1 Measuring DIC . . . 179
7.3 Analyzing aqueous chemistry along the primary flow path . . . 182
7.4 Looking for photosynthesis . . . 187
7.5 Comparing microorganism metabolic activities with inorganic processes 189
7.6 Conclusions . . . 193

8 Minimal models of terrace formation . . . 195
8.1 Introduction . . . 195
8.1.1 Examples of minimal models in geomorphology . . . 197
8.1.2 What are cellular models? . . . 199
8.1.3 Why CDS models? . . . 200
8.2 2D cellular models . . . 201
8.3 The 3D CDS model . . . 204
8.3.1 Step 1: Adding Water . . . 205
8.3.2 Step 2: Fluid Transport . . . 206
8.3.3 Step 3: Water chemistry . . . 208
8.3.4 Step 4: Precipitation . . . 210
8.4 Conclusions . . . 211
8.5 Appendix 1: Pond filling algorithm . . . 212

9 Simple models can explain complex phenomena . . . 214
9.1 Travertine Domes . . . 215
9.1.1 Analytical Results . . . 216
9.1.2 The Role of Surface Tension . . . 218
9.1.3 Simulated Domes . . . 219
9.1.4 θ_c: Combining analytical and numerical results . . . 219
9.2 Terraces . . . 221
9.2.1 The damming instability . . . 222
9.2.2 Simulated terraces . . . 223
9.3 Time-Lapse Photography of Canary Springs . . . 227
9.4 Statistical Analysis of Simulated Landscapes . . . 230
9.4.1 Distribution of Pond Areas . . . 231
9.4.2 Calculating the Terrace Width Distribution . . . 233
9.4.3 Spatial Dependence of the Average Terrace Width . . . 237
9.4.4 Pond Aspect Ratio . . . 238
9.5 Conclusions . . . 240
9.6 Appendix 1 . . . 241

10 Conclusions . . . 243
10.1 Reflections on Interdisciplinary Research . . . 245

References . . . 248

Vita **263**

Chapter 1

Introduction

In this dissertation, we present research which ranges from approximate analytic solutions of the Navier-Stokes equations to biological and chemical measurements at hot springs to numerical models of precipitative pattern formation. These diverse topics are unified by two underlying themes. First, they all involve fluid flow. But more importantly, they involve fluid flow problems which we can solve using *minimal models*. It is this approach, or viewpoint, which truly unifies the analytical, experimental, and numerical results presented in the next eight chapters.

Exactly what we mean by a minimal model will be made explicit by the theoretical work in this dissertation. For the moment, we explain it in contrast to the reductionistic approach which scientists have traditionally used to explain natural phenomena. The reductionist world view aims to understand a system by understanding its parts. This viewpoint goes something like this: All matter is made up of quarks and leptons; these interact through the gravitational, electroweak, and strong forces, which are mediated by gravitons, W and Z bosons, photons, and gluons; through these interactions, these particles form atoms, which in turn form molecules; these molecules combine to form other molecules, which are the basis for proteins. These proteins combine to form organelles, which are the building blocks of cells; groups of cells form tissues, which in turn form organs; these organs are components of organisms; both cells and organisms

can form communities, which then interact to create entire ecosystems. This line of thought continues, but we can stop here, because the very length of this list makes our point. Knowing that all matter is made of quarks and leptons is of very little use to anyone who wants to determine whether microbes control pattern formation in mineral precipitation. Efforts to explain all aspects of all problems in terms of fundamental building blocks suffer from a myopia which Nigel Goldenfeld and Yoshi Oono refer to as "excessive realism." In other words, by focusing on the trees, one cannot see the forest.

The shortcomings of reductionism extend beyond an inability to see important overall processes or emergent properties. In spite of immense increases in computational power, some systems are simply intractable when broken down into smaller parts. For instance, even though the governing principles are well understood, using detailed quantum mechanics to predict the aggregate properties (such as energy levels) of groups of atoms requires vast computational resources. Techniques such as quantum Monte Carlo, for instance, are currently limited to a few hundred atoms.

In contrast to the reductionist viewpoint, this dissertation examines features of complex systems which are *insensitive* to the specific microscopic details and underlying phenomena. For instance, the ideal gas law does not require quarks, leptons, or even any specific kind of atom or molecule. It simply requires that the gas be composed of *a lot* of small particles which only interact by bouncing off each other. Of course, the ideal gas law does not perfectly describe real gases, but it does approximately explain the behavior of many real gases under a broad range of conditions. The ideal gas law is a minimal model which only depends on a crude description of particle interactions and a sufficiently large number of particles.

Minimal models attempt to explain aspects of complex systems through a "top-down" approach. They do this by identifying characteristics of the system which are *universal*[1], e.g., insensitive to most of the details of their microscopic building

blocks. The goal of minimal modelling is to identify the dynamical features of a system which are required to explain the large scale phenomena of interest. The model should only contain the essential "ingredients" — removing any components should render it incapable of explaining the universal features. By building these models, we can explain and predict some large scale system properties.

It is important to note that this approach cannot work in all systems, or explain all phenomena. Some processes depend irreducibly on specific microscopic details. Minimal modelling is a useful approach only with complex systems that exhibit universal behavior. In condensed matter physics, phase transitions have been intensively studied for exactly this reason: very different systems can exhibit the same behavior when in critical regimes[1]. In these regimes, system properties often obey power laws, and are scale-invariant. This dissertation focuses on a problem in geological precipitative pattern formation which also exhibits statistical scale-invariance.

However, the scale invariance is incidental. Many problems which do not exhibit scale invariance can still be understood by minimal models; Chapters 2, 3, and 4 are an example of such a problem, as is recent work in our group on modelling materials microstructure using the phase field crystal models[2, 3, 4]. In earlier work in Nigel Goldenfeld's group, a very simple minimal model, encoded as a five-line computer program, was used to predict the functional form of space-time correlations between topological defects during the kinetics of the isotropic-nematic transition in liquid crystals[5]. Remarkably, this prediction[5], which has no adjustable parameters, agreed with later experiments[6]. The minimal modelling approach to physics is distinct from the conventional scientific method advocated by Popper, Kuhn, and others[7, 8, 9], and this has attracted the attention of philosophers of science[10].

1.1 Biocomplexity

The core of this dissertation involves research at hot springs at Yellowstone National Park (YNP). This project is a broad collaboration which includes geologists, geomicrobiologists, geochemists, microbial ecologists, and physicists. The overarching goal of this collaboration is to explain the unique terraced patterns which form at carbonate hot springs. In particular, we want to understand how microbes impact carbonate deposition, and determine whether they play an important role in the formation of the characteristic terraced architecture. These hot springs are introduced in detail in Chapter 5.

The YNP project falls under the rubric of *biocomplexity*. While ostensibly jargon, this term means something quite specific in the context of our program. The hot springs at Yellowstone are a complicated system: fluid flow, nonequilibrium chemistry, and diverse, abundant microbial communities are all elements of the problem. But it is a system in which the microbial communities change their environment. Their metabolic activities alter spring water chemistry, and their presence can influence precipitation rates. Precipitated minerals, in turn, impact fluid flow patterns, resulting in further changes to the microbial environment. Our project aims to understand the patterns which emerge — both physically and microbiologically — from these layered interactions. Hence the term biocomplexity: we are investigating pattern formation in a complex system where the population dynamics of microbial communities may have a non-trivial impact on their host ecosystem, and vice-versa.

As such, the tools, techniques, and experience gained at YNP may one day be applicable to understanding the global interactions between microbial life and the rest of the terrestrial biosphere. Ultimately, our group would like to initiate a systems biology of the environment, mapping out the coupling between intracellular metabolic pathways, community metabolism, and environmental biogeochemical cycles, with a view toward identifying any universal patterns of flux distribution or global organization.

This dissertation is a first step toward this overall program.

1.1.1 YNP Project Goals

The primary goal of the YNP project is to determine whether the activities of specific living microbes, or microbial communities, are required to create the terraced architectures universally observed in carbonate springs. If specific carbonate motifs form only in the presence of microbes, then finding these motifs in either the ancient rock record or extraterrestrial carbonates implies the existence of these microbes on ancient earth or other planets. To address this question, the YNP principal investigators (Bruce Fouke, Nigel Goldenfeld, and Allison Murray) proposed a tripartite program.

First, we undertake experiments at YNP in which microbes were (a) removed from the spring water and (b) killed but not removed. By comparing the precipitative patterns which form in this treated spring water with those in natural spring outflow, we can determine whether microbes are required for the formation of the characteristic terraced architecture. Moreover, if microbes are needed, we can determine whether they play an active metabolic role, or whether they simply serve as passive nucleation sites.

The second piece of the project is observational. By conducting an extensive, simultaneous survey of microbial communites, spring water chemistry, carbonate crystal form, and aggregate morphology, we can see whether meaningful correlations exist between the aqueous environment and the carbonate features which are preserved in the rock record. These correlations could allow us to determine whether precipitative patterns are uniquely associated with specific depositional environments. They also could constrain the role played by microbial activities and chemical processes in the formation of carbonate terraces.

Finally, the YNP project aims to determine the detailed spatial characteristics of carbonate terraces. To this end, we wish to make minimal models which describe the

growth of terrace patterns. These minimal models could be used to determine which effective dynamical processes are needed to reproduce the observed universal pattern. Moreover, if these models showed that the presence of microorganisms is (or is not) required for the formation of carbonate terraces, they would help fulfill the overall YNP project goals.

1.1.2 My Contributions

I have contributed field work, experimental measurements, extensive data analysis, and numerical modelling to the YNP project. This work has helped complete the last two project aims outlined above. I have helped measure the chemical and physical properties of spring water, as well as concomitant precipitation rates. Analyzing this data set, I looked for systematic correlations between aqueous chemistry and precipitated carbonate, and identified important physical controls on precipitation rates. In conjunction with other project members, I have also examined the correlations between microbial communities and underlying carbonate. Finally, and most importantly, I have built numerical minimal models which couple shallow, turbulent fluid flow to evolving landscapes. I have qualitatively and quantitatively compared the predictions of these models with patterns at real hot springs.

1.1.3 Summary of Main Findings

Our work has shown that microbes do not play an important role in the formation of large scale carbonate terraces. They are highly correlated with underlying mineralogy, but this is due to the tight correlations between aqueous chemistry and precipitated carbonate. These correlations can be used to reconstruct ancient depositional environments. We have also shown that any impact of microbial metabolic activities on precipitation is negligible in comparison to physical controls (such as carbon dioxide degassing), although microbes may be important as passive nucleation sites.

More importantly, we have built a minimal model of fluid flow and precipitation which explains the formation of large scale travertine motifs, including the characteristic terrace architecture. This model does not require any microorganisms, and it predicts dynamical phenomena which have been seen subsequently at real hot springs. It also produces patterns which have the same statistical properties as those seen at real hot springs, and indicates that hot spring terraces are indeed scale invariant. The most significant of these statistical characterizations is the *terrace width distribution* (Chapter 9). This characteristic quantity appears to be universal, applying to all terraces regardless of the details of local chemistry and biology.

1.2 The Renormalization Group and Minimal Models

As mentioned previously, this dissertation presents experimental, analytical, and numerical techniques which lead to minimal models of complex fluid dynamics. Generically, these models describe universal, scale invariant natural phenomena using partial differential equations (PDEs) or computer programs (see Chapter 8). We demonstrate that these models reproduce real features, at least in a statistical sense, by computing quantities such as scaling laws. This approach can also be used to demonstate that analytical, PDE based models capture the same physics as the algorithms which govern numerical simulations.

We can rigorously extract predictions of universal behavior from our PDEs by using the *renormalization group* (RG) [1]. Conceptually, this approach takes a model of small scale dynamics, and *coarse grains* it to obtain an effective description of phenomena at large scales. This process averages out local fluctuations and unimportant microscopic details. As such, it is a "lossy" transformation, in the sense that it cannot be reversed to regain the original microscopic description. However, the effective large scale behavior

which results from this coarse graining describes the universal properties of interest.

In the case of YNP, we would like to model the essential processes, such as fluid flow, precipitation rates, and microbial population dynamics using a set of coupled PDEs. By analyzing these PDEs using RG techniques we can derive approximate solutions that capture universal aspects of hot spring landscapes. For instance, we could predict that the resulting landscape is statistically scale invariant (e.g., [11]). The RG approach formally extracts macroscopic patterns and statistical descriptions (such as power laws) from the governing PDEs. In doing so, it allows us to identify which parts of the governing equations control these large scale patterns, and which parts are irrelevant. These analytical predictions can then be compared with real landscapes and with the results of detailed, numerical solutions of the governing equations. This comparison ultimately validates the minimal model and would indicate that we have captured the essential physical processes.

To build the techniques needed to apply RG to models of YNP, we first used this approach to solve a relatively "easy" problem in fluid mechanics: calculating the drag on a solid body at very low Reynolds number. Chapters 2, 3, and 4 detail how we used RG techniques to extract the coarse grained characteristics of solutions to the Navier-Stokes equations in the limit of highly viscous flow. By removing unimportant microscopic details, the RG procedure produces excellent approximate analytical solutions. In essence, we are using RG to build a minimal model of fluid flow at very small Reynolds number. This first project is only "easy" in comparison to the YNP project, which aims to couple descriptions of fluid flow (like the Navier-Stokes equations) with microbial population dynamics, non-equilibrium chemistry, and a moving boundary problem.

1.2.1 Low Reynolds Number Flow

The oldest, hardest, and perhaps one of the most important problems in *boundary layer* theory was posed by Stokes in 1851[12]. The problem statement is simple: determine the drag on a sphere and an infinite cylinder moving at a fixed speed in a highly viscous fluid. Mathematically, this requires solving the Navier-Stokes equations in the limit of vanishing Reynolds number. This simple problem has given rise to over 150 years of theoretical and experimental confusion; it took over 100 years to obtain a justifiable approximate solution, and that required the invention of a new branch of applied mathematics, *matched asymptotics*. Moreover, although this technique finally put the whole subject on a firm theoretical footing, these results apply only to infinitesimally small Reynolds number. Ian Proudman, who helped obtain the first matched asymptotics result for a sphere, expressed it this way: "It is therefore particularly disappointing that the numerical 'convergence' of the expansion is so poor."[13] In spite of its failings, Proudman's solution was the first mathematically rigorous one; all preceding theoretical efforts were worse.

All solutions to this problem involve approximations which are not always rigorous or explicitly stated. To make matters worse, this field is rife with historical confusion, disagreements over attribution, mysterious factors of two, conflicting terminology, nonstandard definitions, and language barriers. Even a recent article attempting to resolve this quagmire[14] makes mistakes regarding publication dates and scientific priority. This tortured history has left a wake of experiments and numerical calculations which are of widely varying quality, although they can appear to agree when not examined closely. It turns out that the finite size of experimental systems has a dramatic effect on measurements and simulations, a problem not appreciated by early workers.

In Chapter 3, we review all prior theoretical and experimental results. We identify the five calculations and measurements which are accurate enough, and which extend to sufficiently small Reynolds number, to be useful for evaluating theoretical predictions.

Furthermore, we review the history of all theoretical contributions, and clearly present the methodologies and approximations behind previous solutions. In doing so, we eliminate all confusion over chronology and attribution. We conclude by comparing the best experimental results with our "state-of-the-art" RG-based theoretical prediction. This exercise makes the shortcomings that Proudman lamented all too clear.

These theoretical attempts are littered with "paradoxes", revisions, ad-hoc justifications, and failures. These problems arise because the governing equations are singular in the limit of zero Reynolds number. This singularity results in a boundary layer at infinity, a problem which was not resolved until matched asymptotics[15]. Recently, techniques based on perturbative renormalization group have emerged as an alternative to matched asymptotic and multiple scales[16, 17]. These techniques, in contrast to earlier methods, require neither assumptions about the structure of the perturbation series nor asymptotic matching. They have been used to obtain superior approximate solutions to (for example) the Swift-Hohenberg equation, which is used as a model of Rayleigh-Benard convection[16]. However, until now these techniques had not been applied to the low Reynolds numbers problem which gave birth to matched asymptotics. My work shows that RG techniques yield economical expansions and superior solutions to these classic problems.

1.3 Dissertation Outline

The first part of this dissertation is theoretical, focusing on obtaining analytical approximate solutions to the low Reynolds number problems. Chapter 2 introduces the mathematical framework needed to analyze these problems, states them explicitly, and explains how to calculate a drag coefficient given a solution to the Navier-Stokes equation. Chapter 3 presents the experimental and theoretical history of these problems, and chapter 4 explains how to apply RG techniques to these equations, culminating in

a new solution to an old problem.

The second part turns to YNP, beginning with our experimental work. Chapter 5 introduces the hot springs at Yellowstone, explaining the geology and chemistry needed to understand our work. It summarizes our field expeditions, and presents the conceptual tools needed to organize and analyze our extensive data set. Chapter 6 describes the macroscopic correlations between microbial communities, aqueous chemistry, and precipitated carbonate which we found in our data. It concludes by explaining how this result can be used for paleoenvironmental reconstructions. Chapter 7 explains how further data analysis suggests that turbulent degassing is a primary control on precipitation, and demonstrates that the impact of microbial metabolic activities is negligible in comparison to other chemical changes.

The final part of this dissertation presents the minimal models which we have developed to explain macroscopic pattern formation at YNP. Chapter 8 motivates and describes these models, detailing my numerical contributions. The results of these models, including quantitative comparisons with patterns from real hot springs, are given in chapter 9. The modelling program is highly successful, reproducing two unique carbonate patterns, and predicting landscapes which exhibit the same statistical scale invariance seen in nature.

1.4 Publications

Portions of this dissertation have already appeared in print. Part of chapter 6 has appeared in the *Canadian Journal of Earth Science*[18], part is being reviewed for publication in *Geology*[19], and the rest will soon be submitted to *Environmental Microbiology*. Half of chapter 9 is being reviewed for publication in *Physical Review Letters*[20], and the rest will be submitted as quickly as possible. We anticipate submitting the contents of chapters 2—4 for publication as a review article, with a letter

outlining my new RG calculations of low Reynolds number drag.

Chapter 2

Introduction to Low R Flow

In this and the following two chapters, we begin a reexamination of the historical problem of determining the steady state solutions for the flow of a uniform stream past fixed bodies. In particular, we study flow past a sphere and an infinite cylinder. We assume the fluid to be incompressible, and work in the low Reynolds number regime ($R \approx 0$), with the goal of calculating the drag force exerted on the fixed body. The actual physical problem concerns a body moving at constant velocity in an infinite fluid, where the fluid is at rest in the laboratory frame. In practice, it is more convenient to analyze the problem using an inertial frame moving with the fixed body, an approach which is entirely equivalent.[1]

This is a classic problem, with an extensive literature dating back to 1851, when Stokes analyzed flow past a sphere and a cylinder using perturbative expansions in the Reynolds number[12]. He obtained a first order approximation for the drag on a sphere, but was unable to obtain a solution for the cylinder, a failure referred to as *Stokes' paradox*. In 1889, Whitehead found that no second order approximation could be found for flow past a sphere, a failure known as *Whitehead's paradox*. These complications arise because the Navier-Stokes equations, which govern the flow, constitute a singular

[1]Nearly all workers, beginning with Stokes[12], use this approach, which Lindgren[14] refers to as the "steady" flow problem.

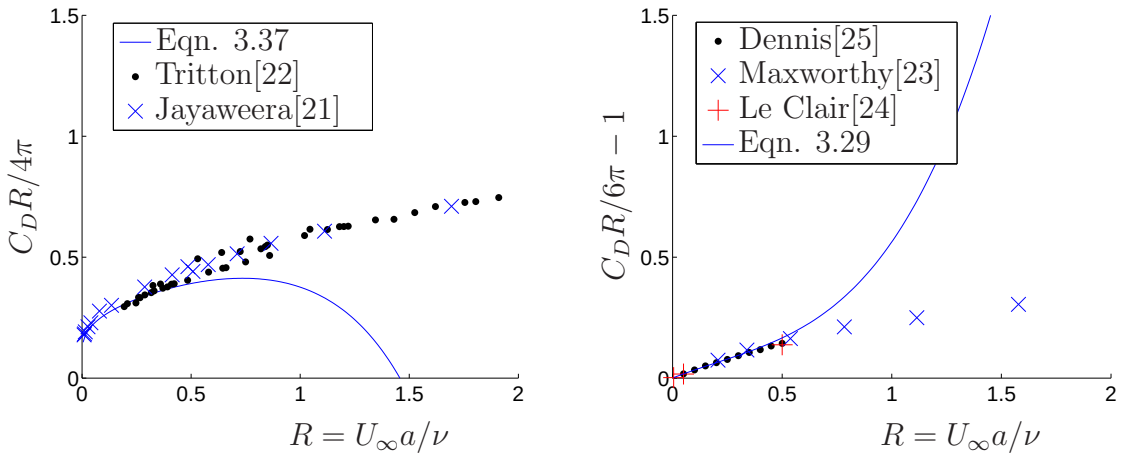

Figure 2.1: Comparing experiment with "state of the art" theoretical predictions for a sphere (right) and a cylinder (left).

perturbation problem (with respect to small Reynolds numbers).

In the early 1950s, a solution of these problems led to the development of a new branch of applied mathematics known as *matched asymptotics*. However, the best current matched asymptotic results have deficiencies, as can be seen from Figure 2.1. This plot compares experimental measurements and theoretically predicted drag coefficients as a function of the Reynolds number. Large deviations for $R \sim 0.5$ demonstrate the need for theoretical predictions which are more robust for small but non-infinitesimal Reynolds numbers.

Although in principle the matched asymptotics results can be systematically extended by working to higher order, this is not practical. The complexity of the governing equations prohibits further improvement. Here we show that techniques based on the renormalization group ameliorate some of the technical difficulties, and result in a more accurate drag coefficient at small but non-infinitesimal Reynolds numbers. Given the historical importance of the techniques developed to solve these problems, we anticipate that our solutions will be of general methodological interest.

Quantity	Description
$\vec{r}$	Coordinate Vector
$\vec{u}(\vec{r})$	Velocity Field
ρ	Fluid Density
$p(\vec{r})$	Pressure
ν	Kinematic Viscosity
a	Characteristic Length of Fixed Body
$\vec{u}_\infty$	The Uniform Stream Velocity

Table 2.1: Quantities needed to characterize low R flow past a rigid body

2.1 Mathematical Formulation of the Problem

Flow past a sphere or circle is shown schematically in Figure 2.2. The body has a characteristic length scale, which we have chosen to be the radius (a), and it is immersed in uniform stream of fluid. At large distances, the undisturbed fluid moves with velocity $\vec{u}_\infty$.

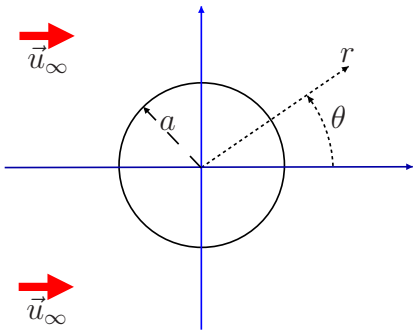

Figure 2.2: Schematic For Flow Past a Sphere or Cylinder

The quantities shown in Table 2.1 characterize the problem. We assume incompressible flow, so $\rho = const.$ The continuity equation (Eqn. 2.1) and the time-independent Navier-Stokes equations (Eqn. 2.2) govern steady-state, incompressible flow.

$$\nabla \cdot \vec{u} = 0 \tag{2.1}$$

$$(\vec{u} \cdot \nabla \vec{u}) = -\frac{\nabla p}{\rho} + \nu \nabla^2 \vec{u} \tag{2.2}$$

These equations must be solved subject to two boundary conditions, given in Eqn.

Dimensionless Quantity	Definition
$\vec{r}^*$	$\vec{r}/a$
$\vec{u}^*(\vec{r})$	$\vec{u}(\vec{r})/\lvert\vec{u}_\infty\rvert$
$p^*(\vec{r})$	$a\ p(\vec{r})/\rho\ \nu\ \lvert\vec{u}_\infty\rvert$
$\vec{\nabla}^*$	$a\ \vec{\nabla}$

Table 2.2: Dimensionless Variables

2.3. First, the *no-slip* conditions are imposed on the surface of the fixed body (Eqn. 2.3a). Secondly, the flow must be a uniform stream far from the body (Eqn. 2.3b). To calculate the pressure, one also needs to specify an appropriate boundary condition (Eqn. 2.3c), although as a matter of practice this is immaterial, as only pressure differences matter when calculating the drag coefficient.

$$\vec{u}(\vec{r}) = 0 \quad \vec{r} \in \{\text{Surface of Fixed Body}\} \tag{2.3a}$$

$$\lim_{|\vec{r}|\to\infty} \vec{u}(\vec{r}) = \vec{u}_\infty \tag{2.3b}$$

$$\lim_{|\vec{r}|\to\infty} p(\vec{r}) = p_\infty \tag{2.3c}$$

It is convenient to analyze the problem using non-dimensional quantities, which are defined in Table 2.2. When using dimensionless variables, the governing equations assume the forms given in Eqns. 2.4 and 2.5, where we have introduced the *Reynolds Number*, $R = |\vec{u}_\infty|a/\nu$, and denoted scaled quantities by an astrisk.

$$\nabla^* \cdot \vec{u^*} = 0 \tag{2.4}$$

$$R(\vec{u}^* \cdot \nabla^*)\vec{u}^* = -\nabla^* p^* + \nabla^{*2}\vec{u}^* \tag{2.5}$$

The boundary conditions also transform, and will later be given separately for both the sphere and the cylinder (Eqns. 2.13, 2.9). Henceforth, the * will be omitted from our notation, except when dimensional quantities are explicitly introduced. It is

useful to eliminate pressure from Eqn. 2.5 by taking the curl and using the identity $\nabla \times \nabla p = 0$, leading to

$$(\vec{u} \cdot \nabla)(\nabla \times \vec{u}) - ((\nabla \times \vec{u}) \cdot \vec{u}) = \frac{1}{R}\nabla^2(\nabla \times \vec{u}) \tag{2.6}$$

2.1.1 Flow Past A Cylinder

For the problem of the infinite cylinder, it is natural to use cylindrical coordinates, $\vec{r} = (r, \theta, z)$. We examine the problem where the uniform flow is in the $\hat{x}$ direction (see Figure 2.2). We will look for 2-d solutions, which satisfy $\partial_z \vec{u} = 0$.

Since the problem is two dimensional, one may reduce the set of governing equations (Eqns. 2.4 and 2.5) to a single equation involving a scalar quantity, the *Lagrangian* stream function, usually denoted $\psi(r, \theta)$. It is defined by Eqn. 2.7.[2]

$$u_r = \frac{1}{r}\frac{\partial \psi}{\partial \theta} \qquad u_\theta = -\frac{\partial \psi}{\partial r} \qquad u_z = 0 \tag{2.7}$$

This definition guarantees that equation (2.4) will be satisfied[27]. Substituting the stream function into equation (2.6), one obtains the governing equation (Eqn. 2.8). Here we follow the compact notation of Proudman and Pearson[26, 28].

$$\nabla_r^4 \psi(r, \theta) = -\frac{R}{r}\frac{\partial(\psi, \nabla_r^2)}{\partial(r, \theta)} \tag{2.8}$$

where

$$\nabla_r^2 \equiv \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}$$

The boundary conditions which fix $\vec{u}(\vec{r})$ (Eqns. 2.3a, 2.3b) also determine $\psi(r, \theta)$ up to an irrelevant additive constant.[3] Eqn. 2.9 gives the boundary conditions expressed

[2] Although many authors prefer to solve the vector equations, I follow Proudman and Pearson[26]

[3] The constant is irrelevant because it vanishes when the derivatives are taken in Eqn. 2.7.

in terms of stream functions.

$$\begin{aligned} \psi(r=1,\theta) &= 0 && \text{(2.9a)} \\ \left.\frac{\partial\psi(r,\theta)}{\partial r}\right|_{r=1} &= 0 && \text{(2.9b)} \\ \lim_{r\to\infty}\frac{\psi(r,\theta)}{r} &= \sin(\theta) && \text{(2.9c)} \end{aligned}$$

To calculate the drag for on a cylinder, we must first solve Equation 2.8 subject to the boundary conditions given in Eqn. 2.9.

2.1.2 Flow Past A Sphere

To study flow past a sphere, we use spherical coordinates: $\vec{r} = (r, \theta, \phi)$. We take the uniform flow to be in the $\hat{z}$ direction. Consequently, we are interested in solutions which are independent of ϕ, because there can be no circulation about the $\hat{z}$ axis.

Since the problem has axial symmetry, one can use the *Stokes' stream function* (or Stokes' current function) to reduce Eqns. 2.4 and 2.5 to a single equation. The stream function is defined through the following relations:

$$v_r = \frac{1}{r^2\sin\theta}\psi_\theta \qquad v_\theta = -\frac{1}{r\sin\theta}\psi_r \qquad v_\phi = 0 \tag{2.10}$$

These definitions guarantee that Eqn. 2.4 will be satisfied. Substituting Eqn. 2.10 into Eqn. 2.6, one obtains the governing equation for $\psi(r,\theta)$[26]:

$$D^4\psi = R\left(\frac{1}{r^2}\frac{\partial(\psi, D^2\psi)}{\partial(r,\mu)} + \frac{2}{r^2}D^2\psi L\psi\right) \tag{2.11}$$

In this equation,

$$\begin{aligned} \mu &\equiv \cos\theta \\ D^2 &\equiv \frac{\partial^2}{\partial r^2} + \frac{1-\mu^2}{r^2}\frac{\partial^2}{\partial\mu^2} \\ L &\equiv \frac{\mu}{1-\mu^2}\frac{\partial}{\partial r} + \frac{1}{r}\frac{\partial}{\partial\mu} \end{aligned}$$

Here we follow the notation of Proudman and Pearson[26]. Other authors, such as Van Dyke[29] and Hinch[28], write their stream function equations in an equivalent, albeit less compact, notation.

As in the case of the cylinder, the boundary conditions which fix $\vec{u}(\vec{r})$ (Eqns. 2.3a, 2.3b) determine ψ up to an irrelevant additive constant. The transformed boundary conditions are given by Eqn. 2.13.

$$\psi(r=1,\mu) = 0 \tag{2.13a}$$

$$\left.\frac{\partial\psi(r,\mu)}{\partial r}\right|_{r=1} = 0 \tag{2.13b}$$

$$\lim_{r\to\infty}\frac{\psi(r,\mu)}{r^2} = \frac{1}{2}\left(1-\mu^2\right) \tag{2.13c}$$

In these chapters, we obtain approximate solutions for Eqn. 2.8 (subject to Eqn. 2.9), and Eqn. 2.11 (subject to Eqn. 2.13). These solutions are then used to calculate drag coefficients, which we compare to experimental results.

2.1.3 Calculating the Drag Coefficient

Once the Navier-Stokes equations have been solved, and the stream function is known, calculating the drag coefficient, C_D, is a mechanical procedure. We follow the methodology described by Chester and Breach[13]. This analysis is consistent with the work done by Kaplun[30] and Proudman[26], although these authors do not detail their calculations.

This methodology is significantly different from that employed by other workers, such as Tomotika[31, 32]. Tomotika calculates C_D approximately, based on a linearized calculation of pressure. Although these approximations are consistent with the approximations inherent in their solution of the Navier-Stokes equations, they are inadequate for the purposes of obtaining a systematic approximation to any desired order of accuracy.

Calculating the drag on the body begins by determining the force exerted on the

body by the moving fluid. Using dimensional variables, the force per unit area is given by[33]:

$$P_i = -\sigma_{ik} n_k \tag{2.14}$$

Here σ_{ik} is the stress tensor, and $\vec{n}$ is a unit vector normal to the surface. For an incompressible fluid, the stress tensor takes the form[33]:

$$\sigma_{ik} = -p\delta_{ik} + \eta(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i}) \tag{2.15}$$

η is the *dynamic viscosity*, related to the kinematic viscosity by $\eta = \nu\rho$. The total force is found by integrating Eqn. 2.14 over the surface of the solid body. We now use these relations to derive explicit formula, expressed in terms of stream functions, for both the sphere and the cylinder.

Cylinder

In the case of the cylinder, the components of the velocity field are given through the definition of the Lagrangian stream function (Eqn. 2.7). Symmetry requires that the net force on the cylinder must be in the same direction as the uniform stream. Because the uniform stream is in the $\hat{x}$ direction, if follows from Eqns. 2.14 and 2.15 that the force[4] on the cylinder per unit length is given by:

$$\begin{aligned} F_{\hat{x}} &= \oint (\sigma_{rr}\cos\theta - \sigma_{r\theta}\sin\theta)\mathrm{d}s \\ &= \left[\int_0^{2\pi} (\sigma_{rr}\cos\theta - \sigma_{r\theta}\sin\theta)\, r\ \mathrm{d}\theta\right]_{r=a} \\ &= \left[\int_0^{2\pi} \left(\left(-p + 2\eta\frac{\partial v_r}{\partial r}\right)\cos\theta - \eta\left(\frac{1}{r}\frac{\partial v_r}{\partial\theta} + \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r}\right)\sin\theta\right) r\ \mathrm{d}\theta\right]_{r=a} \end{aligned} \tag{2.16}$$

The drag coefficient for an infinite cylinder is *defined* as $C_D = F_{\mathrm{Net}}/\rho|\vec{u}_\infty|^2 a$. Note that authors (e.g.,[22, 34]) who define the Reynolds number based on diameter nonetheless use the same definition of C_D, which is based on the radius. For this problem, $F_{\mathrm{Net}} = F_{\hat{x}}$, as given by Eqn. 2.16. Introducing the dimensionless variables defined in

[4]The form of σ_{ik} in cylindrical coordinates is given is Landau[33]

Table 2.2 into Eqn. 2.16, we obtain Eqn. 2.17. Combining this with the definition of C_D, we obtain Eqn. 2.18.

$$F_{\hat{x}} = \frac{\rho|\vec{u}_\infty|^2 a}{R}\left[\int_0^{2\pi}\left(\left(-p(r,\theta)+2\frac{\partial u_r}{\partial r}\right)\cos\theta - \left(\frac{1}{r}\frac{\partial u_r}{\partial \theta}+\frac{\partial u_\theta}{\partial r}-\frac{u_\theta}{r}\right)\sin\theta\right) r\ \mathrm{d}\theta\right]_{r=1} \tag{2.17}$$

$$C_D = \frac{1}{R}\left[\int_0^{2\pi}\left(\left(-p(r,\theta)+2\frac{\partial u_r}{\partial r}\right)\cos\theta - \left(\frac{1}{r}\frac{\partial u_r}{\partial \theta}+\frac{\partial u_\theta}{\partial r}-\frac{u_\theta}{r}\right)\sin\theta\right) r\ \mathrm{d}\theta\right]_{r=1} \tag{2.18}$$

To evaluate this expression, we must first derive $p(r,\theta)$ from the stream function. The pressure can be determined to within an irrelevant additive constant by integrating the $\hat{\theta}$ component of the Navier-Stokes equations (Eqn. 2.5)[13, 33]. The constant is irrelevant because, in Eqn. 2.18, $\int_0^{2\pi} C\cos\theta \mathrm{d}\theta = 0$. Note that all gradient terms involving z vanish by construction.

$$p(r,\theta) = r\int\left[-R\left((\vec{u}\cdot\nabla)\,u_\theta+\frac{u_r u_\theta}{r}\right)+\nabla^2 u_\theta+\frac{2}{r^2}\frac{\partial u_r}{\partial\theta}-\frac{u_\theta}{r^2}\right]\mathrm{d}\theta \tag{2.19}$$

Given a solution for the stream function ψ, the set of dimensionless Eqns. 2.7, 2.18, and 2.19 uniquely determine C_D for a cylinder. Because the velocity field satisfies no-slip boundary conditions, these general formula often simplify considerably.

For instance, consider the class of stream functions which meets the boundary conditions (Eqn. 2.9) and can be expressed as a Fourier sine series: $\psi(r,\theta) = \sum_{n=1}^{\infty} f_n(r)\sin n\theta$. Using the boundary conditions it can be shown that, for these stream functions, Eqn. 2.18 reduces to the simple expression given by Eqn. 2.20.

$$C_D = -\frac{\pi}{R}\left(\frac{\mathrm{d}^3}{\mathrm{d}r^3}f_1(r)\right)_{r=1} \tag{2.20}$$

Sphere

The procedure for calculating C_D in the case of the sphere is nearly identical to that for the cylinder. The components of the velocity field are given through the definition of the Stokes' stream function (Eqn. 2.10). As before, symmetry requires that any net

force on the cylinder must be in the direction of the uniform stream, in this case the $\hat{z}$ direction.

From Eqn. 2.14, the net force on the sphere is given by Eqn. 2.21.

$$\begin{aligned} F_{\hat{z}} &= \oint (\sigma_{rr}\cos\theta - \sigma_{r\theta}\sin\theta)\mathrm{d}s \\ &= 2\pi\left[\int_0^\pi (\sigma_{rr}\cos\theta - \sigma_{r\theta}\sin\theta)\, r^2 \sin\theta \ \mathrm{d}\theta\right]_{r=a} \end{aligned} \tag{2.21}$$

For the sphere, the drag coefficient is *defined* as $C_D \equiv F_{\mathrm{Net}}/\rho|\vec{u}_\infty|^2 a^2$. Often the drag coefficient is given in terms of the *Stokes' Drag*, $D_S \equiv 6\pi\rho|\vec{u}_\infty|a\nu = 6\pi\rho|\vec{u}_\infty|^2 a^2/R$. In these terms, $C_D = F_{\mathrm{Net}} 6\pi/D_S R$. If $F_{\mathrm{Net}} = D_S$, $C_D = 6\pi/R$, the famous result of Stokes[12].

Not all authors follow Stokes' original definition of C_D. For instance, S. Goldstein[27, 35] and H. Liebster[36, 37] define C_D using a factor based on cross-sectional areas: $C_D^{\mathrm{Goldstein}} = C_D 2/\pi$. These authors also define R using the diameter of the sphere rather than the radius. S. Dennis, defines C_D similarly to Goldstein, but without the factor of two: $C_D^{\mathrm{Dennis}} = C_D/\pi$[25].

Using the form of Eqn. 2.15 given in Landau[33] and introducing the dimensionless variables defined in Table 2.2 into Eqn. 2.21, we obtain Eqn. 2.22. Combining this with the definition of C_D, we obtain Eqn. 2.23.

$$F_{\hat{z}} = \frac{D_s}{3}\left[\int_0^\pi \left(\left(-p(r,\theta) + 2\frac{\partial u_r}{\partial r}\right)\cos\theta - \left(\frac{1}{r}\frac{\partial u_r}{\partial\theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}\right)\sin\theta\right) r^2\sin\theta \ \mathrm{d}\theta\right]_{r=1} \tag{2.22}$$

$$C_D = \frac{2\pi}{R}\left[\int_0^\pi \left(\left(-p(r,\theta) + 2\frac{\partial u_r}{\partial r}\right)\cos\theta - \left(\frac{1}{r}\frac{\partial u_r}{\partial\theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}\right)\sin\theta\right) r^2\sin\theta \ \mathrm{d}\theta\right]_{r=1} \tag{2.23}$$

As with the cylinder, the pressure can be determined to within an irrelevant additive constant by integrating the $\hat{\theta}$ component of the Navier-Stokes equations (Eqn. 2.5)[13, 33]. Note that gradient terms involving ϕ must vanish.

$$p(r,\theta) = r\int\left[-R\left((\vec{u}\cdot\nabla)\, u_\theta + \frac{u_r u_\theta}{r}\right) + \nabla^2 u_\theta + \frac{2}{r^2}\frac{\partial u_r}{\partial\theta} - \frac{u_\theta}{r^2\sin\theta^2}\right]\mathrm{d}\theta \tag{2.24}$$

Given a solution for the stream function ψ, the set of dimensionless Eqns. 2.10, 2.23, and 2.24 uniquely determine C_D for a sphere.

As with the cylinder, the imposition of no-slip boundary conditions considerably simplifies these general formula. In particular, consider stream functions of the form $\psi(r,\theta) = \sum_{n=1}^{\infty} f_n(r) Q_n(\cos\theta)$, where $Q_n(x)$ is defined as in Eqn. 3.20. If these stream functions satisfy the boundary conditions, the drag is given by Eqn. 2.25:

$$C_D = \frac{2\pi}{3R}\left(-2f_1''(r) + f_1'''(r)\right)_{r=1} \tag{2.25}$$

A Subtle Point

When applicable, Eqns. 2.20 and 2.25 are the most convenient way to calculate the drag given a stream function. They simply require differentiation of a single angular term's radial coefficient. However, they only apply to functions that can be expressed as a series of harmonic functions. Moreover, for these simple formula to apply, the series expansions *must* meet the boundary conditions exactly. This requirement implies that *each* of the functions $f_i(r)$ independently meets the boundary conditions.

The goal of our work is to derive and understand approximate solutions to the Navier-Stokes' equations. These approximate solutions generally will not satisfy the boundary conditions exactly. What — if any — applicability do Eqns. 2.20 and 2.25 have if the stream function does not exactly meet the boundary conditions?

In some rare cases, the stream function of interest can be expressed in a convenient closed form. In these cases, it is natural to calculate the drag coefficient using the full set of equations. However we will see that the solution for these problems is generally only expressible as a series in harmonic functions. In these cases, it actually preferable to use the simplified equations 2.20 and 2.25.

First, these equations reflect the essential symmetry of the problem, the symmetry imposed by the uniform flow. Equations 2.20 and 2.25 explicitly demonstrate that, given an exact solution, only the lowest harmonic will matter: Only terms which

have the same angular dependence as the uniform stream will contribute to the drag. By utilizing the simplified formula for C_D as opposed to the general procedure, we effectively discard contributions from higher harmonics. This is exactly what we want, since these contributions are artifacts of our approximations, and would not be present in an exact solution.

The contributions from inaccuracies in how the lowest harmonic meets the boundary conditions are more subtle. As long as the boundary conditions are satisfied to the accuracy of the overall approximation, it does not matter whether one uses the full-blown or simplified drag formula. The drag coefficients will agree to within the accuracy of the original approximation.

In general, we will use the simplified formula. This is the approach taken explicitly by many matched asymptotics workers [13, 38], and covertly by other workers[26, 29]. It should be noted that these workers only use the portion[5] of their solutions which can exactly meet the assumptions of the simplified drag formula. However, as we will subsequently discuss, this is an oversimplification.

[5]To be precise, they use only the Stokes' expansion, rather than a uniform expansion.

Chapter 3

History of Low R Flow Studies

3.1 Experiments and Numerical Calculations

Theoretical attempts to determine the drag by solving the Navier-Stokes' equations have been paralleled by an equally intricate set of experiments. In the case of the sphere, experiments usually measured the terminal velocity of small falling spheres in a homogeneous fluid. In the case of the cylinder, workers measured the force exerted on thin wires or fibers immersed in a uniformly flowing viscous fluid.

These experiments, while simple in concept, were difficult undertakings. The regime of interest necessitates some combination of small objects, slow motion, and viscous fluid. Precise measurements are not easy, and neither is insuring that the experiment actually examines the same quantities that the theory predicts. All theoretical drag coefficients concern objects in an infinite fluid, which asymptotically tends to a uniform stream. Any real drag coefficient measurements must take care to avoid affects due to the finite size of the experiment. Due to the wide variety of reported results in the literature, we found it necessary to make a complete survey, as presented in this chapter.

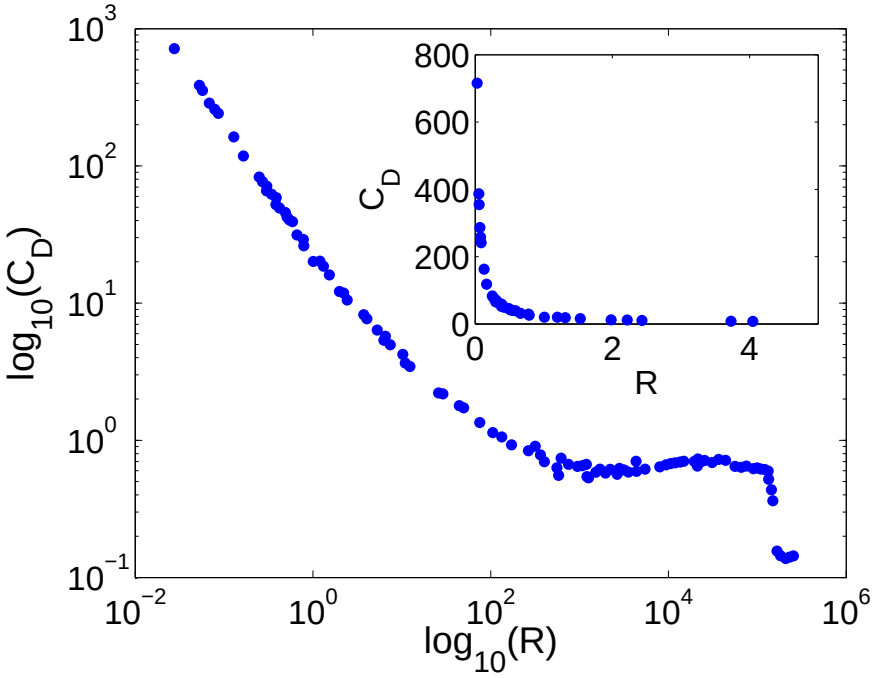

Figure 3.1: Early measurements of the drag on a sphere[35]

3.1.1 Measuring the Drag on A Sphere

As mentioned, experiments measuring the drag on a sphere at low Reynolds number were intertwined with theoretical developments. Early experiments, which essentially confirmed Stokes' law as a reasonable approximation, include those of Allen[39], Arnold[40], Williams[41], and Wieselsberger[42].

The next round of experiments were done in the 1920s, motivated by the theoretical advances begun by C. W. Oseen[32]. These experimentalists included Schmeidel[43], Liebster[36, 37]. The results of Allen, Liebster, and Arnold were analyzed, collated, and averaged by Castleman[44], whose paper is often cited as a summary of prior experiments. The state of affairs after this work is well summarized in plots given by Goldstein (p. 16)[35], and Perry[45]. Figure 3.1 shows Goldstein's plot, digitized and re-expressed in terms of the conventional definitions of C_D and R.

Figure 3.1 shows the experimental data at this point, prior to the next theoretical development, matched asymptotics. Although the experimental data seem to paint

a consistent portrait of the function $C_D(R)$, in reality they are not good enough to distinguish between different theoretical predictions.

Finite geometries cause the most significant experimental errors for these measurements[14, 23, 46]. Tritton notes that "the container diameter must be more than one hundred times the sphere diameter for the error to be less than 2 percent", and Lindgren estimates that a ratio of 50 between the container and sphere diameters will result in a 4% change in drag force.

In 1961, Fidleris et al. experimentally studied the effects of finite container size on drag coefficient measurements[47]. They concluded that there were significant finite size effects in previous experiments, but also proposed corrections to compensate for earlier experimental limitations. Lindgren also conducted some related experiments[14].

T. Maxworthy also realized this problem, and undertook experiments which could be used to evaluate the more precise predictions of matched asymptotics theories. In his own words,

> From the data plotted in Goldstein or Perry, it would appear that the presently available data is sufficient to accurately answer any reasonable question. However, when the data is plotted 'correctly'; that is, the drag is non-dimensionalized with respect to the Stokes drag, startling inaccuracies appear. It is in fact impossible to be sure of the drag to better than $\pm 20\%$ ... The difficulties faced by previous investigators seemed to be mainly due to an inability to accurately compensate for wall effects[23].

Maxworthy refined the falling sphere technique to produce the best experimental measurements yet — 2% error. He also proposed a new way of plotting the data, which removes the R^{-1} divergence in Eqn. 2.23 (as $R \to 0$). His approach makes clear the failings of earlier measurements, as can be seen in Figure 3.2, where the drag measurements are normalized by the Stokes drag, $C_D^{\text{Stokes}} = 6\pi/R$.

In Maxworthy's apparatus, the container diameter is over 700 times the sphere di-

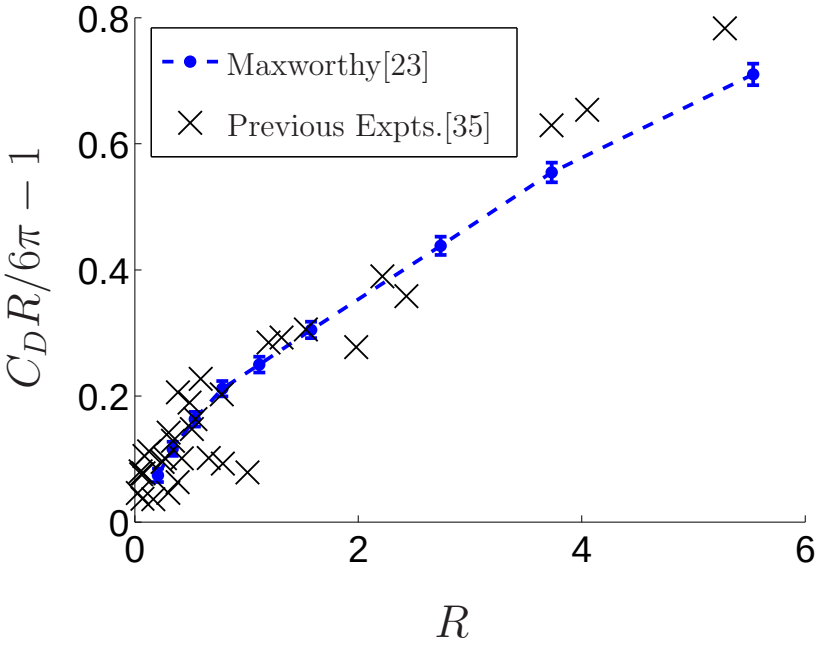

Figure 3.2: Maxworthy's accurate measurements of the drag on a sphere [23].

ameter, and does not contribute significantly to experimental error, which he estimates at better than 2 percent. Note that the data in Figure 3.2 are digitized from his paper, as raw data are not available.

This problem also attracted the attention of atmospheric scientists, who realized its significance in cloud physics, where "cloud drops may well be approximated by rigid spheres[48]." In a series of papers (e.g., [24, 48, 49, 50]), H.R. Pruppacher and others undertook numerical and experimental studies of the drag on the sphere. They were motivated by many of the same reasons as Maxworthy, because his experiments covered only Reynolds numbers between 0.4 and 11, and because "Maxworthy's experimental setup and procedure left considerable room for improvement"[49].

Their results included over 220 measurements, which they binned and averaged. They presented their results in the form of a set of linear fits. Adopting Maxworthy's

normalization, we collate and summarize their findings in Eqn. 3.1.

$$C_D \frac{R}{6\pi} - 1 = \begin{cases} 0.102\,(2R)^{0.955} & 0.005 < R \leq 1.0 \\ 0.115\,(2R)^{0.802} & 1.0 < R \leq 20 \\ 0.189\,(2R)^{0.632} & 20 < R \leq 200 \end{cases} \tag{3.1}$$

Unfortunately, one of their later papers includes the following footnote (in our notation): "At $R < 1$ the most recent values of $C_D R/6\pi - 1$ (Pruppacher, 1969, unpublished) tended to be somewhat higher than those of Pruppacher and Steinberger."[24] Their subsequent papers plot these unpublished data as "experimental scatter." As the unpublished data are in much better agreement with both Maxworthy's measurements and their own numerical analysis[24], it makes us question the validity of the results given in Eqn. 3.1.

There are many numerical calculations of the drag coefficient for a sphere, including: Dennis[25], Le Clair [24, 48], Hamielec [51], Rimon [52], Jenson [53], and Kawaguti [54]. Most of these results are not useful either because of large errors (e.g., Jenson), or because they study ranges of Reynolds number which do not include $R < 1$. Many numerical studies examine only a few (or even just a single) Reynolds numbers. For the purposes of comparing theoretical predictions of C_D at low Reynolds number, only Dennis [25] and Le Clair [24] have useful calculations. Both of these papers report tabulated results which are in very good agreement with both each other and Maxworthy; at $R = 0.5$, the three sets of results agree to within 1% in C_D, and to within 10% in the transformed variable, $C_D R/6\pi - 1$. The agreement is even better for $R < 0.5$.

Figure 3.3 shows all relevant experimental and numerical results for the drag on a sphere. Note the clear disagreement between Pruppacher's results (Eqn. 3.1), and all of the other results for $R < 1$ — including Le Clair and Pruppacher's numerical results[24]. This can be clearly seen in the inset graph. Although Pruppacher's experiment results do agree very well with other data for larger values of R ($R \gtrsim 20$), we will disregard

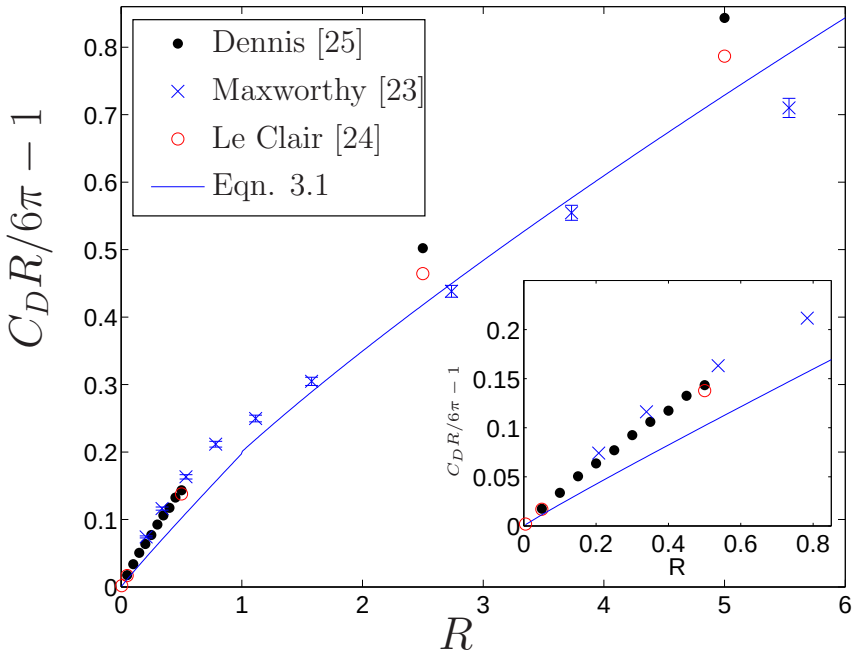

Figure 3.3: A summary of experimental and numerical studies of C_D for a sphere

them for the purposes of evaluating theoretical predictions at low Reynolds number.

It should also be noted that there is a community of researchers interested in sedimentation and settling velocities who have studied the drag on a sphere. In a contribution to this literature, Brown reviews all of the authors discussed here, in an effort to tabulate C_D for $R < 5000$[55]. His report addresses a larger range of Reynolds numbers and he summarizes a number of experiments not treated here. His methodology is to apply the Fidleris' correction[47] to previous experiments where tabulated experimental data was published[1]. While this yields a reasonably well-behaved drag coefficient for a wide range of Reynolds numbers, it is not particularly useful for our purposes, as less accurate work obfuscates the results of the most precise experiments near $R = 0$. It also does not include numerical work or important results which are only available graphically (e.g., Maxworthy[23]).

[1]Brown erroneously reports Dennis' work[25] as experimental.

3.1.2 Measuring the Drag on a Cylinder

Experiments designed to measure the drag on an infinite cylinder in a uniform fluid came later than those for spheres. In addition to being a more difficult experiment — theoretical calculations assume the cylinder is infinite — there were no theoretical predictions to test before Lamb's result in 1911[56].

In 1914, E. F. Relf conducted the first experiments[57]. These looked at the force exerted on long wires in a fluid. Relf measured the drag down to a Reynolds number of about ten. In 1921, Wieselberger measured the drag at still lower Reynods number, reaching $R = 2.11$ by looking at the deflection of a weight suspended on a wire in an air stream[58].

These experiments, combined with others[35, 59] at higher Reynolds number, characterize the drag over a range of Reynolds numbers (see Goldstein, pg. 15). However, they do not probe truly small Reynolds numbers ($R \ll 1$), and are of little use for evaluating theories which are only valid in that range. Curiously, there are no shortage of claims otherwise, such as Lamb, who says "The formula is stated to be in good agreement with experiment for sufficiently small values of $U_\infty a/\nu$; see Wieselsberger"[60].

In 1933, Thom measured the "pressure drag", extending observations down to $R = 1.75$. Thom also notes that this Reynolds number is still too high to compare with calculations: "Actually, Lamb's solution only applies to values of R less than those shown, in fact to values much less than unity, but evidently in most cases the experimental results are converging with them."[61]

In 1946, White undertook a series of measurements, which were flawed due to wall effects[62]. The first high quality experiments which measured the drag at low Reynolds number were done by R. K. Finn[63]. His results, available only in graphical form, are reproduced Figure 3.5. While vastly superior to any previous results, there is considerable scatter in Finn's measurements, and they have largely been surpassed by later experiments.

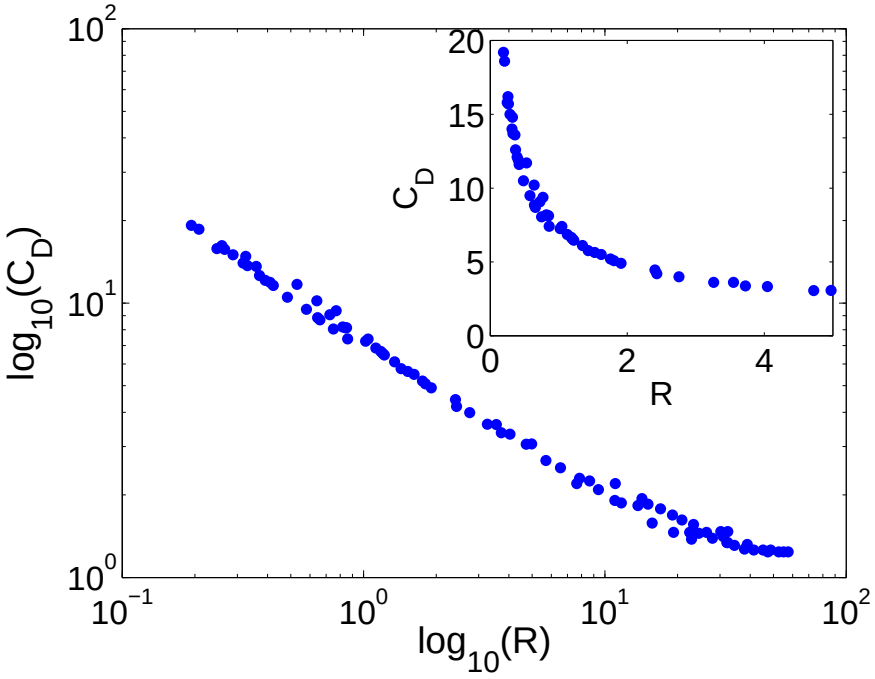

Figure 3.4: Tritton's measurements of the drag on a cylinder[22]

Tritton, in 1959, conducted experiments which reached a Reynolds number of $R = 0.2$, and also filled in some gaps in the $R - C_D$ curve[22]. Tritton estimates his accuracy at $\pm 6\%$, and compares his results favorably to previous work, commenting that, "Probably the lowest R points of the other workers were stretching their techniques a little beyond their limits." Tritton is also the first author to give a discussion of systematic errors.[2] Tritton's results are shown in Figure 3.4. All of his data are available in tabular form.

Maxworthy improved plots of the drag on a sphere (Fig. 3.1), by arguing that the leading divergence must be removed to better compare experiments and predictions (Fig. 3.2). This same criticism applies to plots of the drag on a cylinder. In the case of the cylinder, C_D goes as R^{-1} (with logarithmic corrections) as $R \to 0$ (Eqn. 2.18). This means we ought to plot $C_D R/4\pi$. This function tends to zero as $R \to 0$, so it is

[2]Tritton does caution that his measurements may be negatively biased at higher Reynolds number ($R \gtrsim 30$)

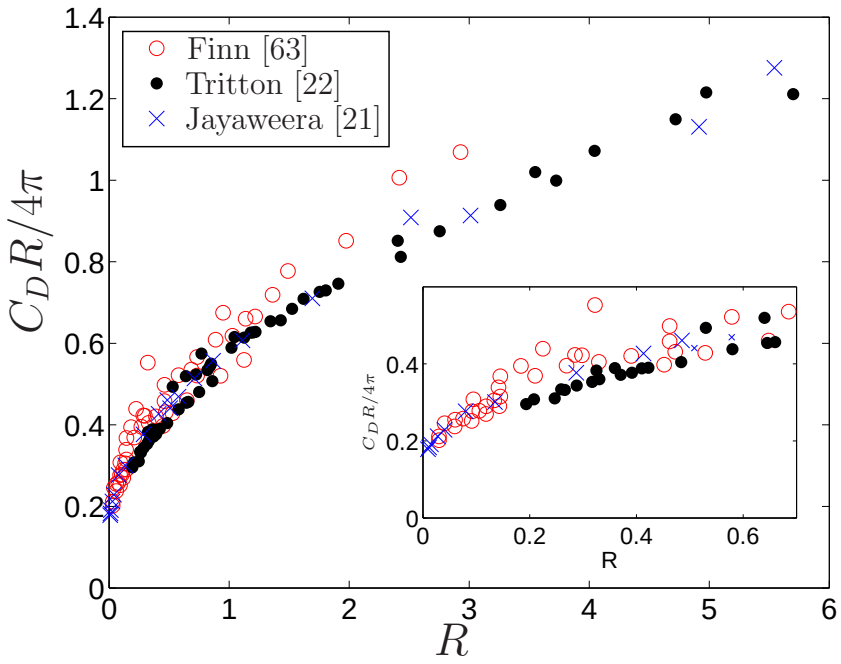

Figure 3.5: Summary of measurements of the drag on a cylinder

not necessary to plot $C_D R/4\pi - 1$, as in the case of the sphere. Figure 3.5 shows both Finn's and Tritton's data re-plotted with the leading divergence removed.

In 1965, K. O. L. F. Jayaweera [21] undertook drag measurements of the drag on very long (but finite) cylinders. At very low Reynolds number ($R \leq 0.135$), his data are available in tabular form. At higher Reynolds number, they had to be digitized. His data, plotted with the leading divergence removed, are also shown in Figure 3.5.

The agreement amongst these experiments is excellent. Henceforth, Finn's data will not be plotted, as it exhibits larger experimental variations, and is surpassed by the experiments of Jayaweera and Tritton. Jayaweera's data exhibit the least scatter, and may be slightly better than Tritton's. However, both experiments have comparable, large ratios of cylinder length to width (the principle source of experimental error), and there is no a priori reason to favor one experimental design over the other. We consider these two experiments to be equivalent for the purposes of evaluating theoretical predictions.

As with the sphere, there are numerical calculations, including: Underwood[64], Son[65], Kawaguti[66], Dennis[67], Thom[61], Apelt[68], and Allen[69]. Of these, most treat only a few Reynolds numbers, none of which are sufficiently small. Others, such as Allen and Dennis, have had their results subsequently questioned[64]. The only applicable studies are Kawaguti[66], and Underwood[64]. Kawaguti has a calculation only for $R = 0.5$, and is omitted. Underwood's results are in principle important and useful, but are only available in a coarse plot, which cannot be digitized with sufficient accuracy. Consequently, no numerical results will be used for evaluating analytical predictions.

There are many different experimental and numerical drag coefficient measurements. We will subsequently use only the most best as benchmarks for evaluating the performance of theoretical predictions. In the case of the sphere, the experimental measurements of Maxworthy[23] as well as the numerical calculations of Dennis[25] and Le Clair[24] all extend to sufficiently small R and possess sufficient accuracy. For the cylinder the experiments of both Tritton[22] and Jayaweera[21] are both excellent. Although they exhibit small differences, we cannot judge either to be superior, and we will compare both with theoretical results.

3.2 Theoretical History

Since these problems were posed by Stokes in 1851, there have been many attempts to solve them. All of these methods involve approximations, which are not always rigorous (or even explicitly stated). There is also considerable historical confusion over contributions and attribution.[3] Here we review and summarize the substantial contributions to the literature, focusing on what approximations are used, in both deriving governing equations and in their subsequent solution. We discuss the validity

[3]For an explanation of confusion over early work, see Lindgren[14]. Proudman and Pearson[26] also begin their article with an insightful, nuanced discussion, although there are some errors[14]

and utility of important results. Finally, we emphasize methodological shortcomings and how they have been surmounted.

3.2.1 Stokes and Paradoxes

In the first paper on the subject, Stokes approximated $R = 0$ in Eqn. 2.5 and solved the resulting equation (a problem equivalent to solving Eqn. 2.11 with $R = 0$)[12]. After applying the boundary conditions (Eqn. 2.13), his solution is given in terms of a stream function by Eqn. 3.2.

$$\psi(r, \mu) = \frac{1}{4}\left(2r^2 - 3r + \frac{1}{r}\right)\left(1 - \mu^2\right) \tag{3.2}$$

By substituting $\psi(r, \mu)$ into Eqns. 2.10, 2.23, and 2.24 (or by using Eqn. 2.25), we reproduce the famous result of Stokes, given by Eqn. 3.3.

$$C_D = \frac{6\pi}{R} \tag{3.3}$$

Stokes also tackled the two dimensional cylinder problem in a similar fashion, but could not obtain a solution. The reason for his failure can be by setting $R = 0$ in Eqn. 2.8, and attempting a direct solution. Enforcing the $\sin\theta$ angular dependence results in a solution of the form $\psi(r, \theta) = (C_1 r^3 + C_2 r \ln r + C_3 r + C_4/r)\sin\theta$. Here C_i are integration constants. No choice of C_i will meet the boundary conditions Eqn. (2.9), as this solution cannot match the uniform flow at large r. The best one can do is to set $C_1 = 0$, resulting in a partial solution:

$$\psi(r, \theta) = C\left(2r \ln r - r + \frac{1}{r}\right)\sin\theta \tag{3.4}$$

Nonetheless, this solution is *not* a description of fluid flow which is valid everywhere. Moreover, due to the indeterminable constant C, Eqn. 3.4 cannot be used to estimate the drag on the cylinder.

A more elegant way to see that no solution may exist is through dimensional analysis[33, 70]. The force per unit length may only depend on the cylinder radius,

Quantity	Description	Dimensions
F_{Net}	Net Force per Unit Length	MT^{-2}
ν	Kinematic Viscosity	L^2T^{-1}
a	Cylinder Radius	L
ρ	Fluid Density	ML^{-3}
$\lvert\vec{u}_\infty\rvert$	The Uniform Stream Speed	LT^{-1}

Table 3.1: Dimensional Analysis of Stokes' Problem

fluid viscosity, fluid density, and uniform stream velocity. These quantities are given in Table 3.1, with M denoting a unit of mass, T a unit of time, and L a unit of length. From these quantities, one may form two dimensionless groups[71]: $\Pi_0 = R = |\vec{u}_\infty|a/\nu$, $\Pi_1 = F_{\text{Net}}/(\rho\nu|\vec{u}_\infty|)$. Buckingham's Π Theorem[71] then tells us that:

$$\Pi_0 = F(R) \tag{3.5}$$

If we make the assumption that the problem does not depend on R, as Stokes did, then we obtain $\Pi_1 = \text{const}$, whence

$$F_{\text{Net}} \propto \rho\nu|\vec{u}_\infty| \tag{3.6}$$

However, Eqn. 3.6 does not depend on the cylinder radius, a! This is physically absurd, and demonstrates that Stokes' assumptions cannot yield a solution. The explanation is that when we take the $R \to 0$ limit in Eqn. 3.5, we made the incorrect assumption that $F(R)$ tended toward a *finite, non-zero limit*. This is an example of *incomplete similarity*, or *similarity of the second kind* (in the Reynolds number)[72]. Note that the problem of flow past a sphere involves force, *not* force per unit length, and therefore is not subject to the same analysis.

Stokes incorrectly took this nonexistence of a solution to mean that steady-state flow past an infinite cylinder could not exist. This problem, which is known as *Stokes' paradox*, has been shown to occur with any unbounded two-dimensional flow[73]. But it is a mathematical problem, not a physical one, and it has since been resolved.

In 1888, Whitehead, attempted to find higher approximations for flow past a sphere, ones which would be valid for small but non-negligible Reynolds numbers[74]. He used Stokes' solution (Eqn. 3.2) to approximate viscous contributions (the LHS of Eqn. 2.11), aiming to iteratively obtain higher approximations for the inertial terms. In principle, this approach can be repeated indefinitely, always using a linear governing equation to obtain higher order approximations. Unfortunately, Whitehead found that his next order solution could not meet all of the boundary conditions (Eqn. 2.13), because he could not match the uniform stream at infinity[29]. These difficulties are analogous to the problems encountered in Stokes' analysis of the infinite cylinder.

Whitehead's approach is equivalent to a perturbative expansion in the Reynolds number, an approach which is "never valid in problems of uniform streaming"[26]. This mathematical difficulty is common to all three-dimensional uniform flow problems, and is known as *Whitehead's paradox*. Whitehead thought this was due to discontinuities in the flow field (a "dead-water wake"), but this is incorrect, and his "paradox" has also since been resolved[29].

3.2.2 Oseen's Equation

Introduction

In 1893, Rayleigh pointed out that Stokes' solution would be uniformly applicable if certain inertial forces were included, and noted that the ratio of those inertial forces to the viscous forces which Stokes considered could be used to estimate the accuracy of Stokes' approximations[75].

Building on these ideas in 1910, C. W. Oseen proposed an ad hoc approximation to the Navier-Stokes equations which resolved both paradoxes. His linearized equations (the *Oseen equations*) attempted to deal with the fact that the equations governing Stokes' perturbative expansion are invalid at large $|\vec{r}|$, where they neglect important inertial terms. In addition to Oseen, a number of workers, have applied his equations

to a wide variety of problems, including both the cylinder and the sphere.[4]

Oseen's governing equation arises independently in several different contexts. Oseen derived the equation in an attempt to obtain an approximate equation which describes the flow everywhere. In modern terminology, he sought a governing equation whose solution is a uniformly valid approximation to the Navier-Stokes equations. Whether he succeeded is a matter of some debate. The short answer is "Yes, but he got lucky."

This story is further complicated by historical confusion. Oseen's equations "are valid but for the wrong reason"[14]; Oseen originally objected to working in the inertial frame where the solid body is at rest, and therefore undertook calculations in the rest frame of uniform stream. This complication is overlooked largely because many subsequent workers have only understood Oseen's intricate three paper analysis through the lens of Lamb's later work[56]. Lamb — in addition to writing in English — presents a clearer, "shorter way of arriving at his [Oseen's] results", which he characterizes as "somewhat long and intricate."[56]

In 1913 Fritz Noether, using both Rayleigh's and Oseen's ideas, analyzed the problem using stream functions[77]. Noether's paper prompted criticisms from Oseen, who then revisited his own work. A few months later, Oseen published another paper, which included a new result for C_D (Eqn. 3.13)[78]. Burgess also explains the development of Oseen's equation, and presents a clear derivation of Oseen's principal results, particularly of Oseen's new formula for C_D[79].

Lindgren offers a detailed discussion of these historical developments[14]. However, he incorrectly reports Noether's publication date as 1911, rather than 1913. As a result, he incorrectly concludes that Noether's work was independent of Oseen's, and contradicts claims made in Burgess[79].

Although the theoretical justification for Oseen's approximations is tenuous, its

[4]Lamb[60] solved the Oseen equations for the cylinder approximately, as Oseen[32] did for the sphere. The cylinder equations have been solved exactly by Faxén[76], as well as by Tomotika and Aoi[31], and the sphere was solved exactly by Goldstein[27].

success at resolving the paradoxes of both Stokes and Whitehead led to widespread use. Oseen's equation has been fruitfully substituted for the Navier-Stokes' equations in a broad array of low Reynolds number problems. Happel and Brenner describe its application to many problems in the dynamics of small particles where interactions can be neglected[70]. Many workers have tried to explain the utility and unexpected accuracy of Oseen's governing equations.

Finally, the Oseen equation, as a partial differential equation, arises in both matched asymptotic calculations and in our new work. In these cases, however, its genesis and interpretation is entirely different, and the similarity is purely formal. Due to its ubiquity and historical significance, we now discuss both Oseen's equation and its *many* different solutions in detail.

Why Stokes' Approximation Breaks Down

Oseen solved the paradoxes of Stokes and Whitehead by using Rayleigh's insight: compare the magnitude of inertial and viscous forces[32, 75]. Stokes and Whitehead had completely negelected inertial terms in the Navier-Stokes equations, working in the regime where the Reynolds number is insignificantly small (so-called "creeping flow"). However, this assumption can only be valid near the surface of the fixed body. It is *never* valid everywhere.

To explain why, we follow here the spirit of Lamb's analysis, presenting Oseen's conclusions "under a slightly different form."[56]

Consider first the case of the sphere. We can estimate the magnitude of the neglected inertial terms by using Stokes' solution (Eqn. 3.2). Substituting this result into the RHS of Eqn. 2.11, we see that the dominant inertial components are convective accelerations arising from the nonlinear terms in Eqn. 2.11. These terms reflect interactions between the uniform stream and the perturbations described by Eqn. 3.2. For large values of $|\vec{r}|$, these terms are of $\mathcal{O}\left(Rr^{-2}\right)$.

Estimating the magnitude of the relevant viscous forces is somewhat trickier. If we substitute Eqn. 3.2 into the LHS of Eqn. 2.11, the LHS vanishes identically. To learn anything, we must consider the terms individually. There are two kinds of terms which arise far from the sphere. Firstly, there are components due solely to the uniform stream. These are of $\mathcal{O}\left(r^{-2}\right)$. However, the uniform stream satisfies Eqn. 2.11 independently, without the new contributions in Stokes' solution. Mathematically, this means that all of the terms of $\mathcal{O}\left(r^{-2}\right)$ necessarily cancel amongst themselves.[5] We are interested in the magnitude of the remaining terms, perturbations which result from the other components of Stokes' solution. These viscous terms (i.e. the ∂_θ^4 term in Eqn. 2.11) are of $\mathcal{O}\left(r^{-3}\right)$ as $r \to \infty$.

Combining these two results, the ratio of inertial to viscous terms, in the $r \to \infty$ limit, is given by Eqn. 3.7.

$$\frac{\text{inertial}}{\text{viscous}} = \mathcal{O}\left(Rr\right) \tag{3.7}$$

This ratio is small near the body (r is small) and justifies neglecting inertial terms in that regime. However, Stokes' implicit assumption that inertial terms are everywhere small compared to viscous terms breaks down when $Rr \sim \mathcal{O}\left(1\right)$, and the two kinds of forces are of the same magnitude. In this regime, Stokes' solution is not valid, and therefore cannot be used to estimate the inertial terms (as Whitehead had done). Technically speaking, Stokes' approximations breaks down because of a singularity at infinity, an indication that this is a *singular perturbation* in the Reynolds' number. As Oseen pointed out, this is the genesis of Whitehead's "paradox".

What does this analysis tell us about the utility of Stokes' solution? Different opinions can be found in the literature. Happel, for instance, claims that it "is not uniformly valid"[70], while Proudman asserts "Stokes' solution is therefore actually a uniform approximation to the total velocity distribution."[26] By a *uniform approxi-*

[5] VanDyke[29] is misleadingly cavalier in his treatment of this issue, and we recommend Proudman[26] or Happel[70] for a more careful discussion.

mation, we mean that the approximation asymptotically approaches the exact solution as the Reynolds' number goes to zero[80]; see Section 3.3 for further discussion.

Proudman and Pearson clarify their comment by noting that although Stokes' solution is a uniform approximation to total velocity distribution, it does not adequately characterize the perturbation to the uniform stream, or the *derivatives* of the velocity. This is a salient point, for the calculations leading to Eqn. 3.7 examine components of the Navier-Stokes equations, not the velocity field itself. These components are forces — derivatives of velocity.

However, Proudman and Pearson offer no proof that Stokes' solution is actually a uniform approximation, and their claim that it is "a valid approximation to many bulk properties of the flow, such as the resistance"[26] goes unsupported. In fact any calculation of the drag necessitates utilizing derivatives of the velocity field, so their analysis is inconsistent.

We are forced to conclude that Stokes' solution is not a uniformly valid approximation, and that his celebrated result, Eqn. 3.3, is the fortuitous result of uncontrolled approximations. Remarkably, Stokes' drag formula is in fact a correct zeroth order approximation, as can be shown using either matched asymptotics or the Oseen equation! This coincidence is essentially due to the fact that the drag is determined by the velocity field and its derivatives at the surface of the sphere, where $r = 1$, and Eqn. 3.7 is $\mathcal{O}(R^1)$. The drag coefficient calculation uses Stokes' solution in the regime where his assumptions are the most valid.

A similar analysis affords insight into the origin of Stokes' paradox in the problem of the cylinder. Although we have seen previously that Stokes' approach must fail for both algebraic and dimensional considerations, examining the ratio between inertial and viscous forces highlights the physical inconsistencies in Stokes' assumptions.

We can use the incomplete solution given by Eqn. 3.4 to estimate the relative contributions of inertial and viscous forces in Eqn. 2.8. More specifically, we examine

the behavior of these forces at large values of r. Substituting Eqn. 3.4 into the RHS of Eqn. 2.8, we find that the inertial forces are $\mathcal{O}\left(RC^2 \log r/r^2\right)$ as $r \to \infty$.

We estimate the viscous forces as in the case of the sphere, again ignoring contributions due solely to the uniform stream. The result is that the viscous forces are $\mathcal{O}\left(C \log r/r^3\right)$.[6] Combining the two estimates, we obtain the result given in Eqn. 3.8.

$$\frac{\text{inertial}}{\text{viscous}} = \mathcal{O}\left(Rr\right) \tag{3.8}$$

This result demonstrates that the paradoxes of Stokes' and Whitehead are the result of the same failures in Stokes' uncontrolled approximation. Far from the solid body, there is a regime where it is incorrect to assume that the inertial terms are negligible in comparison to viscous terms. Although these approximations happened to lead to a solution in the case of the sphere, Stokes' approach is invalid and technically inconsistent in both problems.

How Oseen Resolved the Paradoxes

Not only did Oseen identify the physical origin for the breakdowns in previous approximations, but he also discovered a solution[32]. As explained above, the problems arise far from the solid body, when inertial terms are no longer negligible. However, in this region ($r \gg 1$), the flow field is nearly a uniform stream — it is almost unperturbed by the solid body. Oseen's inspiration was to replace the inertial terms with linearized approximations far from the body. Mathematically, the fluid velocity $\vec{u}$ in Eqn. 2.5 is replaced by the quantity $\vec{u}_\infty + \vec{u}$, where $\vec{u}$ represents the perturbation to the uniform stream, and is considered to be small. Neglecting terms of $\mathcal{O}\left(|\vec{u}|^2\right)$, the viscous forces of the Navier-Stokes' equation — $R\left(\vec{u} \cdot \nabla \vec{u}\right)$ — are approximated by $R\left(\vec{u}_\infty \cdot \nabla \vec{u}\right)$.

This results in Oseen's equation:

$$R\left(\vec{u}_\infty \cdot \nabla \vec{u}\right) = -\nabla p + \nabla^2 \vec{u} \tag{3.9}$$

[6] This result disagrees with the results of Proudman[26] and VanDyke[29].

The lefthand side of this equation is negligible in the region where Stokes' solution applies. One way to see this is by explicitly substituting Eqn. 3.2 or Eqn. 3.4 into the LHS of Eqn. 3.9. The result is of $\mathcal{O}(R)$. This can also be done self-consistently with any of the solutions of Eqn. 3.9; it can thereby be explicitly shown that the LHS can only becomes important when $r \gg 1$, and the ratios in Eqns. 3.7 and 3.8 are of $\mathcal{O}(1)$.

Coupled with the continuity equation (Eqn. 2.4), and the usual boundary conditions, the Oseen equation determines the flow field everywhere. The beautiful thing about Oseen's equation is that it is *linear*, and consequently is solvable in a wide range of geometries. In terms of stream functions, the Oseen equation for a sphere takes on the form given by Eqn. 3.10. The boundary conditions for this equation are still given by Eqn. 2.13.

$$D^4\psi = R\left(\frac{1-\mu^2}{r}\frac{\partial}{\partial\mu} + \mu\frac{\partial}{\partial r}\right)D^2\psi(r,\mu) \tag{3.10}$$

Here, D is defined as in Eqn. 2.11.

For the cylinder, where the boundary conditions are given by Eqn. 2.9, Oseen's equation takes the form given by Eqn. 3.11.

$$\nabla_r^4\psi(r,\theta) = R\left(\cos(\theta)\frac{\partial}{\partial r} - \frac{\sin(\theta)}{r}\frac{\partial}{\partial\theta}\right)\nabla_r^2\psi(r,\theta) \tag{3.11}$$

Here ∇ is defined as in Eqn. 2.8. This equation takes on a particularly simple form in Cartesian coordinates (where $x = r\cos\theta$): $\left(\nabla^2 - R\partial_x\right)\nabla^2\psi(r,\theta) = 0$.

A few historical remarks must be made. First, Oseen and Noether were motivated to refine Stokes' work and include inertial terms because they objected to the analysis being done in the rest frame of the solid body. While their conclusions are valid, there is nothing wrong with solving the problem in any inertial frame. Secondly, Oseen made no use of stream functions; the above equations summarize results from several workers, particularly Lamb.

There are many solutions to Oseen's equations, applying to different geometries and configurations, including some exact solutions. For any useful calculations, such as C_D,

even the exact solutions need to be compromised with approximations. There have been many years of discussion about how to properly interpret Oseen's approximations, and how to understand the limitations of both his approach and concomitant solutions. Before embarking on this analysis, we summarize the important solutions to Eqns. 3.10 and 3.11.

A Plethora of Solutions

Oseen himself provided the first solution to Eqn. 3.10, solving it exactly for flow past a sphere[32]. Eqn. 3.12 reproduces this result in terms of stream functions, a formula first given by Lamb[60].

$$\psi(r,\theta) = \frac{1}{4}\left(2r^2 + \frac{1}{r}\right)\sin^2\theta - \frac{3}{2R}\left(1+\cos\theta\right)\left(1 - e^{-\frac{1}{2}Rr(1-\cos\theta)}\right) \tag{3.12}$$

This solution is reasonably behaved everywhere, and maybe used to obtain Oseen's improved approximation for the drag coefficient (Eqn. 3.13).

$$C_D = \frac{6\pi}{R}\left(1 + \frac{3}{8}R\right) + \mathcal{O}\left(R^2\right) \tag{3.13}$$

Oseen obtained this prediction for C_D after the prompting of Noether, and only presented it in a later paper[78]. Burgess also obtained this result[79]. Oseen's work was hailed as a resolution to Whitehead's paradox. While it *did* resolve the paradoxes (e.g., he explained how to deal with inertial terms), and his solution is uniformly valid, it does *not* posses sufficient accuracy to justify the "$3/8R$" term in Eqn. 3.13. What Oseen really did was to rigorously derive the leading order term, proving the validity of Stokes' result (Eqn. 3.3). Remarkably, his new term is also correct! This is a coincidence which will be carefully considered later.

This solution (Eqn. 3.12) is exact in the sense that it satisfies Eqn. 3.10. However, it does not exactly meet the boundary conditions (Eqn. 2.13) at the surface of the sphere. It satisfies those requirements only approximately, to $\mathcal{O}\left(R^1\right)$. This can readily

be seen by expanding Eqn. 3.12 about $r = 1$:

$$\psi(r,\theta) = \frac{1}{4}\left(2r^2 - 3r + \frac{1}{r}\right)\sin^2\theta + \mathcal{O}\left(R^1\right) \tag{3.14}$$

Up to $\mathcal{O}(R)$ this is simply Stokes' solution (Eqn. 3.2), which vanishes identically at $r = 1$. The new terms fail to satisfy the boundary conditions at the surface, but are higher order in R. Thus Oseen's solution is an exact solution to an approximate governing equation which satisfies boundary conditions approximately. The implications of this confounding hierarchy of approximations will be discussed below.

Lamb contributed a simplified method for both deriving and solving Oseen's equation[56]. His formulation was fruitfully used by later workers (e.g., [27, 31, 76]), and Lamb himself used it both to reproduce Oseen's results and to obtain the first result for the drag on an infinite cylinder.

Lamb's basic solution for flow around an infinite cylinder appears in a number of guises. His original solution was given in terms of velocity components, and relied on expansions of modified Bessel functions which kept only the most important terms in the series. This truncation results in a solution (Eqn. 3.15) which only approximately satisfies the governing equations (Eqn. 3.11), and is only valid near the surface.

$$u_x = 1 + \delta\left(\gamma - \frac{1}{2} + \log\frac{rR}{4} + \frac{1}{2}\left(r^2 - 1\right)\frac{\partial^2}{\partial x^2}\log r\right) \tag{3.15a}$$

$$u_y = \frac{\delta}{2}\left(r^2 - 1\right)\frac{\partial^2}{\partial x \partial y}\log r \tag{3.15b}$$

$$u_z = 0 \tag{3.15c}$$

In this equation, $\delta = \left(\frac{1}{2} - \gamma - \log\frac{R}{4}\right)^{-1}$.

Note that, although it only approximately satisfies Oseen's governing equation, that this result satisfies the boundary conditions (Eqn. 2.3) exactly. Lamb used his solution to derive the first result (Eqn. 3.16) for the drag on an infinite cylinder, ending Stokes' paradox. In his own words, " ... Stokes was led to the conclusion that steady motion is impossible. It will appear that when the inertia terms are partially taken into account

... that a definite value for the resistance is obtained."[56]

$$C_D = \frac{4\pi}{R}(\delta) \tag{3.16}$$

As with all analysis based on the ad-hoc Oseen equation, it is difficult to quantify either the accuracy or the limitations of Lamb's result.

Many authors formulate alternate expressions of Lamb's solution by retaining the modified Bessel functions rather than replacing them with expansions valid for small R and r. This form is given by Eqn. 3.17, and is related to the incomplete form given by VanDyke (p. 162)[29].[7]

$$u_x = 1 + \delta\left(\frac{x^2}{r^4} - \frac{1}{2r^2} + \frac{2x}{Rr^2} - e^{Rx/2}K_0\left(\frac{Rr}{2}\right) - \frac{x}{r}e^{Rx/2}K_1\left(\frac{Rr}{2}\right)\right) \tag{3.17a}$$

$$u_y = \delta\left(\frac{xy}{r^4} + \frac{2y}{Rr^2} - \frac{y}{r}e^{Rx/2}K_1\left(\frac{Rr}{2}\right)\right) \tag{3.17b}$$

$$u_z = 0 \tag{3.17c}$$

Here I_n and K_n are modified Bessel functions.

In contrast to Eqn. 3.15, this solution is an exact solution to Oseen's equation (Eqn. 3.11), but only meets the boundary conditions to first approximation. In particular, it breaks down for harmonics other than $\sin\theta$. Whether Eqn. 3.15 or Eqn. 3.17 is preferred is a matter of some debate, and ultimately depends on the problem one is trying to solve.

Some workers prefer expressions like Eqn. 3.17, which are written in terms of $\vec{u}$. Unlike the solutions for the stream function, these results can be written in closed form. This motivation is somewhat misguided, as applying the boundary conditions nonetheless requires a series expansion.

In terms of stream functions Eqn. 3.17 transforms into Eqn. 3.18[26].

$$\psi(r,\theta) = \left(r + \frac{\delta}{2r}\right)\sin\theta - \sum_{n=1}^{\infty}\delta\phi_n\left(\frac{Rr}{2}\right)\frac{r\sin n\theta}{n} \tag{3.18}$$

[7]Note that VanDyke improperly attributes to this result to Oseen, rather than to Lamb.

Here,

$$\phi_n(x) = 2K_1(x)I_n(x) + K_0(x)\left(I_{n+1}(x) + I_{n-1}(x)\right)$$

This result is most easily derived as a special case of Tomotika's general solution (Eqn. 3.23)[31], although Proudman et al. intimate that it can also be directly derived from Lamb's solution (Eqn. 3.17)[26].

Bairstow et al. were the first to retain Bessel functions while solving Oseen's Eqn. for flow past a cylinder[81]. They followed Lamb's approach, but endeavored to extend it to larger Reynolds' numbers, and obtained the drag coefficient given in Eqn. 3.19. When expanded near $R = 0$, this solution reproduces Lamb's result for C_D (Eqn. 3.16). It can also be obtained from Tomotika's more general solution (Eqn. 3.23).

$$C_D = \frac{4\pi}{R\left(I_0(R/2)K_0(R/2) + I_1(R/2)K_1(R/2)\right)} \tag{3.19}$$

Bairstow also made extensive comparisons between experimental measurements of C_D and theoretical predictions[57]. He concluded, "For the moment it would appear that the maximum use has been made of Oseen's approximation to the equations of viscous fluid motion."

At this point, the "paradoxes" were "resolved" but by an approximate governing equation which had been solved approximately. This unsatisfactory state of affairs was summarized by Lamb in the last edition of his book: " ... even if we accept the equations as adequate the boundary-conditions have only been approximately satisfied."[60] His comment was prompted largely by the work of Hilding Faxén, who initiated the next theoretical development, exact solutions to Oseen's approximate governing equation (Eqn. 3.9) which also exactly satisfy the boundary conditions.

Beginning with his thesis, and spanning a number of papers Faxén systematically investigated the application of boundary conditions to solutions of Oseen's equations[82, 83]. Faxén initially studied low Reynolds number flow around a sphere, and he began by re-examining Oseen's analysis. He derived a formula for C_D which differed from

Oseen's accepted result (Eqn. 3.13). Faxén realized that this was due to differences in the application of approximate boundary conditions; within the limitations of their respective analyses, the results actually agreed.

Faxén next solved Oseen's equation (Eqn. 3.10), but in bounded, finite spaces where the boundary conditions could be satisfied exactly. He initially studied flow near infinite parallel planes, but ultimately focused on flow around a sphere within a cylinder of finite radius. He aimed to calculate the drag force in a finite geometry, and then take the limit of that solution as the radius of the cylinder tends to infinity.

Unfortunately, in the low Reynolds number limit, the problem involves incomplete similarity, and it is incorrect to assume that solutions will be well behaved (e.g., tend to a finite value) as the boundary conditions are moved to infinity.

The drag force which Faxén calculated involved a number of undetermined coefficients, so he also calculated it using solutions to Stokes's governing equations. This solution *also* has unknown coefficients, which he then calculated numerically. Arguing that the two solutions ought to be the same, he matched coefficients between the two results, substituted the numerical coefficients, and thereby arrived at a drag force based on the Oseen governing equation.

This work is noteworthy for two reasons. First, the matching of coefficients between solutions derived from the two different governing equations is prescient, foreshadowing the development of matched asymptotics 30 years later. Secondly, Faxén ultimately concluded that Oseen's "improvement" (Eqn. 3.13) on Stokes' drag coefficient (Eqn. 3.3) is invalid[82]. Faxén's analysis demonstrates that — when properly solved — Oseen's equation yields the same drag coefficient as Stokes', without any additional terms[14].

Studies done by Bohlin and Haberman agreed with Faxén's conclusions[14, 84, 85]. It is not surprising that his results reject Oseen's new term ($3/8R$). We previously explained that Oseen's analysis, although it eliminates the "paradoxes", does not posses

sufficient accuracy to justify more than the lowest order term in Eqn. 3.13.

However, Faxén's results suffer from problems. First, they cannot be systematically used to obtain better approximations. Secondly, Faxén actually solves the problem for bounded flow, with the boundary conditions prescribed by finite geometries. He uses a limiting procedure to extend his solutions to unbounded flow (with boundary conditions imposed on the uniform stream only at infinity, as in Eqn. 2.3). In problems like this, which involve incomplete similarity, it is preferable to work directly in the infinite domain.

Faxén's meticulous devotion to properly applying boundary conditions culminated in the first complete solution to Eqn. 3.11. In 1927, he published a general solution for flow around an infinite cylinder which could exactly satisfy arbitrary boundary conditions[76]. Unfortunately, Faxén's solution contains an infinite number of undetermined integration constants, and approximations must be used to determine these constants. Although this unfortunately destroys the "exact" nature of the solution, these approximations can be made in a controlled, systematic fashion — an improvement over the earlier results of Lamb and Oseen. Although Faxén's solution was the first of its kind, his real insight was realizing that approximations in the application of boundary conditions could be as important as the approximations in the governing equations.

His formal solutions are in essence a difficult extension of Lamb's reformulation of Oseen's equations, and they inspired several similar solutions. In 1929, Goldstein completed a herculean calculation to derive a general solution to Oseen's equation for flow around a sphere[27]. Like Faxén's result for the cylinder, Goldstein's solution can — in principle — exactly satisfy the boundary conditions. Unfortunately, it also suffers from the same problems: It is impossible to determine all of the infinitely many integration constants.

Goldstein's solution is summarized by Tomotika, who also translated it into the

language of stream functions[31]. We combine elements from both papers in quoting the solution given in Eqn. 3.20.

$$\psi(r,\theta) = -r^2 Q_1(\cos\theta) + \sum_{n=1}^{\infty}\left(B_n r^{-n} + \sum_{m=0}^{\infty} X_m r^2 \Phi_{m,n}(rR/2)\right) Q_n(\cos\theta) \tag{3.20}$$

In this equation,

$$\begin{aligned}
Q_n(\mu) &= \int_{-1}^{\mu} P_n(\mu)\mathrm{d}\mu && (3.21a)\\
\Phi_{m,n}(x) &= -\left(\frac{m}{2m-1}\chi_{m-1}(x) + \frac{m+1}{2m+3}\chi_{m+1}(x)\right) f_{m,n}(x) \\
&\quad -\left(\frac{m}{2m+1} f_{m-1,n}(x) + \frac{m+1}{2m+1} f_{m+1,n}(x)\right)\chi_m(x) && (3.21b)\\
\chi_m(x) &= (2m+1)\left(\frac{\pi}{2x}\right)^{\left(\frac{1}{2}\right)} K_{m+\frac{1}{2}}(x) && (3.21c)\\
f_{m,n}(x) &= (2n+1)\sum_{j=0}^{m}\frac{(2j)!(2m-2j)!(2n-2j)!}{(j!)^2(2m+2n-2j+1)!} \\
&\quad \times\left(\frac{(m+n-j)!}{(m-j)!(n-j)!}\right)^2 \phi_{m+n-2j}(x) && (3.21d)\\
\phi_n(x) &= (2n+1)\left(\frac{\pi}{2x}\right)^{\frac{1}{2}} I_{n+\frac{1}{2}}(x) \\
f_{m,n}(x) &= \sum_{j=0}^{m} C_m(k)\frac{\partial^j \phi_n(x)}{\partial x^j} && (3.21e)
\end{aligned}$$

Here $K_n(x)$ and $I_n(x)$ are Bessel functions, $P_m(x)$ are Legendre polynomials, and $C_m(k)$ is the coefficient of x^k in $P_m(x)$. Note that the second expression for $f_{m,n}(x)$, written in terms of derivatives, is computationally convenient[27].

Eqn. 3.20 is given with undetermined constants of integration, B_n and X_m. Methods to determine these constants were discussed by both Tomotika[31] and Goldstein[27]. We will present our own analysis later.

There are many different results which have been obtained using the above general solution. The exact formula for the stream function and the drag coefficient depend on what terms in the solution are retained, and how one meets the boundary conditions. In general, retaining n angular terms in Eqn. 3.20 requires the retention of $m = n-1$ terms in the second sum. In his original paper, Goldstein retains three terms in each

series, and thereby calculates the formula for C_D given in Eqn. 3.22.

$$C_D = \frac{6\pi}{R}\left(1 + \frac{3}{8}R - \frac{19}{320}R^2 + \frac{71}{2560}R^3 - \frac{30179}{2150400}R^4 + \frac{122519}{17203200}R^5 + \mathcal{O}\left(R^6\right)\right) \tag{3.22}$$

The coefficient of the last term reflects a correction due to Shanks[86].

To obtain the result in Eqn. 3.22, Goldstein both truncated his solution for the stream function and then expanded the resulting C_D about $R = 0$. Van Dyke, extended this result to include an additional 24 terms, for purposes of studying the mathematical structure of the series, but not because of any intrinsic physical meaning[87]. Van Dyke does not make it clear whether he was including more harmonics in the stream function solution or simply increasing the length of the power series given in Eqn. 3.22.

In addition to expressing Goldstein's solution for the geometry of a sphere in terms of stream functions, Tomotika derived his own exact solution to Eqn. 3.11 for flow past a cylinder[31]. Tomotika closely followed the spirit of Lamb[56] and Goldstein[27], and his resulting "analysis is quite different from Faxén's."[31]. His solution to Eqn. 3.11 is given in Eqn. 3.23 below, conveniently expressed in terms of stream functions. Note that Tomotika's result suffers from the same problems as his predecessors: An infinite number of undetermined integration constants.

$$\psi(r,\theta) = r\sin\theta + \sum_{n=1}^{\infty}\left(B_n r^{-n} + \sum_{m=0}^{\infty} X_m r \Phi_{m,n}(rR/2)\right)\sin n\theta \tag{3.23}$$

Where

$$\begin{aligned}\Phi_{m,n}(x) &= \left(K_{m+1}(x) + K_{m-1}(x)\right)\left(I_{m-n}(x) + I_{m+n}(x)\right) \\ &\quad + K_m(x)\left(I_{m-n-1}(x) - I_{m-n+1}(x) - I_{m+n-1}(x) + I_{m+n+1}(x)\right)\end{aligned} \tag{3.24}$$

As before, B_n and X_m are constants of integration which need to be determined by the boundary conditions (Eqn. 2.9).

As with Goldstein's solution for the sphere, approximations are necessary in order to actually calculate a drag coefficient. By retaining the $m = 0$ and $n = 1$ terms, Tomotika

reproduced Bairstow's result for C_D (Eqn. 3.19). He also numerically calculated drag coefficients based on retaining more terms. As with the Goldstein solution, the keeping n angular terms requires keeping $m = n - 1$ terms in the second sum.

The solutions given in Eqns. 3.20 and 3.23 represent the culmination of efforts to solve Oseen's equation for both the sphere and the cylinder. These general solutions are also needed in both matched asymptotics and the new techniques presented in this chapter[26].

There is a final noteworthy solution to Eqn. 3.11. In 1954, Imai published a general method method for solving the problem of flow past an arbitrary cylindrical body[88]. His elegant technique, based on analytic functions, applies to more general geometries. Imai calculated a formula for C_D, approximating the functions in his exact solution with power series about $R = 0$. His result (re-expressed in our notation) is given in Eqn. 3.25.

$$C_D = \frac{4\pi}{R}\delta + R\left(-\frac{\pi}{2} + \frac{\pi\delta}{4} - \frac{5\pi\delta^2}{32}\right) \tag{3.25}$$

Note that Imai's result agrees with Eqn. 3.16 at lowest order, the only order to which Oseen's equation really applies. A priori, his result is neither better nor worse than any other solution of Oseen's equation. It is simply different.

Discussion

We have presented Oseen's governing equations for low Reynold number fluid flow. These equations are a linearized approximation to the Navier-Stokes' equations. We have also presented a number of different solutions, for both stream functions and drag coefficients; each of these solutions comes from a unique set of approximations. The approximations which have been made can be put into the following broad categories:

- The governing equation — Oseen's equation approximates the Navier-Stokes equations.

- Solutions which only satisfy the Oseen's equation approximately.
- Solutions which only satisfy the boundary conditions approximately.
- Solutions where the stream function is expanded in a power series about $R = 0$ after its derivation.
- Approximations in the drag coefficient derivation.
- Drag coefficients which were expanded in a power series about $R = 0$ after their derivation.

The first approximation is in the governing equations. Oseen's approximation is an *ad hoc* approximation which, although it can be shown to be self-consistent, requires unusual cleverness to obtain. Because it is not derived systematically, it can be difficult to understand either its applicability or the limitations of its solutions. There have been years of discussion and confusion about both the equation and its solutions. The short answer is this: Oseen's governing equation is a zeroth order uniformly valid approximation to the Navier Stokes equation; the equation and its solutions are valid only at $\mathcal{O}(R^0)$.

It is not easy to prove this claim rigorously[82]. However, it can be easily shown that Oseen's equations are self-consistent with its solutions, and that the error in the solution is of $\mathcal{O}(R^1)$. One way to explicitly demonstrate this is by substituting a solution of Oseen's equation into the LHS of the Navier-Stokes equations (Eqn. 2.5), thereby estimating the contribution of inertial terms for the flow field characterized by the solution. By repeating that substitution into the LHS of Oseen's equation (Eqn. 3.9), one can estimate the contribution of inertial terms under Oseen's approximations. Comparing the two results gives an estimate of the inaccuracies in Oseen's governing equations.

Concretely, for the sphere, we substitute Eqn. 3.12 into the RHS of Eqn. 3.10, and into the RHS of Eqn. 2.11. The difference between the two results is of $\mathcal{O}(R^1)$.

For the cylinder, substitute Eqn. 3.18 into the RHS of Eqns. 3.11 and 2.8. The difference between the exact and approximate inertial terms is of $\mathcal{O}(R\delta)$, where δ is defined as in Eqn. 3.18.

These conclusions do not depend on the choice of solution (or on the number of terms retained in Eqn. 3.18). They explicitly show that the governing equation is only valid to $\mathcal{O}(R)$ (or $\mathcal{O}(R\delta)$). Consequently, the solutions can only be meaningful to the same order, and the boundary conditions need only be satisfied to that order. With these considerations, almost all of the solutions in the preceding section are equivalent. The only ones which are not — such as Eqn. 3.15 — are those in which the solution itself has been further approximated.[8]

Since the formulae for determining C_D (Eqns. 2.18 and 2.23) are of the form $1/R$ + terms linear in stream function + nonlinear terms, a stream function which is valid to $\mathcal{O}(R)$ will result in a drag coefficient which is valid to $\mathcal{O}(R^0)$. Thus, in all of the formula for C_D which have been presented so far, only the first term is meaningful. For a sphere, this is the Stokes' drag (Eqn. 3.3), and for the cylinder, Lamb's results (Eqn. 3.16).

We have concluded that it is only good fortune that Oseen's new "$3/8R$" term is actually correct. This concurs with the analysis of Proudman et al., who wrote, "Strictly, Oseen's method gives only the leading term ... and is scarcely to be counted as superior to Stokes's method for the purpose of obtaining the drag."[26] Proudman and Pearson also note that the vast effort expended finding exact solutions to Oseen's equation is "of limited value." Goldstein's formula for C_D, for instance, is expanded to $\mathcal{O}(R^5)$, well beyond the accuracy of the original governing equations. The reason for Oseen's good fortune is rooted in the symmetry of the problem. Chester and Van Dyke both observe that the non-linear terms which Oseen's calculation neglects,

[8] In this case, the Bessel functions have been expanded near $R = 0$ and are no longer well behaved as $R \to \infty$

while needed for a correct stream function, do not contribute to the drag because of symmetry[29, 89].

Lindgren argues Faxén proved that, when the boundary conditions are met properly and Oseen's equations solved exactly, the resulting C_D is that obtained by Stokes (Eqn. 3.3)[14]. Whether this argument is correct does not matter, as Oseen's additional term is beyond the accuracy of his governing equations.

There is another approximation which arises while computing C_D in the context of Oseen's equation. Many workers (e.g.,[31]) compute the pressure in Eqns. 2.18 and 2.23 by integrating Oseen's equation (Eqn. 3.9, rather than the Navier-Stokes equations (Eqn. 2.5). In Eqns. 2.19 and 2.24, we presented a pressure calculation based on the Navier Stokes equations. Calculating pressure using the linearized Oseen equation introduces an additional approximation into C_D. While not necessarily problematic or inconsistent, this approximation can be difficult to identify.

Two Different Interpretations

One criticism of the Oseen equation is that it may be obtained by linearizing the Navier-Stokes equations, without regard to the magnitude of inertial and viscous terms. By writing $\vec{u} = \vec{U}_\infty + \delta\vec{u}$, treating $\delta\vec{u}$ as a small perturbation, and expanding Eqn. 2.5 one can formally reproduce Oseen's equations. Clearly, the disturbance to the uniform stream is not negligible near the surface of the solid body, and therefore Oseen's equations "would appear to be a poor approximation in the neighborhood of the body where the boundary condition $\vec{u} = 0$ requires that the true inertial term be small."[70]

This incorrect argument, put forth as a reason to use Stokes' solutions, overlooks the origins of Oseen's equations. The point of Oseen's approximation is that inertial terms are *only* significant at large values of $|r|$, where $R|r|$ is no longer negligible. Near the surface of the solid, the approximate inertial terms which Oseen introduced are negligible in comparison to the viscous terms, because they are multiplied by the

factor R (in the LHS of Eqn. 3.9). Hence the difference between Oseen's and Stokes' equations in the neighborhood of the sphere will be of $\mathcal{O}(R)$, and is beyond the scope of either theory.

Better Approximations

The approach of Whitehead was essentially to improve Stokes' solution for the sphere in an iterative fashion[74]. By substituting the first approximation into the governing equations, he estimated the neglected terms. He then tried, and failed, to solve the resulting governing equation. This approach fails because the Stokes' equations are not uniformly valid to zeroth order.

Oseen's equations are uniformly valid, and, as Proudman remarked, "there seems little reason to doubt that Whitehead's iterative method, using Oseen's equation rather than Stokes's equation would yield an expansion, each successive term of which would represent a uniformly valid higher approximation to the flow. In each step of the iteration, a lower-order approximation would be used to calculate those particular inertia terms that are neglected ... the expansion generated in this way would seem to be the most economic expansion possible."[26]

Proudman did not follow through on this idea, instead developing a solution based on matched asymptotics expansions (see below). In an appendix, Van Dyke relates the unpublished work of C. R. Illingworth (1947)[29]. Illingworth carried through Whitehead's program, deriving a new expression (Eqn. 3.26) for C_D, which agrees to $\mathcal{O}(R^2 \ln R)$ with the later results of matched asymptotic calculations (Eqn. 3.29).

$$C_D = \frac{6\pi}{R}\left(1 + \frac{3}{8}R + \frac{9}{40}R^2 \log R + 0.1333R^2 + \frac{81}{320}R^3 \log R - 0.0034R^3 + \ldots\right) \tag{3.26}$$

Although this result has since been subsumed by matched asymptotics, it is nonetheless remarkable, substantially improving any previous drag calculations, and rigorously justifying Oseen's $3/8R$ term.

There have also been efforts (e.g., [86, 87]) to "re-sum" Goldstein's series expansion for C_D (Eqn. 3.22). However, these results have little intrinsic (as opposed to methodological) value, as Goldstein's result is only valid to $\mathcal{O}(R)$. If applied to more accurate approximations, such as Eqn. 3.26, these methods could be worthwhile. Alas, even improved approximations lack a sufficient numbers of terms in the expression for C_D to make this practicable.

Summary

Simply put, Oseen's equations resolved the paradoxes of Stokes and Whitehead, put Stokes' results on firm theoretical ground, and led to the first solution for the drag on a cylinder. Although the Oseen equations happen to provide a uniformly valid first approximation, it is difficult to extend this work to higher order approximations.

Figure 3.6 compares the "predictions" of Oseen theory to experimental and numerical data for the drag on a sphere. Again, Oseen's first order theory is, strictly speaking, not adequate to make the predictions with which it is traditionally credited. The theoretical drag coefficients are roughly valid for $R \lesssim 0.2$, with Goldstein's solution (Eqn. 3.22) being slightly better than Oseen's prediction (Eqn. 3.13). All are clearly superior to Stokes' formula (Eqn. 3.3).

Figure 3.6 also shows the prediction of Illingworth's second order Oseen theory (Eqn. 3.26). Not surprisingly, it gives the best prediction of C_D, particularly when compared to Dennis' numerical results.

Figure 3.7 shows the important predictions of Oseen theory for the drag on an infinite cylinder. As with the sphere, the theory is only truly entitled to predict the lowest order term. Figure 3.7 shows decent agreement with the data. Although more "exact" solutions (such as Bairstow's and Imai's) do better than Lamb's lowest order solution, this is purely coincidental. Tomotika's solutions exhibit similar characteristics to these two solutions.

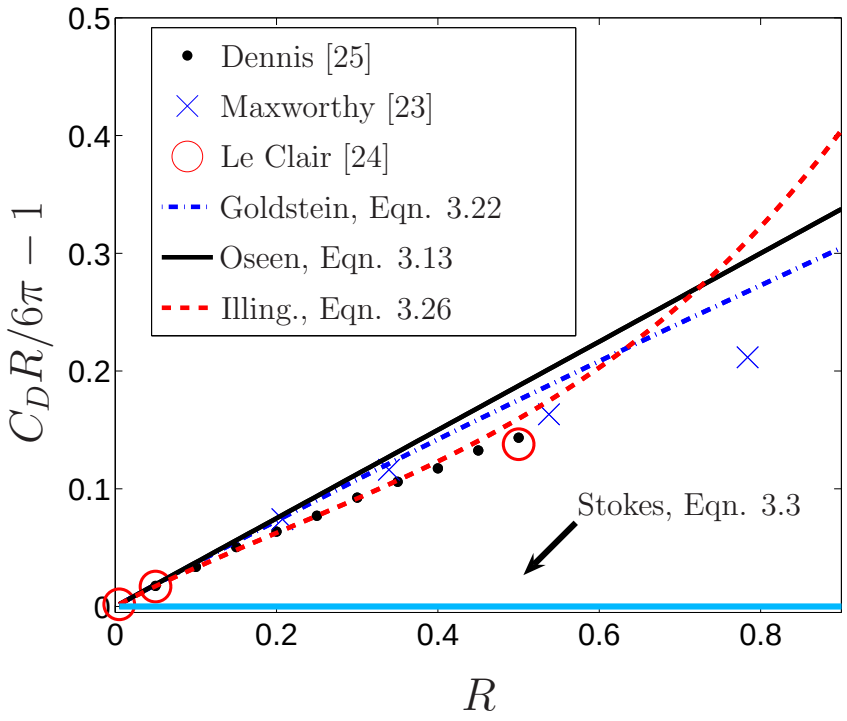

Figure 3.6: Drag on a sphere: expt. vs. Oseen theory. The Stokes' solution (Eqn. 3.3) is shown at the bottom for reference. In these coordinates, it is defined by the line $y = 0$.

3.2.3 Matched Asymptotics

Efforts to systematically improve Oseen's results led to the development of *matched asymptotic expansions*.[9] This branch of applied mathematics was developed gradually, with systematic work beginning with papers by Kaplun and Lagerstrom et al.[90, 91]. Kaplun subsequently used these techniques to calculate the drag on a cylinder, obtaining an entirely new result for C_D[34]. Proudman and Pearson subsequently applied matched asymptotics to both the sphere and the cylinder, deriving a new result for the drag on a sphere[26].

> "The principle of asymptotic matching is simple. The interval on which a boundary-value problem is posed is broken into a sequence of two or more

[9] This technique is also known as the method of *inner and outer expansions* or *double asymptotic expansions*.

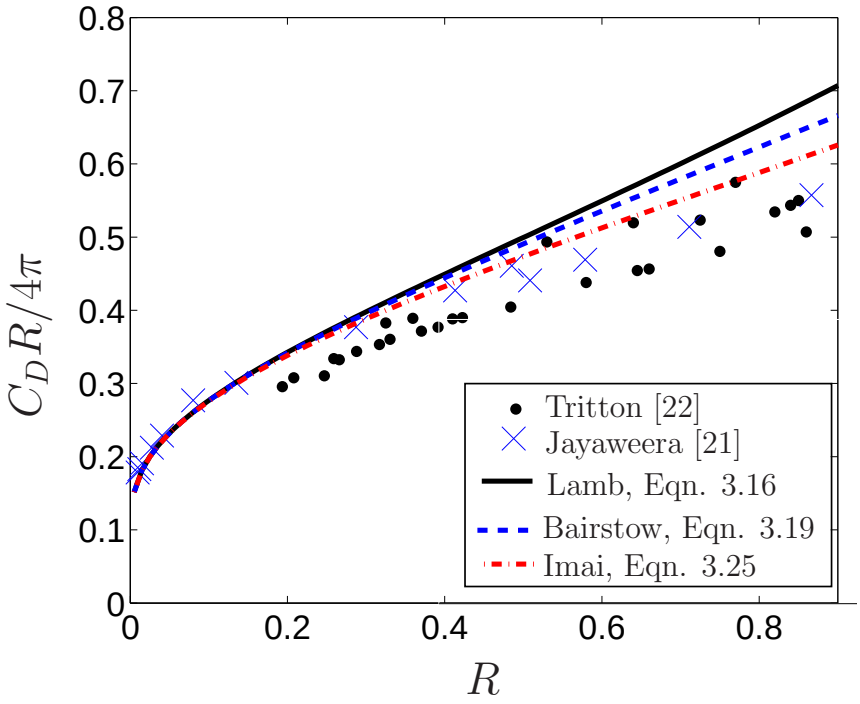

Figure 3.7: Drag on a cylinder: expt. vs. Oseen theory

> *overlapping* subintervals. Then, on each subinterval perturbation theory is used to obtain an asymptotic approximation to the solution of the differential equation valid on that interval. Finally, the matching is done by requiring that the asymptotic approximations have the same functional form on the overlap of every pair or intervals. This gives a sequence of asymptotic approximations ... the end result is an approximate solution to a boundary-value problem valid over the entire interval"[15].

Both of the two low Reynolds number problems are attacked in similar fashion. The problem is divided into only two regions. The first region is near the surface of the solid body. In this region, inertial terms are small, the approximation of Stokes ($R \approx 0$) applies, and the problem is solved perturbatively (in R). At each order in R, the two no-slip boundary conditions at the surface are applied. One undetermined constant remains (at each order in R). Loosely speaking, it is determined by the boundary

condition as $|\vec{r}| \to \infty$. This expansion is referred to as the *Stokes expansion*.

The second region is far from the sphere, where inertial terms are important. In this region, $R|r| \sim \mathcal{O}(1)$, and the approximations which led to Oseen's governing equation apply. The Oseen problem is then solved perturbatively, and the boundary condition as $|\vec{r}| \to \infty$ is applied. There are two undetermined constants remaining; they are related to the boundary conditions on the surface. This perturbative expansion is referred to as the *Oseen expansion*.

The next part of this calculation is *asymptotic matching*, which determines the remaining coefficients.[10] In this process, we expand the Oseen expansion for small $R|\vec{r}|$, and the Stokes expansion for large $|\vec{r}|$. By choosing the three hitherto undetermined coefficients appropriately, these two limiting forms are made to agree order by order in R. For this to be possible, the two asymptotic functional forms must overlap. With the coefficients determined, the two unique, locally valid perturbative approximations are complete. If desired, they can be combined to make a single uniformly valid approximation.

While straightforward in theory, asymptotic matching is difficult in practice, particularly for an equation like the Navier-Stokes equation. However, it is still far simpler than alternatives, such as iteratively solving the Oseen equations. Van Dyke's book is an excellent presentation of the many subtleties which arise in applying matched asymptotics to problems in fluid mechanics[29]. We now present the matched asymptotic solutions for Eqns. 2.8 and 2.11. These solutions result in the "state of the art" drag coefficients for both the sphere and the cylinder.

[10] At this point, there are two unknown coefficients in the Oseen expansion, and one in the Stokes expansion

Sphere

Although Lagerstrom and Cole initially applied matched asymptotics to the problem of the sphere, the seminal work came in an elegant 1957 paper by Proudman and Pearson[26, 90]. Chester and Breach extended this paper via a difficult calculation in 1969[13]. We summarize the results of both papers here. These workers used a perturbative solution in the Stokes regime of the form:

$$\psi(r,\mu) = \psi_0 + R\psi_1 + R^2 \log R\psi_{2L} + R^2\psi_2 + R^3 \log R + R^3\psi_3 + \mathcal{O}\left(R^3\right) \tag{3.27}$$

This rather peculiar perturbative form cannot be determined a priori. Rather, it arose in a fastidious incremental fashion, calculating one term at a time. The procedure of asymptotic matching *required* including terms like $R^2 \log R$ in the expansion; otherwise, no matching is possible. Note that matched asymptotics gives no explanation for the origin of these singular terms.

The first step to finding a perturbative solution in the Oseen region is to define the *Oseen variables*:

$$\rho = Rr, \qquad \Psi(\rho,\mu) = R^2\psi(r,\mu)$$

Part of the reason for this transformation can be understood via the derivation of the Oseen equation. The region where inertial effects become important has been shown to be where $R|r| \sim \mathcal{O}(1)$. Intuitively, the variable $\rho = Rr$ is a natural choice to analyze this regime, as it will be of $\mathcal{O}(1)$. The precise reasons for the selection of these variables is a technique from boundary layer theory known as a *dominant balance* argument. See Bender and Orzag for a detailed explanation. This argument will also be revisited in subsequent sections[15].

The perturbative expansion in the Oseen takes the form:

$$\Psi(\rho,\mu) = \Psi_0 + R\Psi_1 + R^2\Psi_2 + R^3 \log R\Psi_{3L} + \mathcal{O}\left(R^3\right) \tag{3.28}$$

Note that there is no $R^2 \log R$ term in this expansion; none is required to match with

the Stokes expansion. As with the Stokes expansion, this form cannot be determined a priori.

Proudman and Pearson completely solved for the Stokes' expansion through $\mathcal{O}(R \log R)$, and partially solved for the $\mathcal{O}(R^2)$ term. They determined the Oseen expansion through $\mathcal{O}(R)$. Chester and Breach extended these results up to a partial solution for $\mathcal{O}(R^3)$ in the Stokes' expansion, and to $\mathcal{O}(R^3 \log R)$ in the Oseen expansion.

The exact form of these expansions is given in Chester and Breach.[11] Some aspects of these results have been seen before: The leading order in the Stokes' expansion (ψ_0) is simply the Stokes solution (Eqn. 3.2). In the Oseen expansion, Ψ_0 is simply the formula for the uniform stream expressed in Oseen variables. The second term, Ψ_1, is the rotational part of Oseen's solution (Eqn. 3.12).

Both sets of authors then used their result for the Stokes expansion to calculate C_D, which is given in Eqn. 3.29.

$$C_D = \frac{6\pi}{R}\Bigg(\underbrace{1}_{\text{"Stokes"}} \underbrace{+\frac{3}{8}R}_{\text{"Oseen"}} \underbrace{+\frac{9}{40}R^2 \log R}_{\text{Proudman}} \underbrace{+\frac{9}{40}R^2\left(\gamma + \frac{5}{3}\log 2 - \frac{323}{360}\right) + \frac{27}{80}R^3 \log R}_{\text{Chester and Breach}} + \mathcal{O}\left(R^3\right)\Bigg) \tag{3.29}$$

Here γ is Euler's constant. This formula reproduces and extends nearly all earlier work. Eqn. 3.29 shows both the original results of Proudman and Pearson and the higher order contributions of Chester and Breach[13, 26]. The "Stokes" term is Stokes' original result (Eqn. 3.3), which was rigorously justified by Oseen. The "Oseen" term is generally credited to Oseen (Eqn. 3.13), although it is really beyond the accuracy of his work, and is only justified by this calculation.[12]

Figure 3.8 compares the results of matched asymptotics (Eqn. 3.29) with experimental data, numerical results, and the basic prediction of Oseen's equation (Eqn.

[11] Note that the expression for ψ_2 in Proudman, is incorrect[26]. There is also a mistake in Chester and Breach[13], Eqn. 3.5; the coefficient of c_8 should be r^{-3} not r^{-2}.

[12] Illingworth's unpublished result also justifies this term.

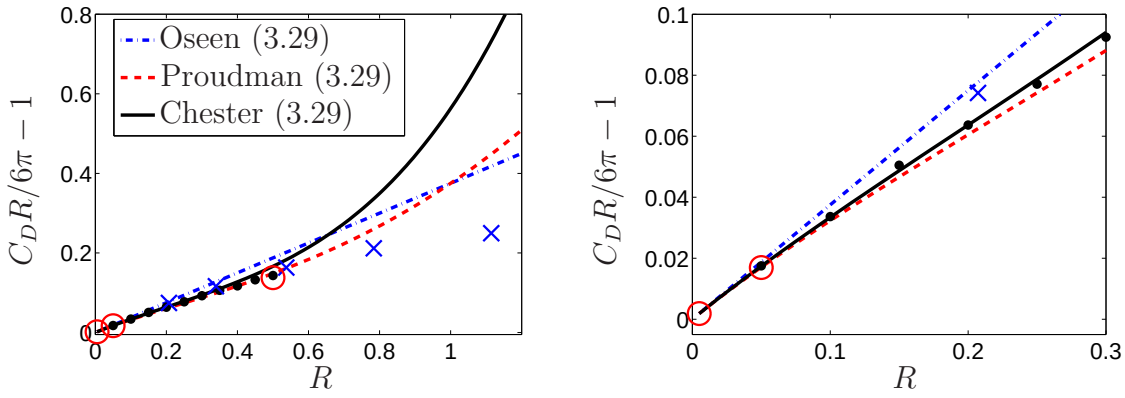

Figure 3.8: Drag on a sphere: experiment vs. matched asymptotic theory. Experimental and numerical results are plotted as in Figure 3.6.

3.13). This plot has been the source of some confusion. Maxworthy examined his data and concluded that C_D as computed by Oseen and Goldstein (Eqn. 3.22) were as good as any matched asymptotics predictions[23]. The calculations of Dennis and Le Clair, however, refute that claim, and demonstrate the systematic improvement that results from matched asymptotics.

Neither is it immediately clear that the extra terms in Eqn. 3.29 due to Chester and Breach are actually any improvement on the work of Proudman and Pearson. Van Dyke notes, "This result is disappointing, because comparison with experiment suggests that the range of applicability has scarcely been increased."[29], and Chester himself remarks that "there is little point in continuing the expansion further." At very low Reynolds number, however, the results of Dennis "indicate that the expression of Chester and Breach gives a better approximation to the drag coefficient than any other asymptotic solution until about $[R = 0.3]$."[25] Figure 3.8 shows the excellent low R agreement between Dennis' numerical results and Eqn. 3.29.

The prediction of matched asymptotics (Eqn. 3.29) is close to Illingworth's second order Oseen theory (Eqn. 3.26). Close examination shows that the matched asymptotics results are slightly closer to the Dennis' calculations in the limit of low

Reynolds number. Strictly speaking, these two theories should only be compared as $r \to 0$, and in this regime matched asymptotics is superior. This is not surprising, as the best matched asymptotic calculation is a higher order approximation than that of Illingworth.

Cylinder

In 1957, Kaplun applied matched asymptotics to the problem of the cylinder, and produced the first new result for C_D[30]. Additional stream function calculations (but without a drag coefficient) were done by Proudman and Pearson[26]. Kaplun's calculations were extended to higher order by Skinner, whose work also explored the structure of the asymptotic expansions[38]. We summarize results from all three papers here.

Near the surface of the cylinder, the Stokes' expansion applies, and the perturbative solution takes the following form.

$$\psi(r,\theta) = \psi_0(r,\theta,\delta) + R\psi_2(r,\theta,\delta) + R^2\psi_3(r,\theta,\delta) + \mathcal{O}\left(R^3\right) \tag{3.30}$$

Here, $\delta = \delta(R)$ is defined as in Eqn. 3.15. What is remarkable about the structure of this expansion is its dependence on δ. To be precise, each function ψ_n is actually another perturbative expansion, in δ:

$$\psi_n(r,\theta,\delta) = \delta F_{n,1}(r,\theta) + \delta^2 F_{n,2}(r,\theta) + \mathcal{O}\left(\delta^3\right) \tag{3.31}$$

This formulation is equivalent to an asymptotic expansion in terms of $\log R^{-1}$, which is used by Proudman and Pearson:

$$\psi_n(r,\theta,\log R) = \frac{\tilde{F}_{n,1}(r,\theta)}{(\log R)^1} + \frac{\tilde{F}_{n,2}(r,\theta)}{(\log R)^2} + \mathcal{O}\left(\frac{1}{(\log R)^3}\right) \tag{3.32}$$

This form is much less efficient than that given in Eqn. 3.31, in the sense that more terms in the Stokes and Oseen expansions are needed to obtain a given number of terms in C_D. For that reason, expansions in δ are used here.

This curious asymptotic form is necessitated by matching requirements. It is also the source of a number of bizarre complications. The first implication is that *all* terms in Eqn. 3.30 of $\mathcal{O}(R)$ and higher will be transcendentally smaller than *any* of the terms in the expansion for ψ_0. This is true asymptotically, as $R \to 0$. The reason for this is that inertial terms *never* enter into any of the governing equations for the Stokes' expansion; they enter only through the matching process with the Oseen expansion.

As with the sphere, the first step to finding a perturbative solution in the Oseen region is to transform into the relevant Oseen variables. In this case,

$$\rho = Rr, \qquad \Psi(\rho, \mu) = R\psi(r, \mu) \tag{3.33}$$

The perturbative expansion which can solve the problem in the Oseen region has the same generic form as Eqn. 3.30.

$$\Psi(\rho, \theta) = \Psi_0(\rho, \theta, \delta) + R\Psi_1(\rho, \theta, \delta) + \mathcal{O}\left(R^2\right) \tag{3.34}$$

The functions $\Psi_n(\rho, \theta, \delta)$ can also be expressed as a series in $\delta(R)$. However, the formula cannot be written down as conveniently as it could in Eqn. 3.31. The first two terms take the forms given in Eqn. 3.35.

$$\Psi_0(\rho, \theta, \delta) = F_{0,0}(\rho, \theta) + \delta F_{0,1}(\rho, \theta) + \mathcal{O}\left(\delta^2\right) \tag{3.35a}$$

$$\Psi_1(\rho, \theta, \delta) = \delta^{-1} F_{1,-1}(\rho, \theta) + F_{0,0}(\rho, \theta) + \mathcal{O}(\delta) \tag{3.35b}$$

Kaplun and Proudman both considered only terms of $\mathcal{O}(R^0)$ in the Stokes' expansion. As $R \to 0$, this is an excellent approximation, as all higher terms are transcendentally smaller. In this limit, the Stokes expansion takes a particularly simple form:

$$\psi(r, \theta) = \psi_1(r, \theta, \delta) = \sum_{n=1}^{\infty} a_n \delta^n \left(2r \log r - r + \frac{1}{r}\right) \sin\theta \tag{3.36}$$

Kaplun obtained terms up to and including $n = 3$. Proudman et al. also obtained expressions for the Oseen expansion, albeit expressed as a series in $\log R^{-1}$. Skinner extended Kaplun's Stokes expansion to include terms up to $\mathcal{O}(\delta^3)$, $\mathcal{O}(R\delta)$, and

$\mathcal{O}\left(R^2\delta\right)$[38]. He obtained approximate solutions for the Oseen expansion, including terms up to $\mathcal{O}\left(\delta\right)$ and $\mathcal{O}\left(R\right)$. The lowest order solutions in the Oseen expansion are related to the expression for a uniform stream and the solution of Lamb (Eqn. 3.15).

Using his solution, Kaplun computed a new result for the drag coefficient (Eqn. 3.37) which agrees with Lamb's result (Eqn. 3.16) at lowest order.

$$C_D = \frac{4\pi}{R}\left(\delta - k\delta^3\right) \tag{3.37}$$

Here, $k = \int_0^\infty K_0(x)K_1(x)\left(x^{-1}I_1(2x) - 4K_1(x)I_1(x) + 1\right)\mathrm{d}x \approx 0.87$. Skinner extended these results, showed that terms of $\mathcal{O}\left(R^0\right)$ do not contribute to the drag, and calculated the first transcendentally smaller contribution, which is of $\mathcal{O}\left(R^1\right)$. His result is given in Eqn. 3.38.

$$C_D = \frac{4\pi}{R}\left(\delta - k\delta^3 + \mathcal{O}\left(\delta^4\right) - \frac{R^2}{32}\left(1 - \frac{\delta}{2} + \mathcal{O}\left(\delta^2\right)\right) + \mathcal{O}\left(R^4\right)\right) \tag{3.38}$$

The value of these new terms is questionable, and Skinner himself noted that they are likely negligible in comparison to the neglected terms of $\mathcal{O}\left(\delta^4\right)$. Asymptotically this is unequivocally true.

Figure 3.9 compares the predictions of matched asymptotic theory with Lamb's result (Eqn. 3.16) based on Oseen's equation. Although both theories agree as $r \to 0$, matched asymptotic results seem no better than Lamb's solution. The comparison is further complicated by the scatter in the different experiments; matched asymptotics agree more closely with Tritton's measurements, while Lamb's solution agrees better with Jayaweera's. We draw two conclusions from Figure 3.9: Both theories break down for $R \gtrsim 0.1$ and neither theory is demonstrably more accurate. Even more disappointingly, Skinner's result is nowhere better than Kaplun's — it is actually worse at higher Reynolds numbers.

Part of the problem with the matched asymptotics approach arises from the need for two expansions, in δ and R. Because infinitely many orders of δ are needed before *any* higher orders in R are relevant means that infinitely many terms in the Oseen

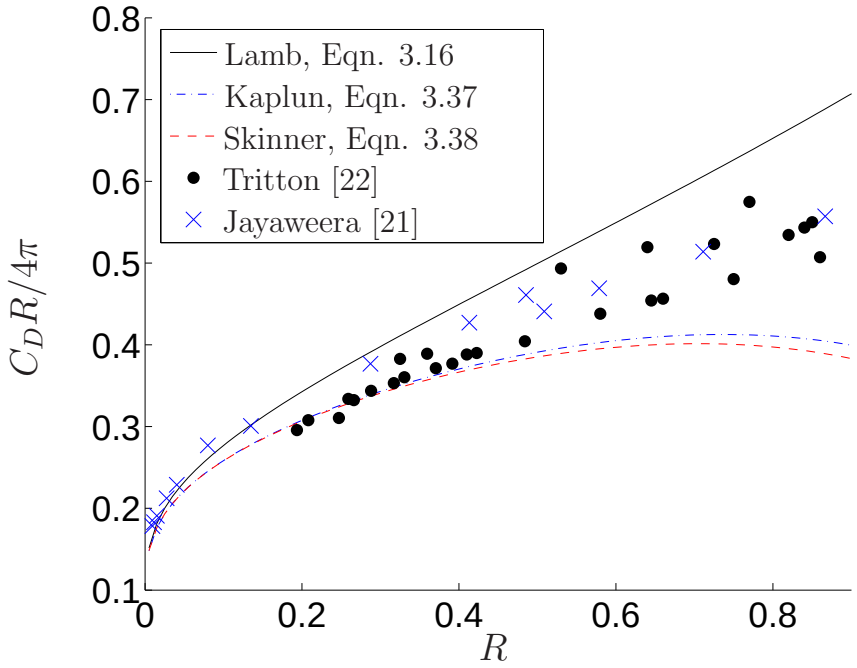

Figure 3.9: Drag on a cylinder: expt. vs. matched asymptotic theory

expansion must be calculated before the second order term in the Stokes expansion. This is inefficient, and is the reason for Skinner's lack of success.

A recent paper by Keller et al. solved this problem numerically[92]. They developed a method to sum *all* of the orders of δ for the first two orders of R. Their "beyond all orders" numerical results prove the importance of these higher order terms. When such terms are accounted for, the resulting C_D is vastly improved from Kaplun's, and is superior to any of the analytic solutions discussed here. Interestingly, it seems to agree very well with the experiments of Tritton, although it is difficult to tell from the plot in their paper, which does not remove the leading order divergence.

3.2.4 Other Theories

Amongst the community interested in settling velocities and sedimentation, there are many theoretical models of the drag on a sphere. These workers specify C_D as a function of R by means of a "sphere drag correlation." An overview of these formula

is given by Brown[55]. These results are generally semi-empirical, relying on a blend of theoretical calculations and phenomenologically fit parameters to predict C_D over a large range of Reynolds number. While practically useful, these results are not specific to low Reynolds numbers, and cannot be derived from the Navier-Stokes equations. They address a different problem, and will not be further considered here.

One other semi-empirical theory is due to Carrier[93]. He argued that the inertial corrections in the Oseen equation were over-weighted, and multiplied them by a coefficient which he constrained to be between 0 and 1. Consequently, his theory is in some sense "in between" that of Stokes and that of Oseen. He ultimately determined this coefficient empirically.

3.2.5 Terminology

Confusing terminology, particularly in the matched asymptotics literature, riddles the history of these problems. We previously detailed discrepancies in the definition of C_D. In this section we explain the sometimes conflicting terms used in the matched asymptotics literature, adopt a convention which eliminates confusion, and also explain how some authors adopt different definitions of the Reynolds' number.

Matched asymptotics literature discusses numerous perturbative expansions, each of which are valid in a different regime, or "domain of validity." Different authors use different labels for these expansions. Most workers define the "inner" expansion to be the expansion which is valid inside the *boundary layer*[15]. A boundary layer is a region of rapid variation in the solution. The "outer" expansion is valid outside of the boundary layer, where the solution is slowly varying[15]. Problems with multiple boundary layers require additional terminology. The outer expansion is based "upon the primary reference quantities in the problem," and the inner expansion is usually obtained by stretching the original variables by dimensionless functions of the perturbation parameter[29]. The appropriate stretching, or scaling functions are obtained

through a *dominant balance* analysis, which can be difficult. After this rescaling, the argument of the inner expansion will be of $\mathcal{O}(1)$ inside the boundary layer. Accompanying these inner and outer expansions are "inner variables", "outer variables", "inner limits", and "outer limits".

The low Reynolds number flow problems are complicated by the fact that some authors, including Van Dyke, also define expansions on the basis of their physical location[29]. The "outer" limit is valid far from the solid body ($|\vec{r}|$ is large), and the "inner" limit is valid near the surface of the body ($|\vec{r}| \approx 1$).

This is consistent with yet another definition, based on proximity to the origin of the chosen coordinate system. In a review paper Lagerstrom and Casten define the "inner limit" as being "valid near the origin," and the "outer limit" as being "valid except near the origin."[94] Part of their motivation for this new definition was to distinguish between the *domain of validity* of an expansion, and the limit process by which it is obtained.

Finally, Kaplun refers to the inner and outer limits based on their correspondence to high Reynolds number flow[30]. He identifies the Stokes' approximation as the "inner" limit, and Oseen's equation as the "outer" limit.

Part of the confusion arises because of disagreements over the location of the boundary layer. Van Dyke claims that "it is the neighborhood of the point at infinity", while Kaplun argues that the boundary layer is near the surface. Definitions referenced to the boundary layer disagree when there are disagreements about its location.

To eliminate this confusion, a preferable alternative notation has emerged from the work of both Kaplun and Proudman et al.[26, 80]. We follow this notation, defining the "Oseen" and "Stokes" expansions, which were used in the previous section. The Oseen expansion is valid far from the surface, and is expressed in stretched coordinates. The Stokes limit is valid near the surface of the sphere, where r is small, and is expressed

in the original variables.[13]

Matched asymptotics workers also discuss *uniform approximations*, *intermediate* expansions, or *composite* expansions[15, 29, 34]. The basic idea is that the Stokes and Oseen expansions can be blended together to form a single expression which is valid everywhere. This result reduces to the two original expansions when expanded asymptotically in the two limits. How to calculate a uniform expansion is discused below.

There are also minor differences in the definition of the Reynolds number, R. Some authors define R based on the diameter of the solid, while others base it on the radius. This factor of 2 can be difficult to track. We define the Reynolds number using the *radius* of the fixed body: $R = |\vec{u}_\infty|a/\nu$. It is worth noting that Kaplun[34], Tomotika[31], Goldstein[35], Liebster[36], Thom[61] and Tritton[22] all use the *diameter*.

3.3 Uniformly valid approximations

As mentioned previously, the inner and outer expansions may be combined into a single, *uniformly valid* approximation, which is applicable everywhere. For a function of one variable, the uniform approximation is constructed as in Eqn. 3.39[15].

$$y_{\text{uniform}}(x) = y_{\text{outer}}(x) + y_{\text{inner}}(x) - y_{\text{overlap}}(x) \tag{3.39}$$

$y_{\text{overlap}}(x)$ consists of the the common "matching" terms between the inner and outer expansions.

Kaplun demonstrates that $y_{\text{uniform}}(x) \to y(x)$ as the expansion variable $R \to 0$, i.e. the uniform approximation tends to the exact solution everywhere[34]. To be more precise, if the matched asymptotics solution is constructed to $\mathcal{O}\left(R^1\right)$, then

$$\lim_{R\to 0} y(x) - y_{\text{uniform}}(x) \sim \mathcal{O}\left(R^1\right)$$

[13]Van Dyke's book is not consistent in relating "inner" and "outer" expansions to the Stokes and Oseen expansions.

As a matter of practice, calculating the uniform solution is mechanistic. First, express the inner and outer expansions in the same coordinates; in our case, express the Oseen expansion in Stokes variables.[14] Alternatively, one can express the Stokes expansion in Oseen variables. Next, express both solutions as a power series in the expansion parameter, R. By construction the Stokes expansion is already in this form but the transformed Oseen expansion is not, and must be expanded to the same power in R as the Stokes solution.

From these two power series we can identify the "overlap" function, y_{overlap}. This function consists of the terms which are in common between the two expansions, and is usually obtained by inspection. Of course, y_{overlap} is only valid to the same order as the original matched asymptotics solution, and higher order terms should be discarded. The uniformly valid approximation is then obtained using y_{overlap} and Eqn. 3.39.

3.3.1 The correct way to calculate C_D

Proudman and Pearson argue that "uniformly valid approximations *per se* are not usually of much physical interest ... In the present problem, for instance, it is the Stokes expansion that gives virtually all the physically interesting information."[26] All matched asymptotics calculations are based solely on the Stokes expansion, and are therefore influenced by the Oseen expansion only via the boundary conditions. For instance, the drag coefficient is calculated using only the Stokes' expansion. Other properties of the stream function, such as the size of the dead water wake directly behind the sphere or cylinder, are also calculated using the Stokes' expansion.

In this section we argue that this approach is incorrect, and that uniformly valid approximation should be used to calculate all quantities of interest. By adopting this viewpoint, we obtain new results for C_D, and demonstrate that these drag coefficients

[14]Note that this transformation affects both the radial coordinates and the stream function, and that it differs for the sphere and cylinder.

systematically improve on previous matched asymptotics results.

Matched asymptotics workers argue that the drag coefficient is calculated at the surface of the solid (Eqns. 2.23, 2.18), where $r = 1$. Since the Oseen solution applies for large r, the Stokes solution applies for small r, and the Stokes solution ought to be used to calculate C_D. In fact, by construction, any uniformly valid approximation must reduce to the Stokes expansion in the limit as $Rr \to 0$.

Curiously, proponents of the Oseen equation argue conversely[70, 76]. They claim that because the Oseen expansion *happens* to apply everywhere, it should be used to calculate all sorts of quantities of interest, including C_D. In fact, Hapel and Brenner wrote a book essentially devoted to this premise[70]. In fairness, it must be mentioned that all of these authors were aware of their choices, and motivated their approach pragmatically: They obtained useful solutions to otherwise intractable problems.

In reality, both approaches converge to the exact solution for suitably small Reynolds' numbers. However, for small but non-infinitesimal R, the best estimate of derivative quantities such as C_D is obtained not by using the Stokes expansion, but by using a uniformly valid approximation calculated with both the Stokes and Oseen expansions. Such a drag coefficient must agree with results derived from the Stokes expansion as $Rr \to 0$, and it can *never* be inferior. Moreover, this approach makes determination of the drag coefficient's accuracy straightforward; it is determined solely by the accuracy of the uniform expansion, without any need to be concerned about its domain of applicability.

We now calculate the drag coefficients for both the sphere and the cylinder using uniformly valid approximations, using previously published inner and outer expansions. These corrections are small but methodologically noteworthy, and are absent from the existing literature.

Cylinder

Although the state-of-the-art matched asymptotics solutions are due to Kaplun, it is more convenient to use stream functions[30]. Skinner conveniently combines previous work, providing a concise summary of Stokes and Oseen stream functions[38]. We base our derivation of a uniformly valid approximation on the results in his paper. The Stokes expansion is given by Eqn. 3.40[38].

$$\psi(r,\theta) = \frac{1}{2}\left(\delta - k\delta^3 + \mathcal{O}\left(\delta^4\right)\right)\left(2r\log r - r + \frac{1}{r}\right)\sin\theta + \mathcal{O}\left(R^1\right) \tag{3.40}$$

The Oseen Expansion is given by Eqn. 3.41.

$$\Psi(\rho,\theta) = \left(\rho\sin\theta - \delta\sum_{n=1}^{\infty}\phi_n\left(\frac{\rho}{2}\right)\frac{\rho}{n}\sin n\theta + \mathcal{O}\left(\delta^2\right) + \mathcal{O}\left(R^1\right)\right) \tag{3.41}$$

With these results, creating the uniform approximation and calculating C_D is straightforward. The only subtlety is the sine series in Eqn. 3.41. However, Eqn. 2.20 tells us that, for the purposes of calculating the drag, only the coefficient of $\sin\theta$ matters. We calculate the overlap between the two functions by expanding Eqn. 3.41 about $\rho = 0$. The result is given by Eqn. 3.42.

$$\psi_{\text{overlap}}(r,\theta) = \delta\frac{r}{2}\left(2\log r - 1\right)\sin\theta + \mathcal{O}\left(\delta^2\right) + \mathcal{O}\left(R^1\right) \tag{3.42}$$

Combining this with the Oseen and Stokes expansions, we obtain the uniformly valid approximation given by Eqn. 3.43.

$$\begin{aligned}\psi_{\text{uniform}}(r,\theta) \;=\; & \left(r + \delta\left(\frac{1}{2r} - r\phi_1(\frac{rR}{2})\right) + k\delta^3\left(\frac{r}{2} - r\log r - \frac{1}{2r}\right)\right)\sin\theta - \\ & \delta\sum_{n=2}^{\infty}\phi_n\left(\frac{Rr}{2}\right)\frac{r}{n}\sin n\theta + \mathcal{O}\left(\delta^2\right) + \mathcal{O}\left(R^1\right)\end{aligned} \tag{3.43}$$

By substituting this result into Eqn. 2.20, we obtain a new result for C_D:

$$C_D = \frac{\pi\delta\left(24 - 32k\delta^3 + 6R^2\phi_1''(R/2) + R^3\phi_1'''(R/2)\right)}{8R} \tag{3.44}$$

Fig. 3.10 compares Eqn. 3.44 with Kaplun's usual result (Eqn. 3.37). The new drag coefficient (Eqn. 3.44) is a small but *systematic* improvement over the results of

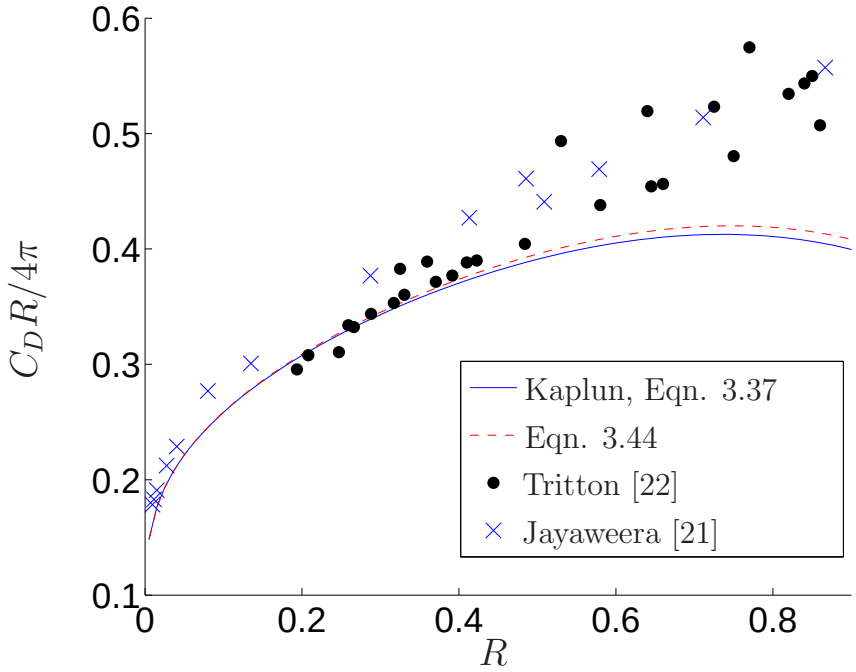

Figure 3.10: Drag on a cylinder, comparing a uniformly valid calculations and matched asymptotics results.

Kaplun. Because they are asymptotically identical up to $\mathcal{O}(\delta^4)$ and $\mathcal{O}(R)$, they agree as $R \to 0$. However, at small but non-infinitesimal R, our new result is superior. Comparing Figures 3.10 and 3.9, we can also see a second surprise: The new result betters Skinner's C_D, even though they were based on the same stream functions. If Skinner had used a uniformly valid approximation, his result would not have misleadingly appeared inferior to Kaplun's.

Sphere

As with the cylinder, calculating C_D from a uniformly valid expansion yields an improved result. However, there is a substantial difference in this case. Although matched asymptotics calculations have been done through $\mathcal{O}(R^3)$ in Eqn. 3.27 and $\mathcal{O}(R^3 \log R)$ in Eqn. 3.28, the higher order terms in the Oseen expansion are impossible to express in a simple analytic form. Asymptotic expressions exist (and have been used for match-

ing), but these cannot be used to construct a uniformly valid expansion. Consequently, we can only compute the uniform expansion through $\mathcal{O}(R)$, and its predictions can only be meaningfully compared to the first two terms in Eqn. 3.29.

The solutions for the Stokes and Oseen expansions are given in Chester and Breach, and are quoted here[13]. The Stokes expansion:

$$\begin{aligned}\psi(r,\mu) &= -\frac{1}{2}\left(2r^2-3r+\frac{1}{r}\right)Q_1(\mu) - R\frac{3}{16}\Bigg(\left(2r^2-3r+\frac{1}{r}\right)Q_1(\mu) - \\ &\quad \left(2r^2-3r+1-\frac{1}{r}+\frac{1}{r^2}\right)Q_2(\mu)\Bigg) + \mathcal{O}\left(R^2\log R\right) \end{aligned} \tag{3.45}$$

The Oseen expansion:

$$\Psi(\rho,\mu) = -\rho^2 Q_1(\mu) - R\frac{3}{2}(1+\mu)\left(1-e^{-\frac{1}{2}\rho(1-\mu)}\right) + \mathcal{O}\left(R^2\right) \tag{3.46}$$

By taking the $\rho \to 0$ limit of Eqn. 3.46, we can calculate the overlap between these two expansions. The result is given in Eqn. 3.47.

$$\psi_{\text{overlap}}(r,\mu) = \frac{r}{8}(12-8r)\,Q_1(\mu) + \frac{rR}{8}\left(3rQ_2(\mu) - 3rQ_1(\mu)\right) + \mathcal{O}\left(R^2\right) \tag{3.47}$$

Eqns. 3.47, 3.46, and 3.45 can be combined to form a uniformly valid approximation:

$$\psi_{\text{uniform}}(r,\mu) = \psi(r,\mu) - \psi_{\text{overlap}}(r,\mu) + \frac{\Psi(rR,\mu)}{R^2} + \mathcal{O}\left(R^2\log R\right) \tag{3.48}$$

Due to the $e^{-\frac{1}{2}\rho(1-\mu)}$ term, we cannot use the simple expression for C_D (Eqn. 2.25). Instead, we must use the full set of Eqns. 2.10, 2.23, and 2.24. After completing this procedure, we obtain a new result for C_D, given by Eqn. 3.49.

$$\begin{aligned} C_D &= \frac{6\pi}{R}\Bigg(\frac{e^{-2R}}{320R^3}\Big(40e^R\left(1728+1140R+335R^2+56R^3+6R^4\right) - 60R(1+R) \\ &\quad +e^{2R}\big(-69120+23580R-2420R^2+20(10+\pi)R^3+10(18-\pi)R^4-8R^5 \\ &\quad -3R^6\big)\Big) - \frac{e^{-R/2}\pi I_1(R/2)}{4R}\Bigg) + \mathcal{O}\left(R^1\right) \end{aligned} \tag{3.49}$$

This result is plotted in Figure 3.11. Asymptotically, it agrees with the matched asymptotics predictions to $\mathcal{O}(1)$, as it must, and reproduces the $3/8R$ "Oseen" term.

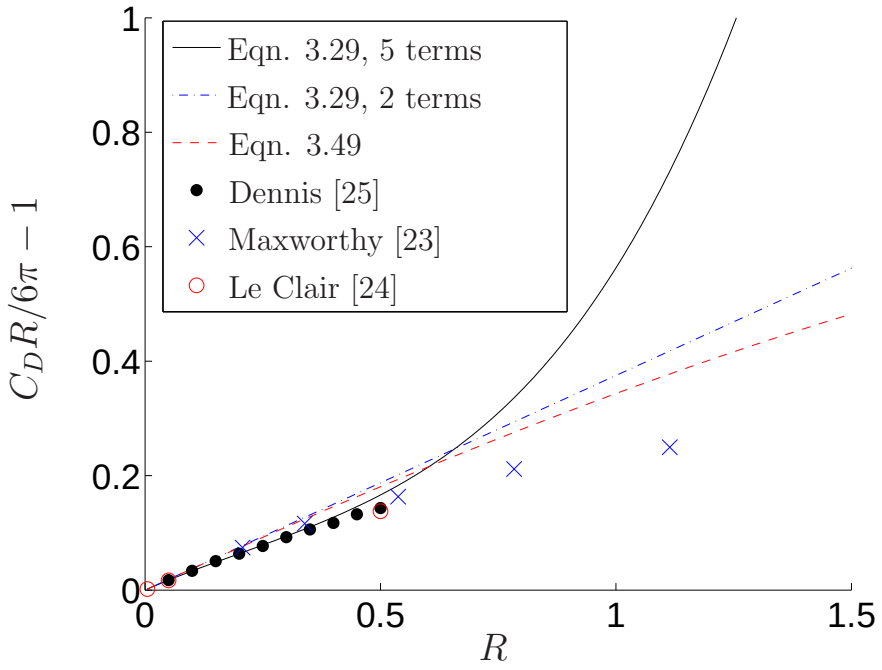

Figure 3.11: Drag on a Sphere — Experiment vs. Theory

As R increases, however, the uniform calculation becomes superior to the first two terms of the matched asymptotic C_D. Although it is a much higher order solution than either of the other two results, we show the full matched asymptotics prediction for comparison.

Chapter 4

The Renormalization Group Applied to Low R Flow

4.1 Introduction to the Renormalization Group

In 1961, Lagerstrom proposed the first of a number of "model problems", ordinary differential equations which exhibited many of the same asymptotic features as the low Reynolds number problems. They were used to study and develop the theory of matched asymptotic expansions. The mathematical solution of these problems is closely analogous to the actual solutions of the Navier-Stokes equations.

A review of these equations, and of their matched asymptotic solutions, is given in Lagerstrom[94]. The relevent models can be summarized by the following equation:

$$\frac{\mathrm{d}^2u}{\mathrm{d}x^2} + \frac{n-1}{x}\frac{\mathrm{d}u}{\mathrm{d}x} + u\frac{\mathrm{d}u}{\mathrm{d}x} + \delta\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 = 0 \tag{4.1}$$

This ODE is subject to the boundary conditions $u(\epsilon) = 1$, $u(\infty) = 0$. In this equation, n corresponds to the number of spatial dimensions ($n = 2$ for the cylinder, $n = 3$ for the sphere). $\delta = 0$ characterizes incompressible flow, and $\delta = 1$ corresponds to compressible flow. This equation is similar to the Navier-Stokes equations expressed in Oseen variables. There are fundamental differences between the structure of the

incompressible and compressible flow equations.

These model problems are posed in Hinch, albeit in terms of "Stokes" (rather than "Oseen") variables[28]. Hinch begins by examining the model describing incompressible flow past a sphere. He next examines incompressible flow past a cylinder, which he calls "A worse problem." Finally, he treats compressible flow past cylinder, which he dubs "A terrible problem."

These problems, which have historically been the proving ground of matched asymptotics, were recently solved using new techniques in two papers by Chen et al.[16, 17]. These papers solved two of these model problems using techniques based on the Renormalization Group (RG). These techniques afford both quantitative and methodological advantages over traditional matched asymptotics. The RG approach derives all of the subtle terms (e.g., $R^2 \log R$) which arise during asymptotic matching, demonstrating that origin of these terms lies in the need to correct flaws inherent in the underlying expansions. Moreover, RG does not require multiple rescalings of variables, and its results, while asymptotically equivalent to those of matched asymptotics, apply over a much larger range (e.g., they extend to higher R).

In particular, Chen et al. solved Hinch's first model, which describes incompressible flow past a sphere ($n = 3$, $\delta = 0$), as well as the model for both kinds of flow past a cylinder ($n = 2$, $\delta = 0, 1$)[16, 17]. In a notation consistent with Hinch, they termed these models the "Stokes-Oseen caricature" and the "terrible problem."

The dramatic success of the RG techniques in solving the model problems inspired their application to the original low Reynold number flow problems. That is our primary purpose here, as the low Reynold number problems are the traditional proving ground for new methodologies. We will show that the RG techniques perform well when applied to these problems. RG produces results superior to the predictions of matched asymptotics. More importantly, the RG calculations are considerably simpler than matched asymptotics, requiring half the work.

The utility of the RG approach is most easily seen through an example, which will also provide a framework for understanding the analysis presented in subsequent sections. Several pedagogical examples can also be found in the references (e.g., [17], [16], and [95]). We begin here with an analysis of the most complicated model problem, the "terrible problem," which caricatures compressible flow past a cylinder.

4.1.1 Detailed Analysis of the "Terrible Problem"

Although the "terrible problem," is solved in a paper by Chen et al., we re-examine it here in considerably more detail, as its solution is closely analogous to those of the low Reynold number flow problems. This switchback problem is exceptionally delicate[1], requiring the calculation of an infinite number of terms for the leading order asymptotic matching.

There are pitfalls and ambiguities in applying RG techniques, even to the "terrible problem," which while terrible, is considerably simpler than the real low Reynolds number problems. Understanding these subtleties in this simpler context provides essential guidance when attacking the Navier-Stokes' equations.

We want to solve the ODE given in Eqn. 4.2a, subject to the boundary conditions 4.2b. This equation can be derived from Eqn. 4.1 by setting $n = 2$, $\delta = 1$, and transforming to the "Stokes" variables, $r = x/\epsilon$. Unlike Eqn. 4.1, Eqn. 4.2 is obviously a singular perturbation in ϵ, which has been removed from the boundary conditions. The last term in the equation vanishes when $\epsilon = 0$.

$$\frac{d^2u(r)}{dr^2} + \frac{1}{r}\frac{du(r)}{dr} + \left(\frac{du(r)}{dr}\right)^2 + \epsilon u(r)\frac{du(r)}{dr} = 0 \tag{4.2a}$$

$$u(1) = 0, \quad u(r = \infty) = 1 \tag{4.2b}$$

This problem cannot be solved exactly, although numerical solution is straightforward. Trouble arises due to the *boundary layer*[2] located near $r = \infty$. RG analysis requires

[1] Hinch notes, "It is unusual to find such a difficult problem ..."[28]

[2] A boundary layer is a region of rapid variation in the solution, $y(t)$

that we work in the "inner" variable for our approximation to capture the correct behavior near the boundary layer. This requirement may also be qualitatively motivated by arguing that one must choose coordinates to "stretch out" the boundary layer so that it can be well characterized by our approximate solution.

To determine the appropriate change of variables, we need to analyze Eqn. (4.2) using a *dominant balance* argument[15]. As it stands, the first three terms of Eqn. (4.2a) will dominate, since ϵ is small. The rescaling $x = \epsilon r$ yields inner[3] Eqn. (4.3). This, of course, is the same equation originally given by Lagerstrom (Eqn. 4.1).

$$\frac{d^2u(x)}{dx^2} + \frac{1}{x}\frac{du(x)}{dx} + \left(\frac{du(x)}{dx}\right)^2 + u(x)\frac{du(x)}{dx} = 0 \tag{4.3a}$$

$$u(\epsilon) = 0, \quad u(x = \infty) = 1 \tag{4.3b}$$

The next step in the RG solution is to begin with the ansatz that the solution to Eqn. (4.3) can be obtained from a perturbation expansion (Eqn. 4.4). We fully expect this ansatz to fail, since we have a singular perturbation in our ODE. We therefore refer to this starting point as the *naïve* perturbation expansion.

$$u(x) = u_0(x) + \epsilon u_1(x) + \epsilon^2 u_2(x) + \mathcal{O}(\epsilon^3) \tag{4.4}$$

Collecting powers of ϵ, we obtain differential equations for $u_0(x)$, $u_1(x)$, etc:

$$\mathcal{O}\left(\epsilon^0\right) \quad : \quad \frac{u_0^{'}(x)}{x} + u_0(x)u_0^{'}(x) + u_0^{'}(x)^2 + u0^{''}(x) = 0 \tag{4.5}$$

$$\mathcal{O}\left(\epsilon^1\right) \quad : \quad u_1u_0^{'} + \frac{u_1^{'}}{x} + u_0u_1^{'} + 2u_0^{'}u_1^{'} + u_1^{''} = 0 \tag{4.6}$$

$$\mathcal{O}\left(\epsilon^2\right) \quad : \quad u_2u_0^{'} + u_1^{'}u_1 + u_0u_2^{'} + (u_1^{'})^2 + 2u_0^{'}u_2^{'} + \frac{u_2^{'}}{x} + u_2^{''} = 0 \tag{4.7}$$

$\mathcal{O}(\epsilon^0)$ solution

The first complication of the terrible problem arises when we attempt to solve Eqn. 4.5, a nonlinear ODE. Although one solution — $u_0(x) = A_0$ — is seen by inspection,

[3]Here we use "inner" in the usual sense[15].

an additional integration constant is not forthcoming, and our solution to the $\mathcal{O}(\epsilon^0)$ problem cannot satisfy both of the boundary conditions (Eqn. 4.3b). The resolution to this quandary is simple: Ignore the problem and it will go away; continue constructing the naïve solution as if $u_0(x) = A_0$ were wholly satisfactory. The qualitative idea is that the $\mathcal{O}(\epsilon^0)$ solution is the uniform field which we have far from any disturbance source. Why is this acceptable?

The RG method is robust against shortcomings in the naïve expansion. We know that singular perturbation problems cannot be solved by a single perturbation expansion. We therefore expect problems, such as secular behavior, to arise in our solution for the naïve expansion. RG techniques can be used to remove these flaws from the perturbative solution, turning it into a uniformly valid approximation[16]. It does not matter whether these defects arise from an incomplete solution for $u_0(x)$, the intrinsic structure of the equation, or a combination of the two. To solve the terrible problem (and later the low Reynolds number problems), we must exploit this flexibility.

For subsequent calculations, there are two ways to proceed. First, we may retain A_0 as an arbitrary constant, one which will ultimately be renormalized in the process of calculating a uniformly valid approximation. Alternatively, we may set $A_0 = 1$, satisfying the boundary condition at $x = \infty$[4]. This unconventional approach to the RG calculation effectively shifts the freedom that usually comes with the $\mathcal{O}(\epsilon^0)$ constants of integration into the $\mathcal{O}(\epsilon^1)$ solution. This artifice greatly simplifies subsequent calculations, and is invaluable in treating the Navier-Stokes equations. Moreover, these two approaches are equivalent, as we now show.

$\mathcal{O}(\epsilon^1)$ solution

If $u_0(x) = A_0$, Eqn. 4.6 simplifies to Eqn. 4.8.

$$\frac{d^2u_1}{dx^2} + \left(\frac{1}{x} + A_0\right)\frac{du_1}{dx} = 0 \tag{4.8}$$

[4]Meeting the boundary condition at $x = \epsilon$ results only in the trivial solution $u_0(x) = 0$

The solution is: $u_1(x) = B_0 + B_1 e_1(A_0 x)$, where $e_n(x) \equiv \int_x^\infty e^{-t} t^{-n} \mathrm{d}t$. Notice that the first term is linearly dependant on the $u_0(x)$ solution. There are many opinions regarding how to utilize this degree of freedom[96, 97]. In our approach, one is free to choose the homogeneous solutions of $u_0, u_1,$ etc. for convenience. The only constraint[5] is that the "naïve" solution (Eqn. 4.4) must have a sufficient number of integration constants to meet the boundary conditions. In this example, that means two constants of integration.

Different choices of particular solutions will ultimately result in different approximate solutions to the ODE. However, all of these solutions will agree within the accuracy limitations of the original approximation (in this case the naïve expansion). This can be shown explicitly. In this example, as in the low Reynolds number problems, we choose a particular solution which simplifies subsequent calculations. Setting $B_0 = 0$ (note that this is *not* the same as a redefinition of the constant A_0), we obtain the solution:

$$u(x) = A_0 + \underbrace{\epsilon B_1 e_1(A_0 x)}_{\text{divergent as } x \to 0} + \mathcal{O}(\epsilon^2) \tag{4.9}$$

The second term in Eqn. 4.9 diverges logarithmically as $x \to 0$. One may argue that this divergence is irrelevant, since the range of the original variable is $r \in [1, \infty)$, and numerical solutions demonstrate that the solutions to Eqn. 4.2 in $[1, \infty)$ diverge when extended to $r < 1$. But the argument that the divergence in Eqn. 4.9 is an intrinsic part of the solution (and therefore should not be considered problematic) is incorrect. Although the original variable, r, is limited to $r \in [1, \infty)$, the transformed variable, $x = \epsilon r$, has the range $x \in [0, \infty)$. This occurs because there are no restrictions on the lower limit of ϵ. The divergence exhibited by the second term of Eqn. 4.9 must be removed via renormalization in order to turn the flawed naïve solution into a uniformly valid approximation.

This divergence arises for two reasons. First, we are perturbing about an $\mathcal{O}(\epsilon^0)$

[5] Of course the solution must also satisfy the governing equation.

solution which is deficient; it is missing the the second integration constant (and concomitant fundamental solution). More fundamentally, Eqn 4.4 attempts to solve a singular perturbation problem with a regular expansion, an approach which must fail. The RG technique solves these problems by restructuring the naïve expansion and eliminating the flaws in $u_0(x)$.

Although A_0 is simply a constant of integration when $\epsilon = 0$, it must be modified when $\epsilon \neq 0$. We will absorb the divergences into a modification, or renormalization, of the constant of integration A_0. Formally, one begins by "splitting" the secular terms, replacing $e_1(A_0 x)$ by $e_1(A_0 x) - e_1(A_0\tau) + e_1(A_0\tau)$, where τ is an arbitrary position. This results in Eqn. 4.10:

$$u(x) = A_0 + \epsilon B_1(e_1(A_0 x) - e_1(A_0\tau) + e_1(A_0\tau)) + \mathcal{O}(\epsilon^2) \tag{4.10}$$

Since τ is arbitrary, it can be chosen such that $e_1(A_0 x) - e_1(A_0\tau)$ is non-secular (for a given x). The divergence is now contained in the last term of Eqn. (4.10), and is exhibited as a function of τ.

It is dealt with by introducing a multiplicative renormalization constant, $Z_1 = 1 + \sum_{i=1}^{\infty} a_i(\tau)\epsilon^i$, and then renormalizing A_0 as $A_0 = Z_1 A_0(\tau)$.[6] The coefficients $a_i(\tau)$ can then be chosen[7] order by order so as to eliminate the secular term in Eqn. (4.10). Subsituting, and choosing a_1 to eliminate the final term of Eqn. 4.10, we obtain

$$u(x) = A_0(\tau) + \epsilon B_1(e_1(A_0(\tau)x) - e_1(A_0(\tau)\tau)) + \mathcal{O}(\epsilon^2) \tag{4.11}$$

Where a_1 satisfies

$$a_1(\tau) = \frac{-B_1 e_1(\tau A_0(\tau)(1 + \sum_{i=1}^{\infty} a_i(\tau)\epsilon^i))}{A_0(\tau)} \tag{4.12}$$

Note that to obtain Eqn. 4.11 we needed to expand e_1 about $\epsilon = 0$. Unusually in this equation, the renormalized constant ($A_0(\tau)$) appears in the argument of the

[6] A_0 is the only constant which can be renormalized to remove the divergences, as B_1 is proportional to the secular terms

[7] Note that the coefficients *must* also be independent of x

exponential integral; this complicates the calculation. We will later show how to avoid this problem by restructuring our calculations.

Qualitatively, the idea underlying Eqn. 4.11 is that boundary conditions far away (from $x = \epsilon$) are unknown to our solution at $x \gg \epsilon$, so that A_0 is undetermined at $x = \tau$. RG determines A_0 in this regime through the renormalization constant Z_1 (which depends on τ). Afterward there will be new constants which can be used to meet the boundary conditions.

The RG condition states that the solution $u(x)$ cannot depend on the arbitrary position τ. This requirement can be implemented in one of two ways. First, since $\partial_\tau u(x) = 0$, apply ∂_τ to the RHS of Eqn. 4.11 and set the result equal to zero:

$$A_0'(\tau) + \epsilon B_1 \left(\frac{e^{-A_0(\tau)\tau}}{\tau} + \frac{A_0'(\tau)}{A_0(\tau)} \left(e^{-A_0(\tau)\tau} - e^{-A_0(\tau)x} \right) \right) + \mathcal{O}(\epsilon^2) = 0 \tag{4.13}$$

The next step in RG is to realize Eqn. 4.13 implies that $A_0'(\tau) \sim \mathcal{O}(\epsilon)$.[8] Retaining only terms of $\mathcal{O}(\epsilon)$, we obtain:

$$\frac{dA_0(\tau)}{d\tau} + \epsilon B_1 \left(\frac{e^{-A_0(\tau)\tau}}{\tau} \right) + \mathcal{O}(\epsilon^2) = 0 \tag{4.14}$$

In principle, we simply solve Eqn. 4.14 for $A_0(\tau)$. Unfortunately, that is not possible, due to the presence of $A_0(\tau)$ in the exponential. This complication also occurs in other switchback problems, as well as in the low Reynolds number problems. Eqn. 4.14 can be solved by an iterative approach: Initially set $\epsilon = 0$, and solve for $A_0(\tau) = \alpha_0$, a constant. Next substitute this result into the $\mathcal{O}(\epsilon)$ term in Eqn. 4.14, solving for $A_0(\tau)$ again:

$$A_0(\tau) = \alpha_0 + \epsilon B_1 e_1(\alpha_0 \tau) \tag{4.15}$$

In this solution, we have a new integration constant, α_0. Having obtained this result, we again must exploit the arbitrary nature of τ. Setting $\tau = x$, and subsituting into

[8] Woodruff argues that these terms vanish because the arbitrariness of x can be used to choose $x = \tau$[98]

Eqn. 4.11, we obtain:

$$u(x) = \alpha_0 + \epsilon B_1 e_1(\alpha_0 x) + \mathcal{O}(\epsilon^2) \tag{4.16}$$

But this is identical to the original solution (Eqn. 4.8)! What have we accomplished? This renormalized result is guaranteed to be a uniformly valid result, for $\forall x$. The renormalization procedure ensures that the logarithmic divergence in Eqn. 4.16 is required by the solution, and is *not* an artifact of our approximations. Obtaining the same answer is a consequence of solving Eqn. 4.13 iteratively. Had we been able to solve that equation exactly, this disconcerting coincidence would have been avoided.

We obtain the final solution to Eqn. 4.2a by applying the boundary conditions (Eqn. 4.3b) to Eqn. 4.16: $\alpha_0 = 1$, $B_1 = -1/(\epsilon e_1(\epsilon))$. Lastly, we undo the initial change of variables ($r = x/\epsilon$), yielding the result given in Eqn. 4.17. As shown in Chen et al., this is a spectacular approximate solution[16].

$$u(r) = 1 - \frac{e_1(r\epsilon)}{e_1(\epsilon)} + \mathcal{O}\left(\epsilon^2\right) \tag{4.17}$$

Furthermore, if we expand the coefficient $B_1 = -1/(\epsilon e_1(\epsilon))$ for $\epsilon \to 0^+$, $B_1(\epsilon)/\epsilon \sim -1/\ln(1/\epsilon) - \gamma/\ln^2(1/\epsilon)$. These logarithmic functions of ϵ are exactly those which are required by asymptotic matching! These "unexpected" orders in ϵ make the solution of this problem via asymptotic matching very difficult. They must be deduced and introduced order by order, so as to make matching possible. In the RG solution, they are seen to arise naturally as a consequence of the term $1/e_1(\epsilon)$.

There are several other equivalent ways to structure this calculation. It is worthwhile to examine these (and to demonstrate their equivalence), in order to streamline our approach for the low Reynolds number problems.

The first variation occurs in how we apply the RG condition. Rather than applying ∂_τ to Eqn. 4.11, we may also realize that the original constants of integration, $A_0 = Z_1(\tau)A_0(\tau)$, must be independent of τ. Hence the "alternative" RG equation:

$$\frac{\partial A_0}{\partial \tau} = \frac{\partial (Z_1(\tau)A_0(\tau))}{\partial \tau} = 0$$

Substituting $Z_1 = 1+\epsilon\left(-B_1 e_1(\tau A_0(\tau)(1+\sum_{i=1}^{\infty} a_i(\tau)\epsilon^i))\right)/A_0(\tau)+\mathcal{O}\left(\epsilon^2\right)$, one obtains:

$$A_0'(\tau) + \epsilon B_1 \left(\frac{e^{-A_0(\tau)\tau}}{\tau} + \frac{A_0'(\tau)}{A_0(\tau)} e^{-A_0(\tau)\tau} \right) + \mathcal{O}\left(\epsilon^2\right) = 0 \tag{4.18}$$

Because this implies $A_0'(\tau) \sim \mathcal{O}\left(\epsilon^1\right)$, Eqn. 4.18 simplifies to Eqn. 4.14 (to within $\mathcal{O}\left(\epsilon^2\right)$), and these two methods of implementing the RG condition are equivalent.

In addition to this dichotomous implementation of the RG condition, there is yet another way to structure the analysis from the outset: We set $A_0 = 1$ in the zeroth order solution, and rely on the robustness of the RG approach to variations in our perturbative solution. With this $u_0(x)$ solution, there is no longer any freedom in our choice of $u_1(x)$ integration constants — both are needed to meet boundary conditions. In this approach, our naïve perturbative solution is:

$$u(x) = 1 + \epsilon(B_0 + \underbrace{B_1 e_1(x)}_{\text{divergent}}) + \mathcal{O}\left(\epsilon^2\right) \tag{4.19}$$

Proceeding as before, replace $e_1(x)$ by $e_1(x) - e_1(\tau) + e_1(\tau)$:

$$u(x) = 1 + \epsilon\left(B_0 + B_1\left(e_1(x) - e_1(\tau) + e_1(\tau)\right)\right) + \mathcal{O}\left(\epsilon^2\right)$$

Again introduce renormalization constants ($Z_1 = 1+\sum_{i=1}^{\infty} a_i(\tau)\epsilon^i$, $Z_2 = 1+\sum_{i=1}^{\infty} b_i(\tau)\epsilon^i$), and renormalize B_0, B_1 as $B_0 = Z_1 B_0(\tau)$ and $B_1 = Z_2 B_1(\tau)$. In fact, only B_0 needs to be renormalized, as the B_1 term multiplies the secular term and consequently cannot absorb that divergence. This can be seen systematically by attempting to renormalize both variables. With an appropriate choice of coefficients, $a_1 = -B_1(\tau)e_1(\tau)$ and $b_1 = 0$, the final term in the last equation is eliminated. $b_1 = 0$ demonstrates that B_1 does not need to be renormalized at $\mathcal{O}\left(\epsilon^1\right)$. The resulting equation is given in Eqn. 4.20.

$$u(x) = 1 + \epsilon\left(B_0(\tau) + B_1(\tau)\left(e_1(x) - e_1(\tau)\right)\right) + \mathcal{O}\left(\epsilon^2\right) \tag{4.20}$$

We did not actually need to determine a_1 or b_1 in order to write the above equation; it could have been done by inspection. Determination of these quantities is useful

for two reasons. First, it helps us see which secular terms are being renormalized by which integration constants. Secondly, it allows the second implementation of the RG condition which was described above. This can sometimes simplify calculations.

Using the first implementation (requiring $\partial_\tau u(x) = 0$), and using Eqn. 4.20, we obtain:

$$B_0^{'}(\tau) + B_1^{'}(\tau)\left(e_1(x) - e_1(\tau)\right) + B_1(\tau)\frac{e^{-\tau}}{\tau} = \mathcal{O}\left(\epsilon^1\right) \tag{4.21}$$

This can only be true $\forall x$ if $B_1^{'}(\tau) = 0$, or $B_1(\tau) = \beta_2$, a constant (as expected). Knowing this, we solve for $B_0(\tau) = \beta_1 + \beta_2 e_1(\tau)$. Substituting this result into Eqn. 4.20, and setting $\tau = x$, we obtain the renormalized solution:

$$u(x) = 1 + \epsilon\left(\beta_1 + \beta_2 e_1(x)\right) \tag{4.22}$$

The boundary conditions in Eqn. 4.3b are satisfied if $\beta_1 = 0$ and $\beta_2 = -1/(\epsilon e_1(\epsilon))$. Returning to the original variable ($r = x/\epsilon$), we obtain:

$$u(r) = 1 - \frac{e_1(r\epsilon)}{e_1(\epsilon)} + \mathcal{O}\left(\epsilon^2\right) \tag{4.23}$$

This is identical to Eqn. 4.17, demonstrating the equivalence of these calculations. The latter method is preferable, as it avoids the nonlinear RG equation (Eqn. 4.14). We will use this second approach for analyzing the low Reynolds number problems.

The RG analysis has shown us that the logarithmic divergences present in Eqn. 4.9 are an essential component of the solution, Eqn. 4.23. However, we must work to $\mathcal{O}\left(\epsilon^2\right)$ in order to see the true utility of RG and to understand all of the nuances of its application.

$\mathcal{O}\left(\epsilon^2\right)$ solution

We base our treatment of the $\mathcal{O}\left(\epsilon^2\right)$ on the second analysis presented above. Through $\mathcal{O}\left(\epsilon^1\right)$, the naïve solution is: $u_0(x) = 1$, $u_1(x) = B_0 + B_1 e_1(x)$. Substituting into Eqn. 4.7, we obtain the governing equation for $u_2(x)$:

$$u_2^{''} + \left(1 + \frac{1}{x}\right)u_2 = \frac{B_0 B_1 e^{-x}}{x} - \frac{B_1^2 e^{-2x}}{x^2} + \frac{B_1^2 e^{-x} e_1(x)}{x} \tag{4.24}$$

This has the same homogeneous solution as $u_1(x)$, $u_2^{(h)}(x) = C_0 + C_1 e_1(x)$. A particular solution is:

$$u_2^{(p)}(x) = -B_1 B_0 e^{-x} + 2B_1^2 e_1(2x) - \frac{1}{2} B_1^2 e_1^2(x) - B_1^2 e^{-x} e_1(x)$$

As discussed previously, we are free to choose C_0, C_1 to simplify subsequent calculations. The constants B_0, B_1 are able to meet the boundary conditions, so there is no need to retain the $\mathcal{O}(\epsilon^2)$ constants: We choose $C_0 = 0$, $C_1 = 0$. In this case, the differing differing choices of C_0, C_1 correspond to a redefinition of B_0, B_1 plus a change of $\mathcal{O}(\epsilon^3)$, i.e. $\tilde{B}_0 = B_0 + \epsilon C_0$.[9] Our naïve solution through $\mathcal{O}(\epsilon^2)$ is thus:

$$\begin{aligned} u(x) &= 1 + \epsilon \left(B_0 + \underline{B_1 e_1(x)} \right) + \\ &\quad \epsilon^2 \left(-B_1 B_0 e^{-x} + \underline{2B_1^2 e_1(2x)} - \underline{\underline{\frac{1}{2} B_1^2 e_1^2(x)}} - \underline{B_1^2 e^{-x} e_1(x)} \right) + \mathcal{O}\left(\epsilon^3\right) \end{aligned} \tag{4.25}$$

The underlined terms in this expression are divergent as $x \to 0$; the doubly underlined term is the most singular ($\sim ln(x)^2$). RG can be used to address the divergences in Eqn. 4.25. However, there is a great deal of flexibility in its implementation; while most tactics yield equivalent approximations, there are significant differences in complexity. We now explore all of the organizational possibilities in the terrible problem, an exercise which will subsequently guide us through the low Reynolds number calculations.

The first possibility is to treat only the most secular term at $\mathcal{O}(\epsilon^2)$. The doubly underlined term dominates the divergent behavior, and contains the most important information needed for RG to construct a uniformly valid approximation. The approximation reached by this approach is necessarily inferior to those obtained utilizing additional terms. However it is nonetheless valid and useful, and eliminating most of the $\mathcal{O}(\epsilon^2)$ terms simplifies our calculations.

Discarding all $\mathcal{O}(\epsilon^2)$ terms except the doubly underlined term, we begin the calculation in the usual manner, but come immediately to the next question: Ought we

[9] This was not true at the previous order.

replace $e_1^2(x)$ by $e_1^2(x)-e_1^2(\tau)+e_1^2(\tau)$ or by $(e_1(x)-e_1(\tau))^2+2e_1(x)e_1(\tau)-e_1^2(\tau)$? Each option eliminates the divergence in x, replacing it with a divergence in τ. Both merit consideration. Beginning with the latter, the renormalized perturbative solution is:

$$\begin{aligned} u(x) &= 1+\epsilon\left(B_0(\tau)+B_1(\tau)\left(e_1(x)-e_1(\tau)\right)\right)-\epsilon^2\left(\frac{1}{2}B_1(\tau)^2\left(e_1(x)-e_1(\tau)\right)^2\right) \\ &\quad +\epsilon^2\left(\text{less divergent terms}\right)+\mathcal{O}\left(\epsilon^3\right) \end{aligned} \tag{4.26}$$

Applying the RG condition ($\partial_\tau u(x)=0$) results in a lengthy differential equation in τ. Because we want our solution to be independent of x, we group terms according to their x dependence. Recognizing that $B_1^{'}(\tau)\sim\mathcal{O}\left(\epsilon^1\right)$, $B_0^{'}(\tau)\sim\mathcal{O}\left(\epsilon^1\right)$, and working to $\mathcal{O}\left(\epsilon^3\right)$, we obtain two equations which must be simultaneously satisfied:

$$B_1^{'}(\tau)-\frac{\epsilon e^{-\tau}B_1^2(\tau)}{\tau} = \mathcal{O}\left(\epsilon^3\right) \tag{4.27a}$$

$$\frac{e^{-\tau}\left(B_1(\tau)+\epsilon B_1^2(\tau)e_1(\tau)\right)}{\tau}-e_1(\tau)B_1^{'}(\tau)+B_0^{'}(\tau) = \mathcal{O}\left(\epsilon^3\right) \tag{4.27b}$$

Eqn. 4.27a has the solution

$$B_1(\tau)=\frac{1}{\beta_1+\epsilon e_1(\tau)}+\mathcal{O}\left(\epsilon^2\right)$$

Substituting this result into Eqn. 4.27b, and solving, we obtain the result

$$B_0(\tau)=\beta_0+\frac{\ln\left(\beta_1+\epsilon e_1(\tau)\right)}{\epsilon}+\mathcal{O}\left(\epsilon^2\right)$$

Both β_0 and β_1 are constants of integration which can be later used to meet the boundary conditions. Substituting these solutions into Eqn. 4.26, setting $\tau=x$, disregarding terms of $\mathcal{O}\left(\epsilon^2\right)$ and higher we obtain the renormalized solution:

$$u(x)=1+\epsilon\left(\beta_0+\frac{\ln\left(\beta_1+\epsilon e_1(x)\right)}{\epsilon}\right)+\mathcal{O}\left(\epsilon^2\right) \tag{4.28}$$

Choosing β_0 and β_1 to satisfy Eqn. 4.3b, results in Eqn. 4.29.

$$u(x)=\ln\left(e+\frac{(1-e)\,e_1(x)}{e_1(\epsilon)}\right)+\mathcal{O}\left(\epsilon^2\right) \tag{4.29}$$

Expressing this in the original variable ($r = x/\epsilon$), results in the final answer (Eqn. 4.30).

$$u(r) = \ln\left(e + \frac{(1-e)\, e_1(\epsilon r)}{e_1(\epsilon)}\right) + \mathcal{O}\left(\epsilon^2\right) \tag{4.30}$$

This is the solution previously obtained by Chen et al., albeit with a typographical error corrected[16]. We will now revisit this analysis, using the alternative "splitting" of the most secular term in Eqn. 4.25, but not yet considering less secular (or non-secular) terms of $\mathcal{O}(\epsilon^2)$.

If we replace replace $e_1^2(x)$ in Eqn. 4.25 by $e_1^2(x) - e_1^2(\tau) + e_1^2(\tau)$, we obtain the new naïve expansion given by Eqn. 4.31.

$$\begin{aligned} u(x) &= 1 + \epsilon\left(B_0(\tau) + B_1(\tau)\left(e_1(x) - e_1(\tau)\right)\right) - \epsilon^2\left(\frac{1}{2}B_1(\tau)^2\left(e_1^2(x) - e_1^2(\tau)\right)\right) \\ &\quad + \epsilon^2\,(\text{less divergent terms}) + \mathcal{O}\left(\epsilon^3\right) \end{aligned} \tag{4.31}$$

We now repeat the same calculations:

1. Apply the RG condition ($\partial_\tau u(x) = 0$).
2. Group the resulting equation according to x dependence. This will result in two equations which must be satisfied independently.
3. Discard terms of $\mathcal{O}(\epsilon^3)$, observing that $B_0'(\tau), B_1'(\tau)$ must be of $\mathcal{O}(\epsilon^1)$.
4. Solve these differential equations simultaneously for $B_0(\tau), B_1(\tau)$.
5. Substitute these solutions into the original equation (i.e. Eqn. 4.31), and set $\tau = x$.
6. Choose the integration constants in this result to satisfy Eqn. 4.3b.
7. Obtain the final solution by returning to the original variable, $r = x/\epsilon$.

For Eqn. 4.31, steps 1 - 4 result in the following solutions for our renormalized constants: $B_1(\tau) = \beta_1 + \mathcal{O}(\epsilon^2)$, $B_0(\tau) = \beta_0 + \beta_1 e_1(\tau) - \frac{\epsilon\beta_1^2 e_1^2(\tau)}{2} + \mathcal{O}(\epsilon^2)$. Completing

step 5, we obtain the renormalized result:

$$u(x) = 1 + \epsilon\left(\beta_0 + \beta_1 e_1(x)\right) - \epsilon^2 \frac{\beta_1^2 e_1^2(x)}{2} + \mathcal{O}\left(\epsilon^2\right) \tag{4.32}$$

This is identical to our starting point, Eqn. 4.25 (retaining only the most secular terms). This should no longer be surprising, as we observed the same phenomena in the $\mathcal{O}\left(\epsilon^1\right)$ analysis (Eqn. 4.16). However, it is worth noticing that we obtained two different results (Eqns. 4.29, 4.32) depending on how we structured our RG calculation. This apparent difficulty is illusory, and the results are equivalent: Expanding Eqn. 4.28 for small ϵ reproduces Eqn. 4.32. Here, as in previous cases, we are free to structure the RG calculation for convenience. This easiest calculation is the second approach — in which only one constant of integration is actually renormalized — and our renormalized result is the same as our naïve starting point.

This simplified analysis (considering only the most secular terms) illustrates some of the pitfalls which can arise in applying RG to switchback problems. However, we must finish the $\mathcal{O}\left(\epsilon^2\right)$ analysis by considering all terms in Eqn. 4.25 to understand the final nuances of these problems. There is a new complication when we attempt to renormalize all terms of Eqn. 4.25: The final term, $-B_1^2 e^{-x} e_1(x)$, has the same kind of "splitting" ambiguity which we encountered in dealing with the doubly underlined term.

We introduce our arbitrary position variable, τ, which we want to choose so as to eliminate the secular term in x by replacing it with a divergence in τ. In many cases, it is clear how to deal with the secular term. For example, a linear divergence — x — can be replaced with $x - \tau + \tau$. The final τ will be absorbed into the renormalized constants of integration, and the $x - \tau$ term (which is now considered non-secular), will ultimately disappear after renormalization. However the term $-B_1^2 e^{-x} e_1(x)$ is confusing. As seen above, there are two ways to "split" the $B_1^2 e_1^2(x)/2$ term. There are *four* different ways to split $e^{-x} e_1(x)$. It may be replaced by any of the following:

1. $\left(e^{-x} - e^{-\tau}\right) e_1(x) + e^{-\tau} e_1(x)$

2. $e^{-x}e_1(x) - e^{-\tau}e_1(\tau) + e^{-\tau}e_1(\tau)$

3. $\left(e^{-x} - e^{-\tau}\right)\left(e_1(x) - e_1(\tau)\right) + e^{-\tau}e_1(x) + e^{-x}e_1(x) - e^{-\tau}e_1(\tau)$

4. $e^{-x}\left(e_1(x) - e_1(\tau)\right) + e^{-x}e_1(\tau)$

All four of these options "cure" the divergent term (i.e. the secular term will vanish when we subsequently set $\tau = x$), and are equal to $e^{-x}e_1(x)$. If handled properly, any of these options can lead to a valid renormalized solution. However, we will show that the fourth and final option is most natural, and results in the simplest algebra.

How do we choose? The first consideration is subtle: *The overall renormalized perturbative result must satisfy the governing equation (Eqn. 4.5) independently for each order in* ϵ. How we renormalize the $\mathcal{O}(\epsilon^1)$ divergences (Eqn. 4.20) has implications for $\mathcal{O}(\epsilon^2)$ calculations. For example, in $\mathcal{O}(\epsilon^1)$ renormalization, there is an important difference between Eqn. 4.20 and Eqn. 4.10. The former has the additional term $-\epsilon B_1(\tau)e_1(\tau)$. This term requires the presence of an additional $\mathcal{O}(\epsilon^2)$ term: $\epsilon^2 e^{-x}B_1^2(\tau)e_1(\tau)$. Without this term the $\mathcal{O}(\epsilon^2)$ renormalized solution will not satisfy Eqn. 4.7, and the renormalization procedure will yield an incorrect solution. We were able to gloss over this before because we were considering only the most secular term at $\mathcal{O}(\epsilon^2)$.

Inspecting the four possible splittings enumerated above, we see that only the last two options provide the necessary $\epsilon^2 e^{-x}B_1^2(\tau)e_1(\tau)$ term, and can satisfy Eqn. 4.7 without contrivances.[10] In examining both of these options, we split the $e_1^2(x)$ term for simplicity, as in the derivation of Eqn. 4.32.[11] Considering the third option first, our

[10]The first two options *can* satisfy the governing equation *if* we carefully choose a different homogeneous solution at $\mathcal{O}(\epsilon^2)$. With the proper non-zero choice of C_0 and C_1 we can use the first two splittings enumerated, and they will result in an equivalent RG solution.

[11]In principle, each of the possible $\mathcal{O}(\epsilon^1)$ splittings could be paired with all possibilities at $\mathcal{O}(\epsilon^2)$, resulting in eight total possibilities.

renormalized perturbation solution becomes:

$$\begin{aligned} u(x) &= 1+\epsilon\left(B_0(\tau)+B_1(\tau)\left(e_1(x)-e_1(\tau)\right)\right)+\epsilon^2\Big(-B_1(\tau)B_0(\tau)e^{-x}- \\ &\quad B_1^2(\tau)\left(e^{-x}-e^{-\tau}\right)\left(e_1(x)-e_1(\tau)\right)-\frac{1}{2}B_1(\tau)^2\left(e_1^2(x)-e_1^2(\tau)\right)+ \\ &\quad 2B_1^2(\tau)\left(e_1(2x)-e_1(2\tau)\right)\Big)+\mathcal{O}\left(\epsilon^3\right) \end{aligned} \tag{4.33}$$

As it must, this result satisfies Eqn. 4.7 to $\mathcal{O}\left(\epsilon^2\right)$. By applying the RG condition ($\partial_\tau u(x)=0$) to this equation, and grouping the resulting equation according to x dependence, we obtain a lengthy equation which can only be satisfied to $\mathcal{O}\left(\epsilon^3\right)$ $\forall x$ if:

$$B_1'(\tau)e^\tau = \epsilon B_1^2(\tau) \tag{4.34}$$

$$e^{2\tau}\tau B_0'(\tau) = e^{2\tau}\tau e_1(\tau)B_1'(\tau)-e^\tau B_1(\tau)-3\epsilon B_1^2(\tau)+e^\tau\epsilon B_1^2(\tau)e_1(\tau)-e^\tau\tau\epsilon B_1^2(\tau)e_1(\tau)$$

$$0 = \epsilon\Big(\epsilon B_1(\tau)+e^\tau\tau\epsilon B_0'(\tau)\Big)$$

Generally, no solution will exist, as we have two unknown functions and three differential equations. In this case, however, the first equation requires that:

$$B_1(\tau)=\frac{e^\tau}{-\epsilon+e^\tau\beta_1} \tag{4.35}$$

For this $B_1(\tau)$ solution, it is actually possible to satisfy the latter equations simultaneously to $\mathcal{O}\left(\epsilon^3\right)$: This occurs because the last equation is simply the lowest order of the second one.[12] There is another noteworthy point regarding the second part of Eqn. 4.34. In all previous calculations, we discarded terms like $\epsilon^2 B_0'(\tau)$, since $B_0'(\tau)$ and $B_1'(\tau)$ had to be of $\mathcal{O}\left(\epsilon^1\right)$. To solve these equations, however, $B_0'(\tau)$ can *not* be $\mathcal{O}\left(\epsilon^1\right)$ (although $B_1'(\tau)$ is). Solving for B_0,

$$B_0(\tau)=\beta_0-\int_\epsilon^\tau \frac{2\epsilon+e^\sigma\beta_1+e^\sigma\left(2\sigma-1\right)\epsilon e_1(\sigma)}{\sigma\left(\epsilon-e^\sigma\beta_1\right)}\mathrm{d}\sigma \tag{4.36}$$

This solution, while valid, is cumbersome. Consider instead the fourth possible "split" enumerated above. Eqn. 4.37 gives our renormalized perturbation solution,

[12] This can be seen explicitly by substituting Eqn. 4.35.

which satisfies Eqn. 4.7.

$$\begin{aligned} u(x) &= 1 + \epsilon\left(B_0(\tau) + B_1(\tau)\left(e_1(x) - e_1(\tau)\right)\right) + \epsilon^2\Big(- B_1(\tau)B_0(\tau)e^{-x} - \\ &\quad B_1^2(\tau)e^{-x}\left(e_1(x) - e_1(\tau)\right) - \frac{1}{2}B_1(\tau)^2\left(e_1^2(x) - e_1^2(\tau)\right) + \\ &\quad 2B_1^2(\tau)\left(e_1(2x) - e_1(2\tau)\right)\Big) + \mathcal{O}\left(\epsilon^3\right) \end{aligned} \tag{4.37}$$

Applying the RG condition ($\partial_\tau u(x) = 0$), and requiring that it be satisfied $\forall x$, we obtain the following solutions for $B_0(\tau)$, and $B_1(\tau)$:

$$B_1(\tau) = \beta_1 + \mathcal{O}\left(\epsilon^3\right) \tag{4.38a}$$

$$B_0(\tau) = \beta_0 + \beta_1 e_1(\tau) + \epsilon\left(-\frac{\beta_1^2 e_1^2(\tau)}{2} + 2\beta_1^2 e_1(2\tau)\right) + \mathcal{O}\left(\epsilon^3\right) \tag{4.38b}$$

Substituting these results into Eqn. 4.37 and setting $\tau = x$, we obtain the final RG result, given by Eqn. 4.39.

$$\begin{aligned} u(x) &= 1 + \epsilon\left(\beta_0 + \beta_1 e_1(x)\right) + \\ &\quad \epsilon^2\left(- \beta_1\beta_0 e^{-x} + 2\beta_1^2 e_1(2x) - \frac{1}{2}\beta_1^2 e_1^2(x) - \beta_1^2 e^{-x} e_1(x)\right) + \mathcal{O}\left(\epsilon^3\right) \end{aligned} \tag{4.39}$$

This is, of course, identical to our naïve staring point, a happenstance we have seen several times previously. It is worth noting that the renormalized solutions obtained using Eqns. 4.35 and 4.36 are asymptotically equivalent to Eqn. 4.39.

It may seem that we have needlessly digressed into the "terrible" problem. However, a clear-cut "best" strategy has emerged from our detailed exploration. Furthermore, we have identified — and resolved — a number of subtleties in the application of RG. Before applying these lessons to the problem of low Reynolds number flow past a cylinder, we summarize our conclusions.

The "best" strategy is the one used to derive Eqn. 4.39, a result which is identical to our naïve solution (Eqn. 4.25). First, transform to the inner equation. Solve the $\mathcal{O}\left(\epsilon^0\right)$ equation incompletely (obtaining just one constant of integration), which can then be set to satisfy the boundary condition at ∞. This "trick" necessitates retention

of integration constants at $\mathcal{O}(\epsilon^1)$, but results in computational simplifications (a non-linear RG equation) which are *essential* in dealing with the Navier-Stokes equations.

At $\mathcal{O}(\epsilon^2)$, the homogeneous solution are identical to those at $\mathcal{O}(\epsilon^1)$. Consequently, the $\mathcal{O}(\epsilon^2)$ integration constants need not be retained, as we can meet the boundary conditions with the $\mathcal{O}(\epsilon^1)$ constants. We just pick a convenient particular solution.

To apply RG to the terrible problem, we first "split" the secular terms. There are several ways to do this, even after requiring that the renormalized perturbation expansions satisfy the governing equations at each order. We can again choose for simplicity, bearing in mind that $\mathcal{O}(\epsilon^1)$ renormalization can impact $\mathcal{O}(\epsilon^2)$ calculations. It is easiest to apply the RG condition to the renormalized perturbation expansion, rather than applying it to the integration constants directly. In solving the resulting equation, we want solutions which are valid $\forall x$. To solve the RG equation, care must be taken to satisfy several conditions simultaneously, and it cannot be assumed that our renormalized constants have a derivative of $\mathcal{O}(\epsilon^1)$.

Although there is quite a bit of flexibility in implementing the RG technique, our results are robust: Regardless of how we structure the calculation, our solutions agree to within an accuracy limited by the original naïve perturbative solution; they are asymptotically equivalent. It is this robustness which makes RG a useful tool for the low Reynolds number problems, where the complexity of the Navier-Stokes equations will constrain our choices.

4.2 Flow Past A Cylinder

4.2.1 Rescaling

To solve Eqn. 2.8 using RG techniques, we begin by transforming the problem to the Oseen variables. As in the terrible problem, to find a solution which is valid for all $\vec{r}$, we need to analyze Eqn. 2.8 using a *dominant balance* argument. As it stands,

different terms of Eqn. 2.8 will dominate in different regimes.[13] Looking for a rescaling of ψ and r which makes all terms of the same magnitude (more precisely, of the same order in R), yields the rescaling given in Eqn. 4.40[26].

$$\rho = Rr, \qquad \Psi = R\psi \tag{4.40}$$

Transforming to these variables, Eqn. 2.8 becomes:

$$\nabla_\rho^4 \Psi(\rho,\theta) = -\frac{1}{\rho}\frac{\partial(\Psi,\nabla_\rho^2)}{\partial(\rho,\theta)} \tag{4.41}$$

The boundary conditions (Eqn. 2.9) become:

$$\Psi(\rho = R,\theta) = 0, \qquad \left.\frac{\partial\Psi(\rho,\theta)}{\partial\rho}\right|_{\rho=R} = 0, \qquad \lim_{\rho\to\infty}\frac{\Psi(\rho,\theta)}{\rho} = sin(\theta) \tag{4.42}$$

4.2.2 Naïve Perturbation Analysis

The next step in obtaining the RG solution is to begin with the ansatz that the solution can be obtained from a perturbation expansion (Eqn. 4.43).

$$\Psi(\rho,\theta) = \Psi_0(\rho,\theta) + R\Psi_1(\rho,\theta) + R^2\Psi_2(\rho,\theta) + \mathcal{O}\left(R^2\right) \tag{4.43}$$

Substituting Eqn. 4.43 into Eqn. 4.41, and collecting powers of R yields a series of equations which must be satisfied:

$$\mathcal{O}\left(R^0\right) : \nabla_\rho^4\Psi_0(\rho,\theta) = \frac{1}{\rho}\left(\frac{\partial\Psi_0}{\partial\theta}\frac{\partial}{\partial\rho} - \frac{\partial\Psi_0}{\partial\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_0 \tag{4.44}$$

$$\mathcal{O}\left(R^1\right) : \nabla_\rho^4\Psi_1(\rho,\theta) = \frac{1}{\rho}\left(\left(\frac{\partial\Psi_1}{\partial\theta}\frac{\partial}{\partial\rho} - \frac{\partial\Psi_1}{\partial\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_0 + \left(\frac{\partial\Psi_0}{\partial\theta}\frac{\partial}{\partial\rho} - \frac{\partial\Psi_0}{\partial\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_1\right)$$

$$\mathcal{O}\left(R^2\right) : \nabla_\rho^4\Psi_2(\rho,\theta) = \frac{1}{\rho}\left(\left(\frac{\partial\Psi_2}{\partial\theta}\frac{\partial}{\partial\rho} - \frac{\partial\Psi_2}{\partial\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_0 + \left(\frac{\partial\Psi_0}{\partial\theta}\frac{\partial}{\partial\rho} - \frac{\partial\Psi_0}{\partial\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_2 + \left(\frac{\partial\Psi_1}{\partial\theta}\frac{\partial}{\partial\rho} - \frac{\partial\Psi_1}{\partial\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_1\right)$$

[13] i.e. the LHS, which is comprised of *inertial terms* dominates for small $|\vec{r}|$ whereas at large $|\vec{r}|$ the *viscous* terms which comprise the RHS are of equal or greater importance.

4.2.3 $\mathcal{O}\left(R^0\right)$ Solution

The zeroth order part of Eqn. 4.45 is the same as Eqn. 4.41, and is equally hard to solve. But RG does not need a complete solution; we just need a starting point. We will begin with the the equation which describes a uniform stream. This is analogous to the constant $\mathcal{O}\left(\epsilon^0\right)$ solution in the "terrible" problem.

A first integral to the $\mathcal{O}\left(R^0\right)$ equation can be obtained by noting that any solutions of $\nabla_\rho^2 \Psi_0(\rho,\theta) = 0$ are also solutions of Eqn. 4.45. This is Laplace's equation in cylindrical coordinates, and has the usual solution (assuming the potential is single-valued):

$$\Psi_0(\rho,\theta) = A_0 + B_0 \ln\rho + \sum_{n=1}^{\infty} \left(\left(A_n\rho^n + B_n\rho^{-n}\right)\sin n\theta + \left(C_n\rho^n + D_n\rho^{-n}\right)\cos n\theta\right) \tag{4.45}$$

We are only interested in solutions with the symmetry imposed by the uniform flow (Eqn. 2.9). Hence $A_0 = B_0 = C_n = D_n = 0$. Furthermore, the boundary conditions at infinity require that $A_n = 0$ for $n > 1$. For simplicity at higher orders, we set $C_n = 0$; this is not required, but these terms will simply re-appear at $\mathcal{O}\left(R^1\right)$. Finally set $A_1 = 1$ to satisfy the boundary condition at ∞ (Eqn. 4.42). As in the "terrible" problem, this is done for technical convenience, but will not change our results. We are left with the potential describing the uniform flow:

$$\Psi_0(\rho,\theta) = \rho\sin(\theta) \tag{4.46}$$

4.2.4 $\mathcal{O}\left(R^1\right)$ Solution

By substituting Eqn. 4.46 into the $\mathcal{O}\left(R^1\right)$ governing equation, we obtain Eqn. 4.47.

$$\nabla_\rho^4 \Psi_1(\rho,\theta) = \left(\cos(\theta)\frac{\partial}{\partial\rho} - \frac{\sin(\theta)}{\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_1 \tag{4.47}$$

This equation is formally identical to Oseen's equation (Eqn. 3.11), albeit derived through a different argument. This is fortuitous, as its solutions are known[31]. Un-

fortunately, when working with stream functions, the solution can only be expressed as an infinite sum involving combinations of modified Bessel functions, K_n, I_n.

The general solution can be obtained either by following Tomotika or by using variation of parameters[26]. It is comprised of two parts, the first being a solution of Laplace's equation (as at $\mathcal{O}(R^0)$). The same considerations of symmetry and boundary conditions limit our solution: In Eqn. 4.45, $A_0 = B_0 = C_n = D_n = 0$; $A_n = 0$, if $n > 1$. Here, however, we retain the constants B_n, and do not fix A_1. This is analogous to what was done with the homogeneous terms at $\mathcal{O}(\epsilon^1)$ in the "terrible" problem. The second part of the general solution is analogous to a particular solution in the "terrible" problem, and can be obtained from Tomotika's solution (Eqn. 3.23). These two results are combined in Eqn. 4.48, which will be the basis for our RG analysis.

$$\Psi_1(\rho,\theta) = A_1\rho\sin\theta + \sum_{n=1}^{\infty}\left(B_n\rho^{-n} + \sum_{m=0}^{\infty} X_m\rho\Phi_{m,n}(\rho/2)\right)\sin n\theta \tag{4.48}$$

Before discussing the application of RG to Eqn. 4.48, it is worthwhile to discuss Eqn. 4.47 in general terms. Eqn. 4.47 may be re-written as:

$$\mathcal{L}\Psi_1 \equiv \left(\nabla_\rho^2 - \cos(\theta)\frac{\partial}{\partial\rho} + \frac{\sin(\theta)}{\rho}\frac{\partial}{\partial\theta}\right)\nabla_\rho^2\Psi_1 = 0 \tag{4.49}$$

We see explicitly that this equation is a linear operator ($\mathcal{L}$) acting on Ψ_1, and that the RHS is zero. This is the *homogeneous Oseen equation*. It is only because of our judicious choice of Ψ_0 that we do not need to deal with the inhomogeneous counterpart, i.e. with a non-zero RHS. However, the inhomogeneous Oseen equation governs Ψ_n at all higher orders. This can be seen for $\mathcal{O}(R^2)$ from Eqn. 4.45.

In general, the solutions to the inhomogeneous Oseen equation are found using the method of variation of parameters. It is worth exploring these solutions, as they provide some insight into the structure of Eqn. 3.23. We now solve Eqn. 4.49 for a particular kind of inhomogeneity, one which can be written as a Fourier sine series.[14]

[14]The symmetry of the problem precludes the possibility of cosine terms in the governing equations for Ψ_n, $\forall n > 1$

We want to solve:

$$\mathcal{L}\Psi_1 = \sum_{n=1}^{\infty} \tilde{F}_n(\rho) \sin n\theta \tag{4.50}$$

The substitution $\nabla^2\Psi_1 = e^{\rho\cos\theta/2}\Pi(\rho,\theta)$[15], allows us to obtain the first integral of Eqn. 4.50. This result is given by Eqn. 4.51[26].

$$\left(\nabla^2 - \frac{1}{4}\right)\Pi(\rho,\theta) = \sum_{n=1}^{\infty} F_n(\rho)\sin n\theta \tag{4.51}$$

Here $F_n(\rho) = e^{-\rho\cos\theta/2}\tilde{F}_n(\rho)$. To solve for $\Pi(\rho,\theta)$, begin by noting that the symmetry of the inhomogeneous terms implies that $\Pi(\rho,\theta)$ can be written as a sine series. Consequently, substitute $\Pi(\rho,\theta) = \sum_{n=1}^{\infty} g_n(\rho)\sin n\theta$ into Eqn. 4.51 to obtain:

$$g_n^{''}(\rho) + \frac{1}{\rho}g_n^{'}(\rho) - \left(\frac{1}{4} + \frac{1}{\rho^2}\right) g_n(\rho) = F_n(\rho) \tag{4.52}$$

The fundamental solutions of this equation are $K_n(\rho/2)$, $I_n(\rho/2)$. Using variation of parameters, the general solution of Eqn. 4.52 may be written:

$$g_n(\rho) = -I_n(\frac{\rho}{2})\left(\alpha_n + \mathcal{J}_1^{(n)}(\rho)\right) + K_n(\frac{\rho}{2})\left(\beta_n + \mathcal{J}_2^{(n)}(\rho)\right) \tag{4.53}$$

Here, $\mathcal{J}_1^{(n)}(\rho) = \int \mathrm{d}\rho\rho F_n(\rho)K_n(\rho/2)$, $\mathcal{J}_2^{(n)}(\rho) = \int \mathrm{d}\rho\rho F_n(\rho)I_n(\rho/2)$, and α_n, β_n are constants. The next step is to undo our original transformation, and to solve the resulting equation:

$$\begin{aligned}\nabla^2\Psi_1(\rho,\theta) &= e^{\frac{\rho\cos\theta}{2}} \sum_{n=1}^{\infty} g_n(\rho)\sin n\theta \\ &= \sum_{n=1}^{\infty} b_n(\rho)\sin n\theta\end{aligned} \tag{4.54}$$

In this equation, $b_n(\rho) = \sum_{m=1}^{\infty} g_m(\rho)\left(I_{n-m}\left(\rho/2\right) - I_{n+m}\left(\rho/2\right)\right)$. We have the unfortunate happenstance that each b_n depends on the *all* of the harmonics of the first integral. This is the origin of the nested sum (over m) in Tomotika's solution (Eqn. 3.23).

[15] $\nabla^2\Psi_1(\rho,\theta)$ is the *vorticity*.

As before, symmetry will require that $\Psi_1(\rho,\theta)$ be representable as a sine series: $\Psi_1(\rho,\theta) = \sum_{m=1}^{\infty} X_m(\rho)\sin m\theta$. With this substitution we obtain (for each m), the radial component of Poisson's equation in cylindrical coordinates:

$$X_m''(\rho) + \frac{1}{\rho}X_m'(\rho) - \frac{m^2}{r^2}X_m(\rho) = b_m(\rho) \tag{4.55}$$

The fundamental solutions were discussed before in the context of Laplace's equation: ρ^m, ρ^{-m}. As before, a particular integral is obtained through variation of parameters, and the general solution may be written:

$$X_n(\rho) = -\rho^n\left(A_n + \mathcal{I}_1^{(n)}(\rho)\right) + \frac{1}{\rho^n}\left(B_n + \mathcal{I}_2^{(n)}(\rho)\right) \tag{4.56}$$

Here $\mathcal{I}_1^{(n)}(\rho) = \int \mathrm{d}\rho - \rho b_n(\rho)/(2n\rho^n)$, $\mathcal{I}_2^{(n)}(\rho) = \int \mathrm{d}\rho - \rho b_n(\rho)\rho^n/(2n)$, and A_n, B_n are integration constants.

It is useful to relate Eqn. 4.56 to Tomotika's solution (Eqn. 3.23). There are four integration constants for each angular harmonic. Two are obvious: A_n,B_n. The other two arise in the first integral (the vorticity solution), Eqn. 4.53. However, every vorticity integration constant appears in each harmonic of Eqn. 4.56. For example, you cannot uniquely assign α_1 and β_1 to the $\sin\theta$ harmonic of Eqn. 4.56. However, if one considers n terms from Eqn. 4.53 and n terms from Eqn. 4.56, there will be $4n$ integration constants — four per retained harmonic of Eqn. 4.56. In passing we note that matched asymptotics workers avoid this problem by using the vorticity directly, and thereby simplify their treatment of boundary conditions. This approach does not work in conjunction with RG.

It is mildly disconcerting to have four integration constants, as there are only three boundary conditions for each harmonic (Eqn. 4.42). However, two of the constants — A_n and α_n — will be determined by the boundary conditions at infinity. This claim is not obvious, particularly since terms which are divergent prior to renormalization might not be present after the renormalization procedure. We outline here an argument

which can be made rigorous. There are two kinds of divergences in Eqn. 4.56: Terms which are secular as $\rho \to 0$, and terms which diverge too quickly as $\rho \to \infty$.[16]

After renormalization, we will try to need to meet the boundary conditions (Eqn. 4.42). As in the case of the "terrible" problem, it will turn out that the simplest approach to renormalization yields a renormalized perturbation solution which is the same as the naïve series. Consider Eqn. 4.56. The terms which are secular as $\rho \to 0$ will not preclude satisfying the boundary conditions. Those which diverge too quickly as $\rho \to \infty$, however, will conflict with Eqn. 4.42.

These terms must be eliminated by a suitable choice of integration constants. It turns out not to matter whether we do this before or after the renormalization procedure. For simplicity, we will do it before renormalizing. First, the coefficient of ρ^n must vanish for all $n > 1$. This can happen, with an appropriate choice of A_n, if

$$\lim_{\rho\to\infty} \mathcal{I}_1^{(n)}(\rho) \sim \mathcal{O}(1)$$

For this requirement to be met, the coefficient of $I_n(\rho/2)$ in Eqn. 4.53 must vanish (e.g., $\alpha_n = \lim_{\rho\to\infty} \mathcal{J}_1^n(\rho)$). It is always possible to choose α_n appropriately, because the following condition is satisfied for all n:

$$\lim_{\rho\to\infty} \mathcal{J}_1^{(n)}(\rho) \sim \mathcal{O}(1)$$

In our problem this is true because $F_n(\rho)$ is based on solutions to the lower order governing equations. By construction, these are well-behaved as $\rho \to \infty$. Therefore, for the inhomogeneous Oseen equation under consideration (Eqn. 4.51), we see that two of the four integration constants — A_n, α_n — are needed to satisfy the boundary conditions at infinity.

More specifically, the immediate problem requires us to consider the homogeneous Oseen's equation (Eqn. 4.49), and Tomotika's solution (Eqn. 3.23). For this problem,

[16] To be precise, terms which diverge faster than ρ as $\rho \to \infty$ are problematic, and prevent satisfying the boundary conditions (Eqn. 4.42)

$F_n(\rho) = 0$, and the coefficient of $I_n(\rho/2)$ in Eqn. 4.53 has no ρ dependence. So we simply choose α_n such that this coefficient vanishes. Simplifying Eqn. 4.53, we then have the following solution for the vorticity:

$$\nabla^2\Psi_1(\rho,\theta) = e^{\frac{\rho\cos\theta}{2}} \sum_{n=1}^{\infty} K_n(\frac{\rho}{2})\,(\beta_n)\sin n\theta \tag{4.57}$$

Since this solution for the vorticity is well-behaved as $\rho \to \infty$, it follows that we can choose A_n $(n > 1)$ in Eqn. 4.56 so that the coefficient of ρ^n vanishes as $\rho \to \infty$. We are left with the solution

$$X_n(\rho) = A_n\rho\delta_{n,1} + \rho^n\left(\mathcal{I}_1^{(n)}(\rho) - \mathcal{I}_1^{(n)}(\infty)\right) + \rho^{-n}\left(B_n + \mathcal{I}_2^{(n)}(\rho)\right) \tag{4.58}$$

For the homogeneous Oseen's equation, $\mathcal{I}_1^{(n)}(\rho)$ and $\mathcal{I}_2^{(n)}(\rho)$ simplify to:

$$\mathcal{I}_1^{(n)}(\rho) = \int \mathrm{d}\rho \frac{-\rho}{2n}\rho^{-n}\left(\sum_{m=1}^{\infty}\beta_m K_m(\frac{\rho}{2})\left(I_{n-m}(\frac{\rho}{2}) - I_{n+m}(\frac{\rho}{2})\right)\right) \tag{4.59}$$

$$\mathcal{I}_2^{(n)}(\rho) = \int \mathrm{d}\rho \frac{-\rho}{2n}\rho^{n}\left(\sum_{m=1}^{\infty}\beta_m K_m(\frac{\rho}{2})\left(I_{n-m}(\frac{\rho}{2}) - I_{n+m}(\frac{\rho}{2})\right)\right) \tag{4.60}$$

This result is fundamentally the same as Tomotika's (Eqn. 3.23). However, his solution is more useful, as he accomplished the integrals in Eqn. 4.59. What is the point of all this work? Firstly, the approach based on the variation of parameters may be applied to the inhomogeneous Oseen's equation, which must be solved for orders higher than $\mathcal{O}(R^1)$. Secondly, we see explicitly what happens to the two sets of integration constants α_n and A_n. Tomotika's solution has but two integration constants[17] — B_n and β_n. The other constants have already been chosen so as to satisfy the boundary conditions at ∞. We have shown explicitly how they must be determined, and stated without proof that this may be done prior to renormalization. In short, we have explained why Eqn. 4.48 is the appropriately general $\mathcal{O}(R^1)$ solution for our naïve perturbation analysis.

In addition to explaining why Tomotika's solution is a suitable starting point for RG, our analysis also connects with the $\mathcal{O}(R^1)$ solution of Proudman and Pearson[26].

[17] There is also A_1, but this is a obviously a special case.

We have shown that the vorticity must be well-behaved at $\rho = \infty$ if the overall solution is to satisfy the boundary conditions.

Secular Behavior

Combining Eqns. 4.46, 4.48, we begin the following naïve solution:

$$\Psi_(\rho,\theta) = \rho\sin(\theta)+R\left(A_1\rho\sin\theta+\sum_{n=1}^{\infty}\left(B_n\rho^{-n}+\sum_{m=0}^{\infty}X_m\rho\Phi_{m,n}(\rho/2)\right)\sin n\theta\right)+\mathcal{O}\left(R^2\right) \tag{4.61}$$

Although intimidating, this is conceptually equivalent to Eqn. 4.10 (in the terrible problem). The first step in our analysis is identifying which terms are divergent. As explained above, Eqn. 4.61 is specifically constructed to be of $\mathcal{O}\left(\rho^1\right)$ as $\rho \to \infty$. In fact, only the $\mathcal{O}\left(R^0\right)$ and A_1 terms matter at large ρ. As $\rho \to 0$, however, many other terms in Eqn. 4.61 diverge. All of the B_n terms diverge. Most of the $\Phi_{m,n}(\rho)$ terms are also secular.

Rather than enumerating and sorting through the different divergences, we simply treat the problem abstractly. Eqn. 4.61 can be rewritten as:

$$\Psi_(\rho,\theta) = \rho\sin(\theta) + R\left(A_1\rho\sin\theta + \mathcal{R}(\rho,\theta;\{B_i\};\{X_j\}) + \mathcal{S}(\rho,\theta;\{B_m\};\{X_n\})\right) \tag{4.62}$$

Here, $\mathcal{S}$ includes the terms which are secular as $\rho \to 0$, and $\mathcal{R}$ includes regular terms.

4.2.5 Renormalization

Equation 4.62 is renormalized just like the terrible problem. We begin with the renormalized perturbation expansion given in Eqn. 4.63. Note that we are not specifying the details of which terms are secular, or how we are "splitting" these terms. The only term we are explicitly considering is A_1. This is a trick built on consideration of the terrible problem. Our "best" solution (Eqn. 4.39) to that problem was built on the renormalization of just *one* constant, B_0 in Eqn. 4.38a. Essentially, we will repeat that

procedure here, using A_1 as that constant.

$$\Psi(\rho,\theta) = \rho\sin(\theta) + R\Big(A_1(\tau)\rho\sin\theta + \mathcal{R}(\rho,\theta;\{B_i(\tau)\};\{X_j(\tau)\}) + \mathcal{S}(\rho,\theta;\{B_m(\tau)\};\{X_n(\tau)\}) - \mathcal{S}(\tau,\theta;\{B_m(\tau)\};\{X_n(\tau)\}) + \mathcal{O}\left(R^2\right)\Big) \tag{4.63}$$

We will now apply the RG condition — $\partial_\tau\Psi(\rho,\theta) = 0$ — to this equation. Accomplishing this in complete generality is difficult. However, using our experience from the terrible problem, we can see that this is not necessary. The RG condition may be satisfied as follows: First, suppose that $X_n^{'}(\tau) = \mathcal{O}\left(R^2\right) \quad \forall n$, $B_m^{'}(\tau) = \mathcal{O}\left(R^2\right) \quad \forall m$. These equations are satisfied by $X_n(\tau) = \chi_n$, $B_m(\tau) = \beta_m$. Substituting these results into Eqn. 4.63, and applying the RG condition results in:

$$0 = R\left(A_1^{'}(\tau)\rho\sin\theta - \mathcal{S}^{'}(\tau,\theta;\{\beta_m\};\{\chi_n\})\right) \tag{4.64}$$

This is easily solved for $A_1(\tau)$.

$$A_1(\tau) = \frac{\mathcal{S}(\tau,\theta;\{\beta_m\};\{\chi_n\})}{\rho\sin\theta} + \alpha_1 \tag{4.65}$$

We have explicitly validated our supposition that $\{X_n(\tau)\}$ and $\{B_m(\tau)\}$ can be constants. With this supposition, we have shown that the RG condition applied to Eqn. 4.63 can be satisfied with an appropriate choice of $A_1(\tau)$. We have satisfied the RG condition through clever tricks derived from our experience with the terrible problem. However, this solution is entirely valid, and our experience with the terrible problem has shown us that more complicated solutions are asymptotically equivalent.

Substituting Eqn. 4.65 into Eqn. 4.63, and setting $\tau = \rho$, we obtain our renormalized solution:

$$\Psi_(\rho,\theta) = \rho\sin(\theta) + R\left(\alpha_1\rho\sin\theta + \mathcal{R}(\rho,\theta;\{\beta_i\};\{\chi_j\}) + \mathcal{S}(\rho,\theta;\{\beta_m\};\{\chi_n\})\right) \tag{4.66}$$

By now it should not be surprising that this is the same equation as our naive perturbation solution (Eqn. 4.62), and by extension the same solution obtained by Tomotika[31]. As in the case of the terrible problem, however, we now know that this is a uniformly

valid approximation. We now may choose the integration constants to satisfy the boundary conditions, and then calculate the drag coefficient.

Truncation

Unfortunately, there are infinitely many integration constants, and it is impossible to apply the boundary conditions to our renormalized solution (or Eqn. 4.61). To progress further, we must make the same sort of uncontrolled approximations made by previous workers[26, 31].[18]

Our approximation consists of a careful truncation, in both m and n, of the series in Eqn. 4.61. There are two important points to consider. First is the $\sin\theta$ symmetry of the overall problem: terms proportional to $\sin\theta$ reflect the symmetries exhibited by the uniform flow which are imposed on our solution via the boundary conditions at infinity. The importance of the this harmonic is further seen in Eqn. 2.20: Only the coefficient of $\sin\theta$ will be needed for the computation of C_D.

Secondly we recall that the remaining boundary conditions are imposed at the surface of the sphere, at $\rho = R$ in Oseen coordinates. When applying the boundary conditions, terms which are secular as $\rho \to 0$ will therefore be most important. Specifically, we cannot truncate any terms which are divergent, although we are at liberty to set their coefficients equal to zero.

These considerations allow exactly one solution. First, set all $B_n = 0 \quad n > 1$. Secondly, set all $X_m = 0 \quad m > 0$. We retain three coefficients: A_1, B_1, X_0, which will permit the boundary conditions to be satisfied for the $\sin\theta$ harmonic. What about the higher harmonics? These terms are truncated in an *uncontrolled* approximation. However, as we will show, the discarded terms are $\mathcal{O}\left(R^3 \log R\right)$ or higher at the surface of the sphere. They are regular terms, and thus negligible in comparason to the secular

[18]Kaplun was able to avoid this difficulty by using the velocity field instead of stream functions, although his approach brings other problems: the solution cannot be expressed in closed form, and must be approximated to apply the boundary conditions (see section 3.2.3).

n =	1	2	3	4
$\Psi^{(n)}_{\text{discard}}(\rho,\theta)$	$\mathcal{O}\left(R^3 \log R\right)$	$\mathcal{O}\left(R^2\right)$	$\mathcal{O}\left(R^1\right)$	$\mathcal{O}\left(R^0\right)$
$\Psi'^{(n)}_{\text{discard}}(\rho,\theta)$	$\mathcal{O}\left(R^2 \log R\right)$	$\mathcal{O}\left(R^1\right)$	$\mathcal{O}\left(R^1\right)$	$\mathcal{O}\left(R^{-1}\right)$

Table 4.1: Relative importance of discarded terms at $\rho = R$

terms retained (which are $\mathcal{O}\left(R^{-1}\right)$).

Now, suppose we follow Tomotika, and try to extand this approach, by retaining a few more terms. The next step would be to retain the B_2, X_1 terms, and to try to satisfy the boundary conditions for the $\sin 2\theta$ harmonic. As before, all the higher B_n, X_m are set to zero. Why not include the next harmonic or two?

The answer lies in the terms we discard. If we satisfy the boundary conditions at $\rho = R$ for the first n harmonics, we must retain the coefficients $X_0, ..., X_n - 1$. To minimize the amount of truncation we do, first set $X_m = 0$ for $\forall m > n-1$ and $B_k = 0$ for $\forall k > n$. What, then, is the form of the terms which are discarded from our solution?

$$\Psi^{(n)}_{\text{discard}}(\rho,\theta) = R\left(\sum_{k=n+1}^{\infty}\sum_{m=0}^{n-1} X_m \Phi_{m,k}(\rho/2)\rho \sin k\theta\right) \tag{4.67}$$

$\Psi^{(n)}_{\text{discard}}(\rho,\theta)$ is largest as $\rho \to 0$, and will be most important at $\rho = R$, on the surface of the cylinder. If we retain only the $n = 1$ harmonic, $\Psi^{(1)}_{\text{discard}}(\rho,\theta) \sim \mathcal{O}\left(R^3 \log R\right)$. Since we are only working to $\mathcal{O}\left(R^1\right)$, this is fine. We must also consider the derivative, since we want to satisfy all of the boundary conditions (Eqn. 2.9) to the same order. $\Psi'^{(1)}_{\text{discard}}(\rho,\theta) \sim \mathcal{O}\left(R^2 \log R\right)$ Therefore, in the case where we retain only the $\sin\theta$ harmonic, the discarded terms are negligible, as we are working to $\mathcal{O}\left(R^1\right)$.[19] When we retain higher harmonics, everything changes. Table 4.1 shows the magnitude of the discarded terms at $\rho = R$ for the first four harmonics.

From Table 4.1, we see immediately that to retain $\sin 2\theta$ harmonics, we must have

[19]This argument is somewhat simplistic: The neglected terms also contribute, when meeting the boundary conditions, to the values of the retained coefficients. i.e. All non-zero X_m affect X_0. But these are lower order effects.

an error in our derivative boundary condition of $\mathcal{O}(R^1)$ — the order to which we are trying to work. If we retain higher harmonics, this situation gets worse. First we have an $\mathcal{O}(R^1)$ error in the stream function itself, and then we begin to have errors which are *divergent* in R! For $n > 4$, both $\Psi^{(n)}_{\text{discard}}(\rho,\theta)$ and $\Psi^{'(n)}_{\text{discard}}(\rho,\theta)$ are increasingly divergent functions of R.

Since it is in practice impossible to fit the boundary conditions to Eqn. 4.61, we must truncate the series expansion. We have shown that there is only one truncation consistent with both the symmetry requirements of the problem and the demand that we satisfy the boundary conditions to $\mathcal{O}(R^1)$:

$$\Psi_(\rho,\theta) = \rho\sin(\theta) + R\left(A_1\rho + B_1\rho^{-1} + X_0\rho\Phi_{0,1}(\rho/2)\right)\sin\theta \tag{4.68}$$

This result is identical to Proudman's $\mathcal{O}(R^1)$ result for the Oseen stream function[26]. However, he arrives at this result by considering matching requirements with the $\mathcal{O}(R^0)$ Stokes expansion and by imposing $\sin\theta$ symmetry on the first integral (Eqn. 4.55). Our approach arrives at the same conclusion, but without the need for asymptotic matching or the two expansions it requires. Moreover, we did not need the expertise and finesse which matched asymptotics workers needed to deduce the unusual form of their expansions (e.g., the $1/\log R$ term in Eqn. 3.32). Finally, we note that Tomotika's numerical results support our truncation[31].

Meeting Boundary Conditions

It is straightforward to apply the boundary conditions (Eqn. 2.9) to Eqn. 4.68. To satisfy the condition at infinity, $A_1 = 0$. The other two requirements are met by the following choice of coefficients:

$$B_1 = \frac{-R^2\Phi'_{0,1}(R/2)}{4\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)} \tag{4.69}$$

$$X_0 = \frac{-4}{R\left(4\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)\right)} \tag{4.70}$$

Notice that we are using the Oseen stream function. The Stokes' function is related by: $\psi(r,\theta) = \Psi(rR,\theta)/R$. Putting everything together, we have the new result given by Eqn. 4.71. Remember also although our truncated solution satisfies the boundary conditions *exactly*, it only satisfies the governing equations *approximately*.

$$\Psi_(\rho,\theta) = \rho\sin(\theta) + R\left(\frac{-R^2\Phi'_{0,1}(R/2)}{4\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)}\rho^{-1} + \frac{-4}{R\left(4\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)\right)}\rho\Phi_{0,1}(\rho/2)\right)\sin\theta \tag{4.71}$$

4.2.6 Calculating the Drag Coefficient

We now transform Eqn. 4.71 into Stokes' coordinates, and substitute the result into Eqn. 2.20.[20] We thereby obtain a new result for C_D, given by Eqn. 4.72.

$$C_D = \frac{\pi\left(-12\Phi'_{0,1}(R/2) + R\left(6\Phi''_{0,1}(R/2) + R\Phi'''_{0,1}(R/2)\right)\right)}{8\Phi_{0,1}(R/2) + 2R\Phi'_{0,1}(R/2)} \tag{4.72}$$

This result is plotted in Figure 4.1, where it is compared against the principal results of Oseen theory, matched asymptotic theory, and experiments. When compared asymptotically, all of these theoretical predictions agree. At small but not infinitesimal Reynolds number, the largest difference is seen between Kaplun's second order result and the first order predictions, including Eqn. 4.72. As explained previously, current experimental data cannot determine whether Kaplun's second order matched asymptotics solution is actually superior.

The RG result lies among the first order predictions. Fundamentally, the RG calculation begins with an equation similar to Oseen's, so this is not too surprising. Within this group Eqn. 4.72 performs very well, and is only slightly bettered by Imai's prediction (Eqn. 3.25). These two results are very close over the range $0 < R < 1$.

The real strength of Eqn. 4.72 can be seen in in Figure 4.2. As the Reynolds number increases beyond $R = 1$, all other theories begin to behave pathologically. They

[20] Or, alternatively, into Eqns. 2.7, 2.18, and 2.19

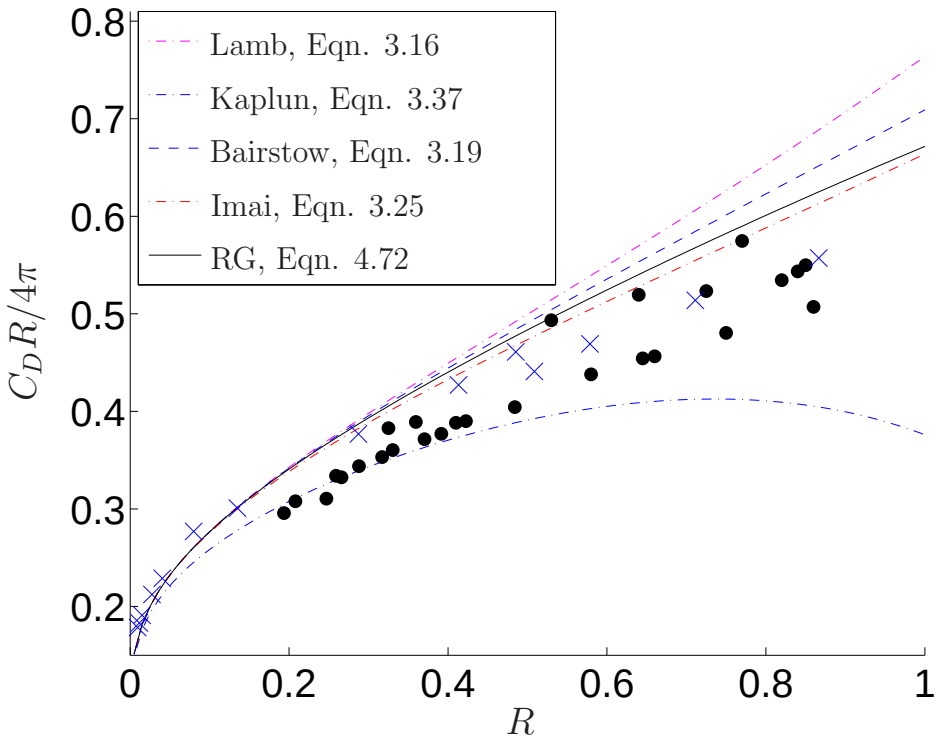

Figure 4.1: Drag on cylinder, comparing RG predictions to other theories at low R.

diverge from experimental measurements and behave non-physically (e.g., a negative drag coefficient). The RG prediction suffers from none of these problems; it is well behaved for all R. As it is still based on a perturbative solution, it does becomes less accurate as R increases.

4.3 Flow Past a Sphere

4.3.1 Rescaling

Our analysis of low Reynolds number flow past a sphere closely follows both the cylinder problem and the terrible problem. We omit redundant explanations. As before, the first step is a rescaling of both r and ψ — the transformation into Oseen coordinates.

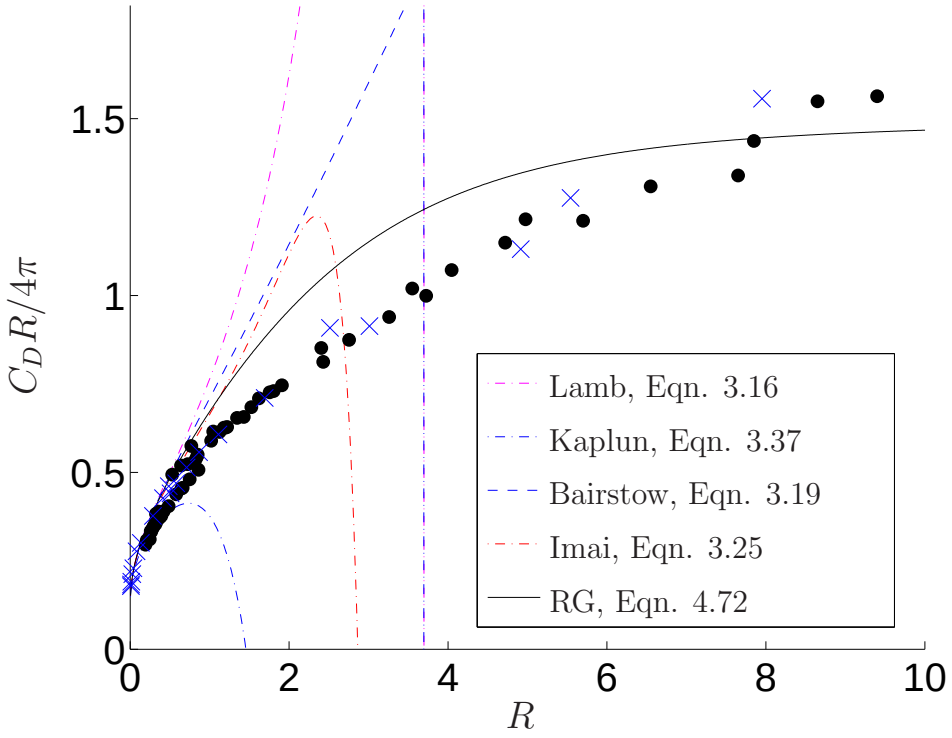

Figure 4.2: Drag on a Cylinder, comparing RG predictions to other theories at higher R.

A dominant balance analysis identifies the rescaling given in Eqn. 4.73.

$$\rho = Rr, \quad \Psi = R^2\psi \tag{4.73}$$

In Oseen variables, the governing equation (Eqn. 2.11) becomes:

$$D_\rho^4\Psi(\rho,\mu) = \frac{1}{\rho^2}\left(\frac{\partial(\Psi(\rho,\mu), D_\rho^2\Psi(\rho,\mu))}{\partial(\rho,\mu)} + 2D_\rho^2\Psi(\rho,\mu)L_\rho\Psi(\rho,\mu)\right) \tag{4.74}$$

Where

$$\mu \equiv \cos\theta, \qquad D_\rho^2 \equiv \frac{\partial^2}{\partial\rho^2} + \frac{1-\mu^2}{\rho^2}\frac{\partial^2}{\partial\mu^2}, \qquad L_\rho \equiv \frac{\mu}{1-\mu^2}\frac{\partial}{\partial\rho} + \frac{1}{\rho}\frac{\partial}{\partial\mu} \tag{4.75}$$

The boundary conditions (Eqn. 2.13) transform into:

$$\Psi(\rho = R,\mu) = 0, \qquad \left.\frac{\partial\Psi(\rho,\mu)}{\partial\rho}\right|_{\rho=R} = 0, \qquad \lim_{\rho\to\infty}\frac{\Psi(\rho,\mu)}{\rho^2} = \frac{1}{2}\left(1-\mu^2\right) \tag{4.76}$$

4.3.2 Naïve Perturbation Analysis

We continue by substituting our naïve perturbation assumption (Eqn. 4.43) into Eqn. 4.74, and then collecting powers of R.

$$\mathcal{O}\left(R^0\right) : D_\rho^4\Psi_0(\rho,\mu) = \frac{1}{\rho^2}\left(\frac{\partial(\Psi_0(\rho,\mu), D_\rho^2\Psi_0(\rho,\mu))}{\partial(\rho,\mu)} + 2D_\rho^2\Psi_0(\rho,\mu)L_\rho\Psi_0(\rho,\mu)\right) \quad (4.77a)$$

$$\mathcal{O}\left(R^1\right) : D_\rho^4\Psi_1(\rho,\mu) = \frac{1}{\rho^2}\left(\frac{\partial(\Psi_0, D_\rho^2\Psi_1)}{\partial(\rho,\mu)} + \frac{\partial(\Psi_1, D_\rho^2\Psi_0)}{\partial(\rho,\mu)} + 2\left(D_\rho^2\Psi_0 L_\rho\Psi_1 + D_\rho^2\Psi_1 L_\rho\Psi_0\right)\right) \quad (4.77b)$$

$$\mathcal{O}\left(R^2\right) : D_\rho^4\Psi_2(\rho,\mu) = \frac{1}{\rho^2}\left(\frac{\partial(\Psi_0, D_\rho^2\Psi_2)}{\partial(\rho,\mu)} + \frac{\partial(\Psi_1, D_\rho^2\Psi_1)}{\partial(\rho,\mu)} + \frac{\partial(\Psi_2, D_\rho^2\Psi_0)}{\partial(\rho,\mu)} + 2\left(D_\rho^2\Psi_0 L_\rho\Psi_2 + D_\rho^2\Psi_1 L_\rho\Psi_1 + D_\rho^2\Psi_2 L_\rho\Psi_0\right)\right) \quad (4.77c)$$

4.3.3 $\mathcal{O}\left(R^0\right)$ Solution

As seen with both the cylinder problem and the terrible problem, Eqn. 4.77a is the same as the original governing equation (Eqn. 4.74). As before, we proceed using an incomplete solution for Ψ_0: the uniform stream which describes flow far from any disturbances. Analogously to the cylinder, we notice that Eqn. 4.77a is satisfied if $\Psi_0(\rho,\mu)$ obeys $D_\rho^2\Psi_0(\rho,\mu) = 0$. The general solution of this equation which also satisfies the appropriate symmetry requirement ($\Psi_0(\rho,\mu=\pm1) = 0$) is given by Eqn. 4.78.

$$\Psi_0(\rho,\mu) = \sum_{n=0}^{\infty}\left(A_n\rho^{n+1} + B_n\rho^{-n}\right)Q_n(\mu) \quad (4.78)$$

Here $Q_n(\mu)$ is defined as in Eqn. 3.20. Following the analysis used for the cylinder, we set all of the coefficients to zero, excepting $A_1 = -1/2$. This choice of A_1 satisfies the uniform stream boundary condition (Eqn. 4.76) at $\rho = \infty$. We thereby obtain:

$$\Psi_0(\rho,\mu) = -\rho^2 Q_1(\mu) \quad (4.79)$$

4.3.4 $\mathcal{O}\left(R^1\right)$ Solution

Substituting Eqn. 4.79 into Eqn. 4.77b, we obtain Eqn. 4.80:

$$D_\rho^4 \Psi_1(\rho,\mu) = \left(\frac{1-\mu^2}{\rho}\frac{\partial}{\partial\mu} + \mu\frac{\partial}{\partial\rho}\right) D_\rho^2 \Psi_1(\rho,\mu) \tag{4.80}$$

This result is also derived in matched asymptotic analysis, and is formally identical to the Oseen equation for a sphere (Eqn. 3.9). Structurally, this problem is similar to what we have seen previously, and is solved in two steps[27]. First use the transformation $D_\rho^2\Psi_1 = e^{\rho\mu/2}\Phi(\rho,\mu)$ to obtain Eqn. 4.81.[21]

$$\left(D_\rho^2 - \frac{1}{4}\right)\Phi(\rho,\mu) = 0 \tag{4.81}$$

This may be solved to obtain the first integral:

$$D_\rho^2\Psi_1(\rho,\mu) = e^{\frac{1}{2}\rho\mu}\sum_{n=1}^{\infty}\left(\mathcal{A}_n\left(\frac{\rho}{2}\right)^{\frac{1}{2}} K_{n+\frac{1}{2}}\left(\frac{\rho}{2}\right) + B_n\left(\frac{\rho}{2}\right)^{\frac{1}{2}} I_{n+\frac{1}{2}}\left(\frac{\rho}{2}\right)\right) Q_n(\mu) \tag{4.82}$$

As in the case of the cylinder, the inhomogeneous terms on the RHS of Eqn. 4.82 consists of integration constants which multiply the two modified Bessel functions. We are beset by the same considerations, which (properly speaking) must be resolved by applying boundary conditions (Eqn. 4.76) to the renormalized solution. Following the same arguments given for the cylinder, we set the coefficients $B_n = 0$, which will later make it possible to satisfy the boundary conditions at infinity.

Completing the second integration is difficult, but was accomplished by Goldstein[27]. The requisite solution is essentially the second term in Eqn. 3.20:

$$\Psi_1^{(\mathrm{a})}(\rho,\theta) = A_1\rho^2 Q_1(\mu) + \sum_{n=1}^{\infty}\left(B_n\rho^{-n} + \sum_{m=0}^{\infty} X_m\rho^2\Phi_{m,n}(\rho/2)\right) Q_n(\mu) \tag{4.83}$$

Note that we have omitted the terms $A_n r^n Q_n(\mu)$ which diverge too quickly at infinity (this was also done for the cylinder).

Alternatively, one may simplify the series in Eqn. 4.82, by retaining only the $n = 1$ term (setting all other $A_n = 0$). It is then possible to complete the second integration

[21] $D_\rho^2\Psi_1(\rho,\mu)$ is the vorticity.

with a closed form solution:

$$\Psi_1^{(\mathrm{b})}(\rho,\theta) = A_1\rho^2 Q_1(\mu) + \mathcal{A}_1\,(1+\mu)\left(1 - e^{-\frac{1}{2}\rho(1-\mu)}\right) + \sum_{n=1}^{\infty} B_n \rho^{-n} Q_n(\mu) \qquad (4.84)$$

As before, we neglect the $A_n r^n Q_n(\mu)$ solutions. This is essentially Oseen's solution (Eqn. 3.12), expressed in the appropriate variables and with undetermined coefficients.

We therefore have two solutions (Eqns. 4.83, 4.84) which can be used for Ψ_1. For the moment, we will consider both. We will later demonstrate that the former is the preferred choice by considering boundary conditions.

4.3.5 Secular Behavior

We consider our $\mathcal{O}(R^1)$ naïve solution abstractly:

$$\Psi(\rho,\mu) = -\rho^2 Q_1(\mu) + R\left(A_1\rho^2 Q_1(\mu) + \sum_{n=1}^{\infty} B_n\rho^{-n} Q_n(\mu) + \cdots\right) + \mathcal{O}\left(R^2\right) \qquad (4.85)$$

This generic form encompasses both Eqn. 4.84 and Eqn. 4.83. It also possesses two key similarities with both the terrible and the cylinder problems. First, there is a term at $\mathcal{O}(R^1)$ which is a multiple of the $\mathcal{O}(R^0)$ solution ($A_1\rho^2 Q_1(\mu)$). Secondly, the secular behavior in our naïve solution occurs *at the same order* as the integration constants which we hope to renormalize.[22] This fact is in essence related to equations like Eqn. 4.14, which must be solved iteratively. We avoided that kind of RG equation by introducing the constant which could have been associated with the $\mathcal{O}(R^0)$ solution at $\mathcal{O}(R^1)$. But renormalizing divergences into integration constants *at the same order* limits the ability of RG to "re-sum" our naïve series. In all of these cases, the real power of RG techniques could be seen by extending our analysis to $\mathcal{O}(R^2)$.

Because of the similarities between Eqn. 4.85 and Eqn. 4.61, we can tackle this problem in a manner formally the same as the cylinder. By construction, Eqn. 4.85 is $\mathcal{O}(\rho^2)$ as $\rho \to \infty$. Hence the only terms with problematic secular behavior occurs

[22] These secular terms are not written explicitly in Eqn. 4.85. They can be found in Eqns. 4.84 and 4.83.

in the limit $\rho \to 0$. As before, these divergences need not even be explicitly identified. We write:

$$\Psi_(\rho,\mu) = -\rho^2 Q_1(\mu) + R\left(A_1\rho^2 Q_1(\mu) + \mathcal{R}(\rho,\mu;\{B_i\};\{X_j\}) + \mathcal{S}(\rho,\mu;\{B_m\};\{X_n\})\right) \tag{4.86}$$

Here, $\mathcal{S}$ includes the terms which are secular as $\rho \to 0$, and $\mathcal{R}$ includes regular terms.

4.3.6 Renormalization

Eqn. 4.86 is only cosmetically different from Eqn. 4.62. Renormalizing the two equations can proceed in *exactly* the same fashion. Therefore, we may immediately write the renormalized solution:

$$\Psi_(\rho,\mu) = -\rho^2 Q_1(\mu) + R\left(\alpha_1\rho^2 Q_1(\mu) + \mathcal{R}(\rho,\theta;\{\beta_i\};\{\chi_j\}) + \mathcal{S}(\rho,\theta;\{\beta_m\};\{\chi_n\})\right) \tag{4.87}$$

This is, of course, the same solution from which we began. As in the previous two problems, we now know that it is a uniformly valid solution, and turn to the application of the boundary conditions.

4.3.7 Meeting the Boundary Conditions

We have two possible solutions for $\Psi_1(\rho,\mu)$. Considering the boundary conditions on the surface of the sphere (Eqn. 4.76) will demonstrate why Eqn. 4.83 is preferential. Eqn. 4.84 can never satisfy the two requirements for all of the angular harmonics. Expanding the exponential term, we see that although it has but one integration constant, it contributes to *all* of the powers of μ. The second solution, Eqn. 4.83, can meet both of the boundary conditions — in principle. However, as in the case of the cylinder, this is practically impossible, and we must consider truncating our solution.

It is clear that we will need to approximate our solutions in order to apply the boundary conditions. Our procedure is governed by the following considerations. First,

we demand that our approximate solution satisfy the boundary conditions as accurately as possible. This requirement is necessary because our goal is to calculate the drag coefficient, C_D, a calculation which is done by evaluating quantities derived from the stream function *on the surface of the sphere*. Hence it is necessary that the stream function be as accurate as possible in that regime. Secondly, we want the difference between our modified solution and the exact solution (one which satisfies the governing equations) to be as small as possible.

Oseen's Solution

First, consider trying to satisfy these requirements starting from Eqn. 4.84. Although this is the less general solution to Oseen's equation, we consider Oseen's solution because of (1) its historical importance, including widespread use as a starting point for matched asymptotics work and (2) the appealling simplicity of a closed-form solution.

We combine Eqns. 4.84 and 4.79 to begin from the solution: $\Psi(\rho,\mu) = \Psi_0(\rho,\mu) + R\Psi_1^{(\mathrm{b})}(\rho,\mu)$. Since we are interested in the solution near the surface of the sphere ($\rho = R$), and because there is no other way to determine the integration constants, we expand the exponential in that vicinity. Retaining terms up to $\mathcal{O}\left(R\rho^1\right) \sim \mathcal{O}\left(\rho^2\right)$, we obtain:

$$\Psi(\rho,\mu) = \left(-\rho^2 + R\left(A_1\rho^2 - \mathcal{A}_1\rho\right)\right) Q_1(\mu) + R\sum_{n=1}^{\infty} B_n \rho^{-n} Q_n(\mu) \tag{4.88}$$

The boundary conditions are satisfied if $B_n = 0\ \forall n > 1$, $A_1 = 0$, $\mathcal{A}_1 = -3/2$, and $B_1 = -R^2/2$. In passing, we note that substituting these values into Eqn. 4.84 reproduces Oseen's original solution[32]. Continuing, we substitute these values into Eqn. 4.88, obtaining:

$$\Psi(\rho,\mu) = \left(-\rho^2 + \frac{3R\rho}{2} - \frac{R^3}{2\rho}\right) Q_1(\mu) \tag{4.89}$$

This is nothing more than Stokes' solution (Eqn. 3.2), albeit expressed in Oseen variables. Consequently, when substituted into Eqns. 2.10, 2.23, and 2.24 Eqn. 4.89

reproduces $C_D = 6\pi/R$.

How accurate is our approximate solution? The difference between Eqn. 4.89 and Eqn. 4.84 is given by:

$$\Delta\Psi = -\frac{3}{4}R\left(1+\mu\right)\left(-2+2e^{-\frac{1}{2}\rho(1-\mu)}+\rho\left(1-\mu\right)\right) \tag{4.90}$$

At the surface of the sphere ($\rho = R$), this equates to an $\mathcal{O}\left(R^3\right)$ error in the stream function, and an $\mathcal{O}\left(R^2\right)$ error in the derivative. That is entirely acceptable. However, at large ρ, $\Delta\Psi$ grows unbounded, being of $\mathcal{O}\left(\rho^1\right)$. This is the fundamental problem with the solution given by Eqn. 4.89. By beginning from Eqn. 4.83, we can avoid this difficulty.

It is at first a little disconcerting that Oseen used his solution to obtain the next approximation to C_D (Eqn. 3.13)[78]. How can our results be worse? As remarked previously, Oseen got lucky making approximations in how he calculated the drag coefficient. As Proudman put it, "Strictly, Oseen's method gives only the leading term ... and is scarcely to be counted as superior to Stokes' method for the purpose of obtaining the drag."[26]

Goldstein's Solution

We now apply the boundary conditions to Eqn. 4.83. By starting from the more general solution to Oseen's equation, we can remedy the difficulties encountered above. This analysis will be very similar to the truncation performed on Tomotika's solution for the cylinder problem.

We combine Eqns. 4.83 and 4.79 to begin from the solution: $\Psi(\rho,\mu) = \Psi_0(\rho,\mu) + R\Psi_1^{(\mathrm{a})}(\rho,\mu)$. As with the cylinder, we will approximate the full solution by truncating the series in both m and n. Our first consideration is again symmetry: The uniform flow imposes a $\sin\theta$, or $Q_1(\mu)$ symmetry on the problem. Hence we must retain the $n = 1$ term in Eqn. 4.83. The importance of this term is clearly seen from Eqn. 2.25:

n =	1	2	3	4
$\Psi^{(n)}_{\text{discard}}(\rho,\mu)$	$\mathcal{O}(R^3)$	$\mathcal{O}(R^2)$	$\mathcal{O}(R^1)$	$\mathcal{O}(R^0)$
$\Psi'^{(n)}_{\text{discard}}(\rho,\mu)$	$\mathcal{O}(R^2)$	$\mathcal{O}(R^1)$	$\mathcal{O}(R^1)$	$\mathcal{O}(R^{-1})$

Table 4.2: Importance of discarded terms at $\rho = R$

Only the coefficient of $Q_1(\mu)$ is needed to calculate the drag if the stream function satisfies the boundary conditions.

As in the case of the cylinder, if we retain n harmonics, we must retain $m = n-1$ terms in the second sum (the sum over m) in order to meet both boundary conditions. To minimize the error introduced by our approximations we set all *other* B_n, X_m equal to zero. The remaining terms, those which would violate the boundary conditions and must be truncated, are then given by Eqn. 4.91.

$$\Psi^{(n)}_{\text{discard}}(\rho,\mu) = R\left(\sum_{k=n+1}^{\infty}\sum_{m=0}^{n-1} X_m \Phi_{m,k}(\rho/2)\rho^2 Q_k(\mu)\right) \tag{4.91}$$

We want to estimate the magnitude of the error in our approximation, both overall and at the surface (the error in the boundary conditions). The error is given by Eqn. 4.91. First, we calculate the magnitude of both $\Psi^{(n)}_{\text{discard}}(\rho,\mu)$ and its derivative at the surface ($\rho = R$) with n retained harmonics. The results are given in Table 4.2.

From Table 4.2, we see that to retain the $Q_2(\mu)$ harmonics, we must have an error in our derivative boundary condition of $\mathcal{O}(R^1)$ — the order to which we are trying to work. If we retain higher harmonics, this situation gets worse.

Since it is in practice impossible to fit the boundary conditions to all harmonics, we must truncate the series expansion. We see that there is only one truncation consistent with both the symmetry requirements of the problem and the demand that we satisfy the boundary conditions to $\mathcal{O}(R^1)$:

$$\Psi(\rho,\mu) = -\rho^2 Q_1(\mu) + R\left(A_1\rho^2 + B_1\rho^{-1} + X_0\Phi_{0,1}(\rho/2)\rho^2\right)Q_1(\mu) + \mathcal{O}\left(R^2\right) \tag{4.92}$$

We also must consider the overall error, e.g., how big can $\Psi^{(1)}_{\text{discard}}(\rho,\mu)$ get? Although,

at the surface of the sphere, Eqn. 4.92 is no better than Eqn. 4.89, it is superior for $\rho \neq R$. The magnitude of the error is maximized as as $\rho \to \infty$. It can be shown that $\Phi_{m,n}(x \to \infty) \sim x^{-2}$. Therefore,

$$\lim_{\rho \to \infty} \Psi^{(1)}_{\text{discard}}(\rho, \mu) = \mathcal{O}\left(R^1\right)$$

Although this is somewhat unsatisfactory, this solution does not suffer from the same shortcomings as Eqn. 4.88. The error remains bounded.

Eqn. 4.92 will satisfy the boundary conditions (Eqn. 4.76) if $A_1 = 0$ and

$$X_0 = \frac{6}{6R\Phi_{0,1}(R/2) + R^2\Phi'_{0,1}(R/2)} \tag{4.93}$$

$$B_1 = \frac{R^3\Phi'_{0,1}(R/2)}{6\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)} \tag{4.94}$$

As in the case of the cylinder, the resulting stream function satisfies the boundary conditions exactly, and the governing equations approximately. Our final solution is:

$$\Psi(\rho, \mu) = -\rho^2 Q_1(\mu) + R\left(\frac{R^3\Phi'_{0,1}(R/2)}{6\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)}\rho^{-1} + \frac{R^3\Phi'_{0,1}(R/2)}{6\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)}\Phi_{0,1}(\rho/2)\rho^2\right)Q_1(\mu) + \mathcal{O}\left(R^2\right) \tag{4.95}$$

For reference,

$$\Phi_{0,1}(x) = -\frac{3\pi}{4x^2}\left(2 - \frac{2}{x} + \frac{1}{x^2} - \frac{e^{-2x}}{x^2}\right)$$

4.3.8 Calculating the Drag Coefficient

We calculated the drag coefficient by substituting Eqn. 4.95 into Eqn. 2.25, giving this new result:

$$C_D = \frac{\pi\left(-16\Phi'_{0,1}(R/2) + R\left(8\Phi''_{0,1}(R/2) + R\Phi'''_{0,1}(R/2)\right)\right)}{2\left(6\Phi_{0,1}(R/2) + R\Phi'_{0,1}(R/2)\right)} \tag{4.96}$$

This can be expressed in terms of more conventional functions by substituting for $\Phi_{0,1}(x)$, resulting in the drag coefficient given by Eqn. 4.97.

$$C_D = \frac{4\pi\left(24 + 24R + 8R^2 + R^3 + 4e^R\left(R^2 - 6\right)\right)}{R\left(2\left(R + 1\right) + e^R\left(R^2 - 2\right)\right)} \tag{4.97}$$

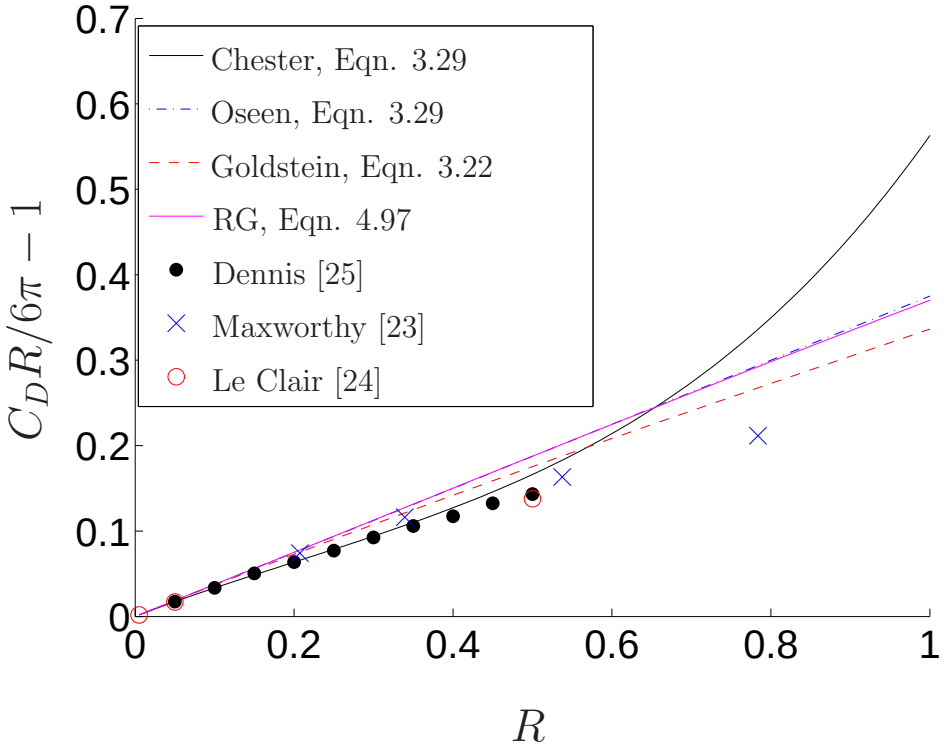

Figure 4.3: Drag on a sphere, comparing RG to other theories.

This result is plotted in Figure 4.3, where it is compared against the principal results of Oseen theory, matched asymptotic theory, numerical results, and experiments. As $R \to 0$, there is excellent agreement. At small but non-infinitesimal Reynolds numbers, RG is nearly identical to Oseen's prediction (Eqn. 3.13), which is disappointing. It is surprising that Goldstein's result is better than the RG result, as they are calculations of the same order in R, and are a series approximation. That the matched asymptotics predictions are superior is not surprising; Chester and Breach's result began with a much higher order perturbative approximation. If a higher order RG calculation were possible, RG ought to be better than the same order matched asymptotics prediction.

As in the case of the cylinder, the real strength of Eqn. 4.97 can be seen as the Reynolds number increases. Figure 4.4 demonstrates that all other theories diverge from experimental measurements for $R \gtrsim 1$. This is an unavoidable aspect of their structure and derivation — they are only valid asymptotically. The RG prediction

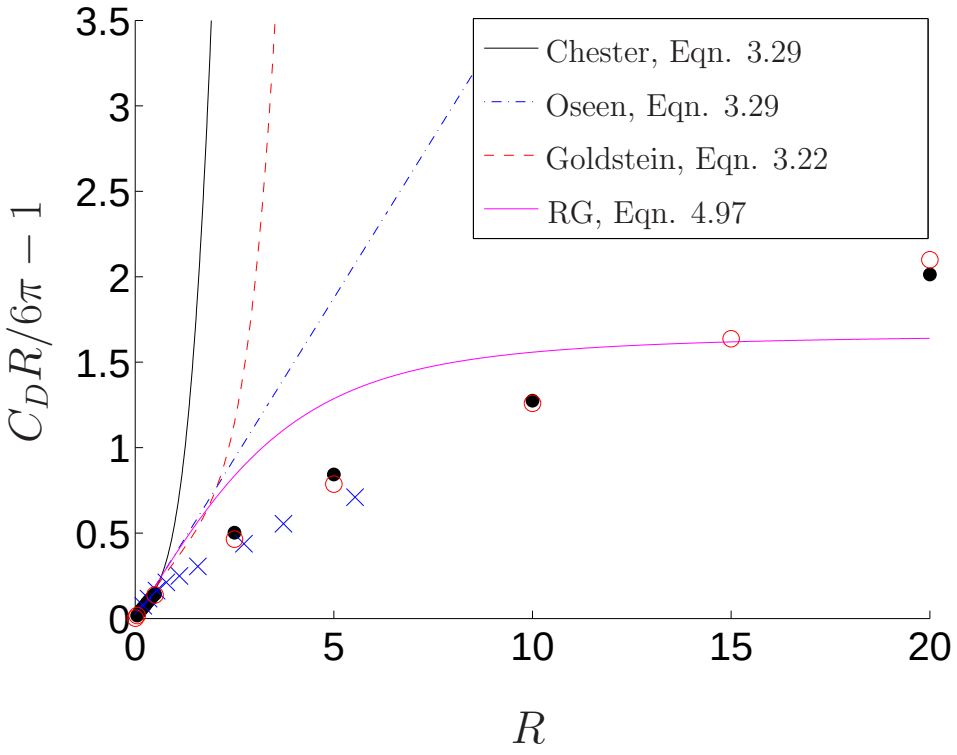

Figure 4.4: Drag on a sphere, comparing RG at larger R.

suffers from none of these problems. Eqn. 4.97 is well behaved for all R, although it does become less accurate at larger Reynolds numbers.

4.4 Conclusions

We have devoted a substantial effort to the historical problem of calculating the drag coefficient for flow around a cylinder and a sphere at low Reynolds number. We report four principal accomplishments. First, we have untangled over 150 years of diffuse, confusing, and sometimes contradictory experimental, numerical, and theoretical results. We have expressed all important previous work within a consistent mathematical framework, and explained the approximations and assumptions which have gone into previous calculations. Moreover, by plotting experimental results and theoretical predictions with the leading order divergence removed (an idea originally due to Maxwor-

thy), we have consistently and critically compared all available measurements. There are no other such exhaustive comparative reviews available in the existing literature, which is piecemeal at best. This review is an invaluable asset for anyone else studying these problems.

Secondly, we have extended traditional matched asymptotics calculations. We advance and justify the idea that *uniformly valid approximations*, not the Stokes or Oseen expansions[23], should be used to calculate derivative quantities such as C_D. By combining this approach with previously published matched asymptotics results, we obtain new results for the drag coefficients. These results *systematically* improve on published drag coefficients, which relied only on the Stokes expansion. This methodology also resolved a problem in the existing literature: the most accurate calculations for a cylinder, due to Skinner, had failed to improve C_D[38]. When treated via a uniformly valid approximation, our new result based on Skinner's solutions betters all matched asymptotics predictions.

We have also explored the structure and subtleties involved in applying renormalization group techniques to the "terrible" problem posed by Hinch and Lagerstrom[28, 94]. This problem, previously solved by Chen et al.[16], contains a rich and henceforth unexplored collection of hidden subtleties. We exhaustively examined all possible complications which can arise while solving this problem with the renormalization group. To treat some of these possibilities, we identified and implemented a new constraint on the RG calculation; the renormalized perturbation solution itself, not just the expansion on which it is based, must satisfy the governing equations to the appropriate order in ϵ. While this had been done implicitly in previous calculations, we had to deal with it explicitly (e.g., by appropriate choices of homogeneous solutions). In the process of doing so, we obtained several new second order solutions to the "terrible" problem, and demonstrated their equivalence.

[23] "Inner" and "Outer" approximations in the usual matched asymptotics literature.

The work with the "terrible" problem laid the foundation for our most significant new calculation. In close analogy with the "terrible" problem, we used the RG to derive new results for the drag coefficients for both a sphere and a cylinder (Eqns. 4.97 and 4.72, respectively). These new results agree asymptotically with previous theoretical predictions, but greatly surpass them at larger R. Other theories diverge pathologically, while the results from the RG calculation remain well behaved.

We demonstrated that these new techniques could reproduce and improve upon the results of matched asymptotics — when applied to the very problem which that discipline was created to solve! Matched asymptotics requires the use of two ingenious and intricate expansions, replete with strange terms (like $R \log R$) which must be introduced while solving the problem via a painful iterative process. RG requires only a single generic expansion, which can always be written down a priori, even in complicated singular perturbation problems with boundary layers. It therefore gives rise to a much more economical solution, requiring half the work and yielding a superior result. It is hoped that demonstrating of the utility of these techniques with this seminal problem will result in increased interest and more widespread application.

Chapter 5

Biocomplexity at Yellowstone National Park

5.1 Introduction

The remaining chapters present my work related to carbonate hot springs at Yellowstone National Park (YNP). This and the following two chapters focus on my field work, experimental data, and subsequent analysis. This chapter motivates our work, introduces the Yellowstone hot springs, broadly outlines our field work, and builds the theoretical framework needed to organize our data. Chapters 8 and 9 present numerical models of our system which build on experimental observations and conclusions. This work has produced three papers which have been submitted for publication[19, 20, 99].

The YNP project is part of a broad collaboration which includes geologists, geomicrobiologists, geochemists, microbial ecologists, and physicists. The geologists are led by Professor Bruce Fouke (UIUC), the physicists by Professor Nigel Goldenfeld (UIUC), and the microbial ecologists by Professor Alison Murray (Desert Research Institute). The overarching goal of this project is to explain the unique terraced patterns which form at carbonate hot springs (Figure 5.1). In particular, we want to understand how microbes effect carbonate deposition, and determine whether they play an

important role in the formation of the characteristic patterns.

At locations throughout the world[100], geothermal hotsprings flow upward through underground limestone formations. As the hot water flows to the surface, it becomes saturated with calcium carbonate ($CaCO_3$). Once the saturated water reaches the surface, it often flows downhill. Outside of the geothermal vent, the water undergoes numerous changes, including cooling, outgassing of CO_2, and the precipitation of carbonate. This process can create spectacular carbonate ponds and terraces which are shown in Figure 5.1. These terraces are composed of *travertine*, a general term which refers to a non-marine terrestrial hot spring carbonate deposit. The terraces appear to have the same morphology regardless of size; they seem to be scale-invariant over a range of scales from millimeters to hundreds of meters[100]. In other words, small ponds appear to be a simple re-scaling of larger ponds. Scale-invariance is a common occurrence in geology[101], and is almost always accompanied by universality[102]. In these cases, scale invariant features can be described by simple models, which are insensitive to the specific microscopic parameters in the system.

5.2 Why Yellowstone?

This project is interesting for us, as physicists, for several reasons. Firstly, in the context of Nigel Goldenfeld's program of research into pattern formation, it represents a new level of complication — shallow, turbulent fluid flow coupled to a moving boundary problem. Can we compute the correlation functions and dynamics of such structures? Secondly, this problem is one of the simpler, non-trivial examples of the interaction between organisms and their host ecosystem. As such, the tools, techniques, and experience gained at YNP may one day be applicable to understanding the global interactions between microbial life and the rest of the terrestrial biosphere, and leading to an improved understanding of global climate change, for example. Thirdly, our group

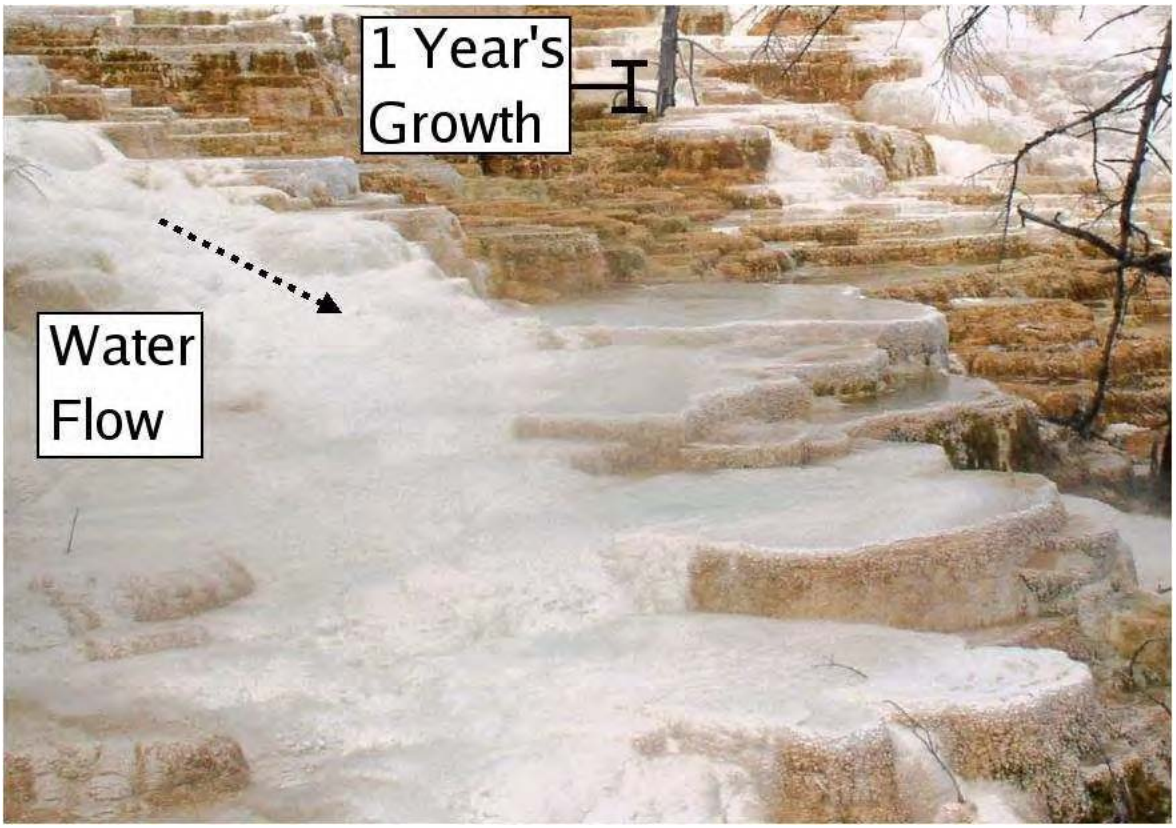

Figure 5.1: Canary Terrace, Yellowstone National Park. This depicts the typical terraced architecture seen at Yellowstone National Park. A channel of water is seen entering into ponds in the foreground. The background shows the amount (almost 0.5 m) of carbonate deposited in a one year period after this photo was taken. Photo by Brian Suderman.

would like to initiate a systems biology of the environment, mapping out the coupling between intracellular metabolic pathways, community metabolism, and environmental biogeochemical cycles, with a view toward identifying any universal patterns of flux distribution or global organization. My project is part the first steps in this overall program.

The Yellowstone hot springs are ideal for our study of interactions between microbial activity, fluid flow, and the underlying landscape. Unlike most geologic processes, the hot springs evolve on a time scale of days to weeks; most rocks form over eons, but carbonate precipitation rates at YNP can be as high as 1-5mm per day[103, 104, 105]. For geology, a science which regularly measures time in millions of years, this is

incredibly fast. One can — and we have! — make a time-lapse movie which shows rocks *growing* over a time period of a year. Figure 5.1 depicts the almost half a meter of travertine which precipitated in a single year.

Our studies at Yellowstone enable us to relate a rock to the exact depositional environment in which it was produced. Travertine deposits are characterized by specific crystal fabrics[103], which are preserved in the rock record[106, 107]. Working in the modern environment, we correlate travertine fabrics with the exact depositional environments from which they precipitated. If those correlations are significant, we can infer the depositional environment of any identifiable fabrics in the rock record.

This seemingly simple statement has profound consequences. Two examples are astrobiology and environmental reconstructions. An example of the former is the "Martian Meteorite" (ALH84001) found in Antarctica in 1984[109]. It has been claimed that markings on this rock indicate that it is biogenic in origin, implying the existence of life on Mars[110, 111, 112]. Figure 5.2 shows a picture containing these putative markings. Many other authors argued that these markings do *not* indicate biogenic origins[113]. This confusion arose because it is not well understood how bacteria can influence the formation of rocks. By characterizing the microbial environment in which carbonate rocks actually form, we can help resolve this question. Correlating physical, chemical, and microbial attributes of the modern depositional environment with unique travertine features also allows us to extrapolate ancient environments from the rock record here on Earth. This process is known as *environmental reconstruction*.

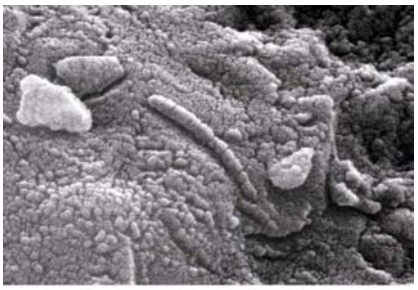

Figure 5.2: High-resolution scanning electron microscopy image of the martian meteorite ALH84001. Markings on this rock may resemble the shape of terrestrial microfossils. Picture taken from NASA website[108].

Part of the reason we are focusing on hot springs is because they are an excel-

lent laboratory in which to study in detail the interactions between microbes and precipitating travertine. It has been suggested[114, 115] that microbial activity plays a significant role in travertine precipitation. For instance, given that the hot spring water is supersaturated with respect to $CaCO_3$, precipitation can be limited by the availability of nucleation sites[116]. As such, microbial cells could serve as nucleation sites, thereby enhancing the rate of precipitation. It is even possible that microbial metabolic activities increase deposition by removing CO_2 from the water (e.g. by photosynthesis), or that microbial cells inhibit precipitation to avoid death by entombment. While microbe-rock interactions remain a subject of ongoing investigations, we will present evidence here that macroscopic carbonate patterns, such as terraces, can develop purely abiotically.

Even if feedback with microbial communities does not control the formation of large scale travertine morphology, this system contains a fascinating interplay between the underlying topography and the fluid which flows over it. The fluid advects the dissolved materials that then precipitate, forming travertine, which subsequently alters the fluid flow path. By coupling simple models of spring water flow with models of aqueous chemistry, we explain in Chapters 8 and 9 how the terrace morphology can emerge as a result of the interactions between thin film flows and underlying topography. These minimal models support the qualitative observation that the terraces are scale invariant.

We also study Yellowstone because its hot springs have been protected from human contamination by the National Park Service. As a result, these springs offer an untrammeled laboratory in which to investigate both microbial communities and spring water chemistry.

5.3 Geologic setting

Our field studies took place at two hot-springs located within the Mammoth Hot Springs Complex at Yellowstone National Park, in Wyoming. Figure 5.9 shows the location of the park, the hot springs complex, and the two specific hot springs at Angel Terrace that we have focused on, named AT-1 and AT-3. Mammoth Hot Springs lies on the northern flank of the Yellowstone caldera, and contains a succession of travertine deposits that range from 0 to approximately 8,000 years old, are 73 m in thickness and cover more than 4 km^2[117, 118, 119, 120, 121]. The springs expel $Ca - Na - HCO_3 - SO_4$ type hot waters derived from a subsurface reservoir at temperatures greater than 100^{o}C[122, 123]. Because the water at all of the springs in the Mammoth Complex comes from a common reservoir, and percolates through similar deposits (particularly Mission Canyon Limestone), the chemistry of the emergent spring water is nearly identical. We have explicitly verified this at AT-1 and AT-3.

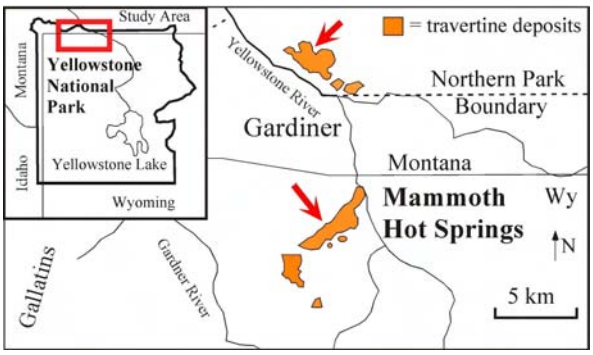

Figure 5.3: Geographic location of Mammoth Hot, with inset map showing of all of YNP.

The Mammoth complex contains an estimated 100 springs, and travertine deposits which range from zero to 400,000 years old[118, 124, 125, 126]. Angel Terrace, near the top of the complex, contains several active small springs, including the two which we have intensively studied (AT-1 and AT-3). Figure 5.4 shows the location of Angel Terrace, as well as the AT-1 and AT-3. Photos of these two springs are shown in Figure 5.5. The hydrologic system is dynamic, with multiple vents appearing, sealing, and reopening on Angel Terrace at a frequency of months to tens of years[118, 123]. During my 4 years working on these springs, AT-1 dried up completely, and the location of the vent at AT-3 has moved about 5 meters. These changes are accompanied by significant changes in total flux, and have a large impact on the patterns which form

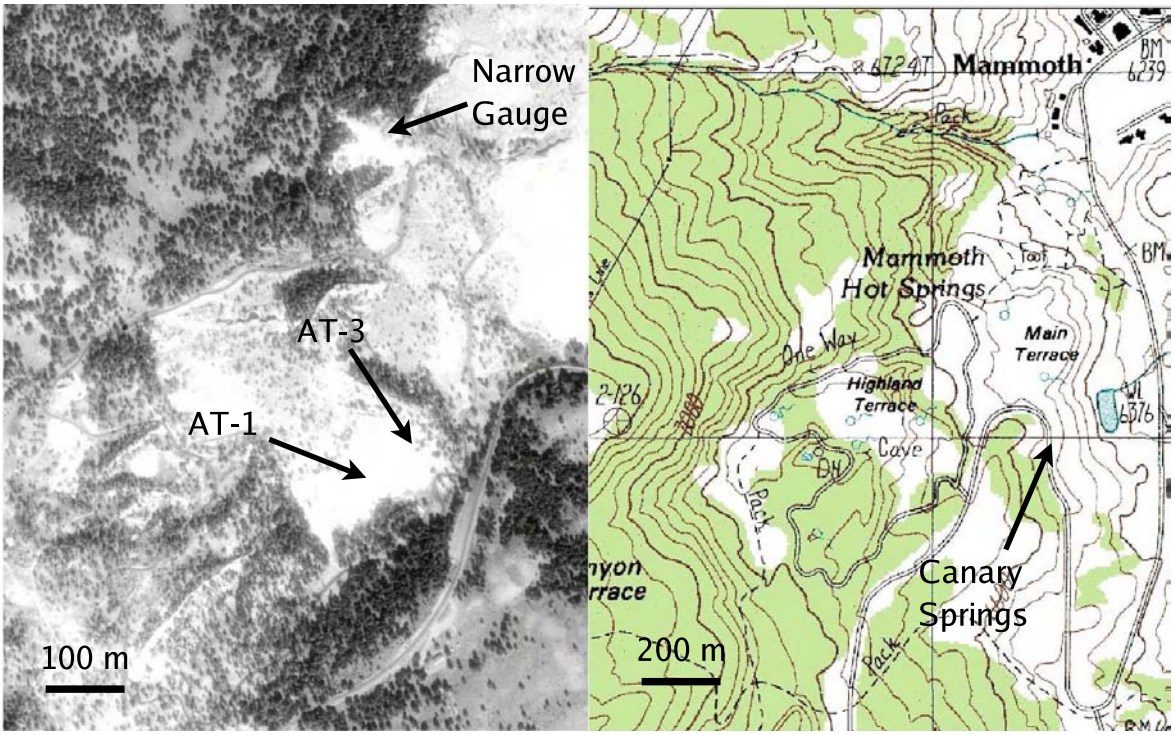

Figure 5.4: Overhead imagery (left) and topographic map of Angel Terrace, showing the location of spring AT-1 and AT-3. The overhead shot shows bright white carbonate thermal features. The topographic map is a digital raster graphic (DRG) produced by the USGS from a 1986 USGS National Mapping Division map. The overhead image was extracted from a Color Infra-Red (CIR) digital orthophoto quadrangle provided by the USGS. It was photographed on 03 Sep 1991 by the National Aerial Photography Program and digitized by the USGS DOQ Program on 02 Dec 1992.

in the drainage system.

Several distinct morphologies are found in the complex. In addition to the *terraces* mentioned, there are *fissure ridges*, which arise from a line (as opposed to a point) source of spring water, and *caps*, which are towers of travertine with a rounded top. Finally, we have reported and explained a new feature, which we term a *dome*[20]. An example of this kind of feature is shown in Figure 5.6, and they are discussed in detail in Chapter 9.

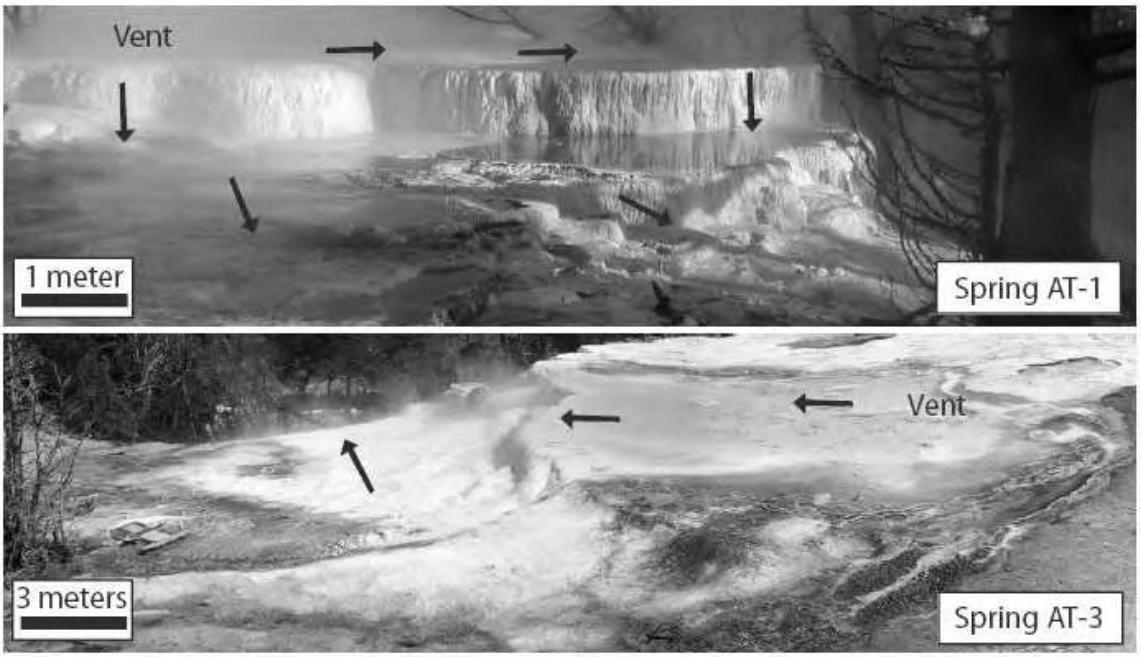

Figure 5.5: Spring AT-1 and AT-3 showing vent and general direction of spring water flow.

5.4 Carbonate Chemistry Primer

The spring water emerges from the ground at approximately 71^oC with a pH of about 6.3. This water advects a number of dissolved chemical species, most importantly Ca^{2+} and HCO_3^-. Na^{2+}, K^+, SO_4^-, Cl^- and H_2S are also present in lesser concentrations. Over the course of the drainage system, this water comes into equilibrium with $CaCO_3$, atmospheric CO_2 fugacity, and atmospheric temperatures. As it does so, it exsolves CO_2 and precipitates carbonate. Initially the precipitated carbonate is in the form of *aragonite*; futher downstream it becomes predominantly *calcite*. Both of these crystals have the same chemistry ($CaCO_3$), but different *unit cells*.

We provide here a general introduction to the chemistry involved in carbonate precipitation. The overall precipitation reaction is described by Eqn. 5.1. Among other things, it demonstrates that one molecule of CO_2 must be released for each

Figure 5.6: An example of a travertine dome. Photo taken by Nigel Goldenfeld near Narrow Gauge in Figure 5.4.

molecule of $CaCO_3$ which precipitates.

$$Ca^{2+} + 2HCO_3^- \rightarrow CaCO_3(s) + H_2O + CO_2(g) \uparrow \tag{5.1}$$

This is an *equilibrium* equation, but do not be mislead: our system is controlled by non-equilibrium (kinetic) processes[116, 127, 128, 129, 130, 131]. Nonetheless, the equilibrium equation gives a qualitative understanding of carbonate precipitation and its controls. In particular, precipitation can be controlled by two factors. On the LHS of Eqn. 5.1, supersaturation with respect to calcium or bicarbonate ions can drive precipitation forward. On the RHS, exsolution of carbon dioxide will also drive precipitation.

The equilibrium process described in Eqn. 5.1 is often characterized by the quantities Ω and SI, which describe the *saturation state*, or the equilibrium ability of a fluid to precipitate. Ω is the ratio of the ion activity (or concentration) product to the thermodynamic or stoichiometric solubility product, and is defined in Eqn. 5.2

$$\Omega \equiv \frac{[Ca^{2+}][CO_3^{2-}]}{K_{\rm sp}} \tag{5.2}$$

Here K_{sp} is the activity solubility product constant, which differs for calcite and aragonite. $\Omega < 1$ indicates undersaturation (and dissolution can occur), $\Omega > 1$ supersaturation (precipitation is possible), and $\Omega = 1$ corresponds to equilibrium. Workers also discuss the *saturation index*, SI, which is defined in Eqn. 5.3.

$$SI_C \equiv \log_{10} \Omega \tag{5.3}$$

Positive SI indicates precipitation, negative SI dissolution, and $SI = 0$ equilibrium.

In our system, Ω is slightly negative when the water emerges from the vent, but it rapidly becomes (and stays) positive. Notice that these two measures of saturation state differ for calcite and aragonite because these minerals have different solubility product constants (K_{sp}). If we use our measurements of aqueous chemistry, and calculate SI_{calcite} and $SI_{\mathrm{aragonite}}$, equilibrium chemistry implies that SI_{calcite} is almost always larger than SI_{calcite}, and that calcite precipitation is always favored in our system. This contradicts with observations, and clearly indicates that precipitation is a non-equilibrium process.

The overall reaction (Eqn. 5.1) describes equilibria amongst liquid, solid, and gaseous phases. Reducing it into its constituent reactions gives some insight into the kinetic processes which control precipitation. There are two established models which predict calcite precipitation rates. The first is commonly called the *PWP model*[131], and the second is a *boundary layer* model[132]. The latter model explains precipitation using Eqns. 5.4 and 5.5.

$$H^+ + HCO_3^- \rightarrow H_2CO_3 \rightarrow CO_2(g) \uparrow + H_2O \tag{5.4}$$

This equation describes the relatively slow conversion of HCO_3^- and H^+ to CO_2[132]. The second reaction describes the formation of carbonate:

$$Ca^+ + 2HCO_3^- \rightarrow CaCO_3 + HCO_3^- + H^+ \tag{5.5}$$

In their model of carbonate kinetics, Dreybrodt et al. demonstrate that there are two limiting regimes of carbonate precipitation kinetics[132]. If V denotes the volume

of fluid, A the area of the mineral surface where precipitation is occurring, R the precipitation rate, and $\mathrm{d}CO_2/\mathrm{d}t$ the rate of CO_2 outgassing, then

$$V\frac{\mathrm{d}CO_2}{\mathrm{d}t} = AR \tag{5.6}$$

In the limit of small V/A, the rate of precipitation is controlled by Eqn. 5.4. There is plenty of surface area available for nucleation, and the precipitation rate is controlled by the formation and outgassing of CO_2. In the other limit (large V/A), precipitation rates are controlled by the availability of nucleation sites. Dreybrodt demonstrated that the crossover between these regimes occurs for $V/A \sim 0.1$cm[132]. Both regimes likely play a role in different parts of our system.

In addition to the above studies, there have been numerous laboratory and field studies of cool water calcite precipitation rates (e.g.,[133, 134, 135]), and some studies of hot spring precipitation (e.g., [103, 104, 136]). In Chapter 7, we will relate our own work to some of these experiments, and summarize the general conclusions about the processes which control hot spring precipitation. These conclusions form the basis for the numerical modelling in Chapter 8.

One interesting aspect of Dreybrodt's study was his use of *carbonic anhydrase* in laboratory experiments. This enzyme enhances Eqn. 5.4, and increases precipitation rates fifteenfold. This provides evidence that bacterially produced enzymes could greatly effect carbonate precipitation.

Spring water chemistry is dominated by Ca^{2+} and CO_2. It is useful to characterize the latter via *dissolved inorganic carbon*, or DIC. DIC subsumes all of the species which result when CO_2 is dissolved in water, and is defined by: $DIC \equiv [CO_3^{2+}] + [CO_2] + [H_2CO_3] + [HCO_3^-]$.

There is no complete theory which describes detailed precipitation rates at hot springs. Most previous work has focused exclusively on calcite precipitating cool water springs, and has limited applicability to our system. Other members of our research group are working on a fundamental theory of explaining carbonate nucleation and

precipitation kinetics, as well as determining how these rates depend on microbial communities. This work is beyond the scope of this thesis.

The most fascinating aspect of this precipitation process is that it is essentially irreversible. Once CO_2 exsolves into the atmosphere, and carbonate precipitates, there is no way for it to dissolve. For all intents and purposes, the carbon dioxide cannot re-enter the spring water. This stands in marked contrast to most other process which couple fluid dynamics to an evolving landscape (e.g., [137]), which must consider erosion. Even other process which do not necessarily involve sediment transport, such as salt precipitation, are easily reversible. This novel attribute helps explain the unique geomorphology of carbonate terraces.

5.5 Introduction to the facies model

Most of our experimental work has been within the context of a *depositional facies*. In geology, a *facies* is a body of sedimentary rock which is identifiable and differentiable on the basis of its unique characteristics[138, 139, 140]. It is a concept which is particularly useful for organizing and classifying sedimentary deposits. Traditionally, a facies was defined purely on the basis of the physical (crystal structure and aggregate morphology) and chemical (including isotopic) aspects of the body of rock. The definition of a facies is explored further in Chapter 6.

In 2001, Fouke et al. developed a five component facies model for carbonate hot springs[103]. We have extensively used this model to organize our microbiological, physical, and chemical studies at YNP. More importantly, it has enabled us to systematically compare measurements taken from different springs and times, even though the springs themselves are constantly changing. Here we summarize and define these facies, which are shown schematically in Figure 5.7. The five facies in the Fouke model are:

Vent (V): Springwater emerges from the ground, typically into a pool of water where $\Omega \sim 1$. Travertine consists of aragonite needle botryoids[103].

Apron Channel (AC): Macroscopically, this facies consists of a narrow channel of flow which connects the vent pool to the rest of the system. Microscopically, the facies is "floored by hollow tubes composed of aragonite needle botryoids that encrust sulfide-oxidizing *Aquificales* bacteria."[103] (e.g. Figure 5.8).

Pond (P): Macroscopically, this facies is easily identified by a pool of water 10s of centimeters to meters in diameter which forms behind a travertine dam. These pools of water lead to the characteristic terraces. Its diverse microscopic structure includes aragonite needle shrubs, calcite "ice sheets", calcified bubbles, and aggregated aragonite needles ("fuzzy dumbells")[103].

Proximal Slope (PS): This facies is macroscopically characterized by small (centimeter sized) "microterracettes", which often form on a steep slope. Microscopically these are composed of "arcuate aragonite needle shrubs."[103]

Distal Slope (DS): This facies is easily identified, as it occurs near the end of precipitation. It is the only facies dominated by calcite mineralogy. Microscopically, the calcite forms spherules and "feather" crystals[103].

During the course of our studies and simulations, we have realized that the characteristic morphology is "terrace-like" rather than "pond-like". This is explained further in Chapter 9. Nonetheless, we use the previously established nomenclature when referring to the pond facies.

Remarkably, we find that the physical structures characteristic of each facies develop sharp boundaries instead of gradual transitional zones[18]. Although any given travertine facies may be as much as 10s of meters long and cover 100s of square meters in area, the boundary between facies is relatively abrupt, occurring over as little as 1

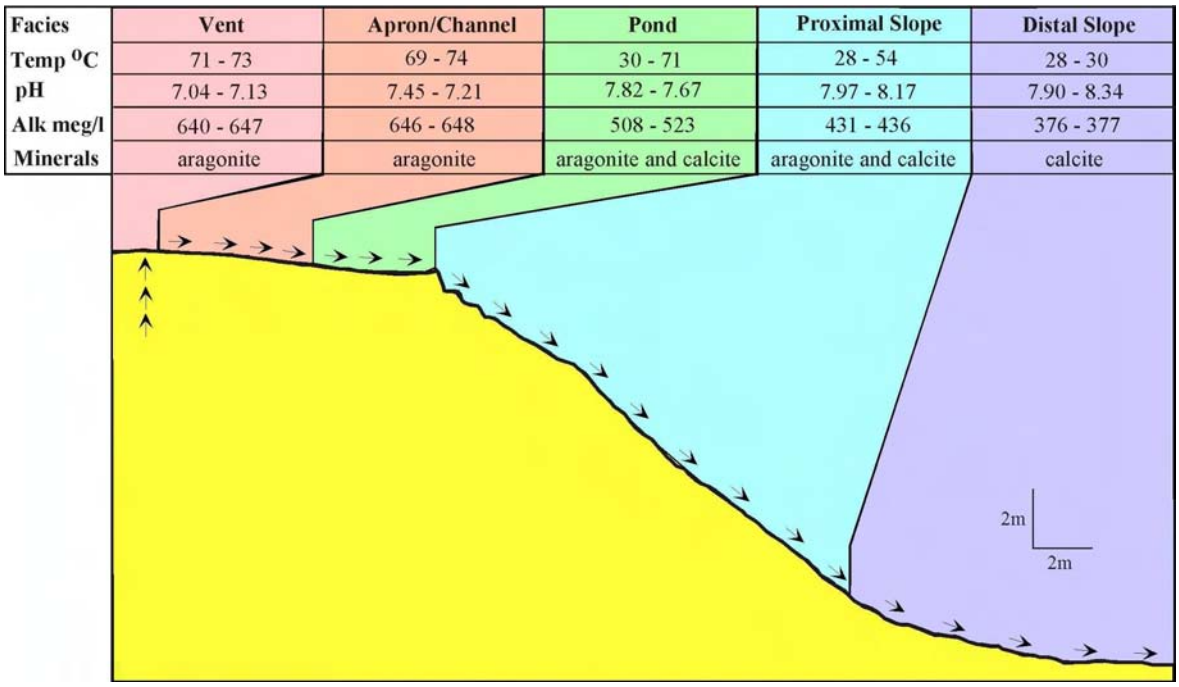

Figure 5.7: The five facies which describe carbonate hot springs are indicated, along with typical minerology, temperature, and pH. The water erupts at the vent at high temperature and flows into the apron channel. It then flows into the pond or proximal slope facies. It ultimately ends up in the distal slope, where chemical equilibrium has nearly been reached and precipitation rates have fallen. Figure taken from ref. [18].

cm in distance between the pond (which is 1-3 m in length) and proximal slope (10-15 m in length) or up to 10 cm between the proximal slope and distal slope (10-15 m in length).

Chapter 6 reports significant and reproducible correlations between the underlying depositional facies and the spring water from which they precipitated. The distinct facies are also invaluable guides to modelling these hot springs; we will show that our minimal numerical models can reproduce all five facies.

5.6 Many Important Length Scales

A facies is defined by unique travertine features on all length scales, and cannot necessarily be identified by its characteristics on any one scale. For example, on the scale

of microns, the apron channel facies is comprised of aragonite needles radiating from the outer walls of filamentous bacteria. But on the scale of centimeters to meters (Fig. 5.8), this facies is a 1-meter wide channel floored by streamer pavements[18, 103]. Because different processes control carbonate precipitation at each scale, we have analyzed the physical, chemical, and microbiological attributes of the depositional environment within the following spatial hierarchy: (1) "microscopic" on the scale of microns; (2) "mesoscopic" on the scale of millimeters to centimeters; (3) "macroscopic" on a scale of meters to 10s of meters; and, (4) "system-level" on the scale larger than 10s of meters.

Each of the five facies defined in Section 5.5 incorporates the first three levels of spatial description, whereas the system-level scale subsumes *all* five of the facies. Figure 5.1 gives an example of a macroscopic pattern, which includes both pond and proximal slope facies. Figure 5.8 gives an example of carbonate patterns on the mesoscopic level, and Figure 5.9 gives an example of carbonate precipitation on the microscopic level.

Figs. 5.8 and 5.9 both evidence signs of microbial influence on carbonate features. The latter shows examples of microbial inclusions on the smallest level and of the influence of extracellular polysaccharide (EPS) on a slightly larger level. However, microbes impact carbonate patterns differently at each of these three length scales. When discussing the process which control deposition, especially microbial processes, we must be very careful to specify the length scale which these processes influence. Process which are very important at the microscopic level may have no impact on macroscopic pattern formation.

5.7 Characterization of spring water

Correlation of solid-phase travertine facies with the spring water has required quantification of physical and chemical aqueous parameters within a complex, heterogeneous natural environment that exhibits large spatial variations and temporal fluctuations.

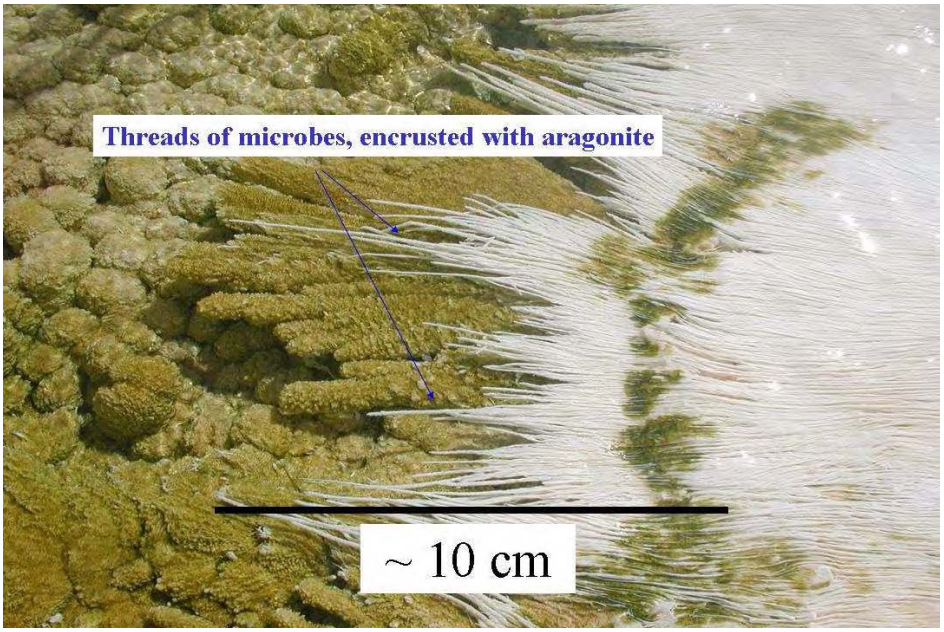

Figure 5.8: This shows the floor of the apron channel. Bacterial filaments are encrusted in aragonite. Picture courtesy of Bruce Fouke and George Bonheyo.

Over a period of seven years, our research group has collected measurements, summarized in Table 5.1, of the physical, chemical, and biological properties of two hot springs (AT-1 and AT-3) at Angel Terrace. I took part in the 2002, 2003, and 2004 expeditions.

The springs were surveyed using a Brunton compass, a 30 m steel tape measure, and a Garmin Model 12 GPS unit. The locations of all sample sites were determined with respect to the vent. During each trip, samples were collected along *transects*, which we define as groups of measurements taken at nominally the same time at locations beginning with the vent and proceeding downstream through the drainage system. All measurements were taken in triplicate; the mean is taken as our best estimate of the true value, and the standard deviation quantifies measurement accuracy.

343 triplicate pH measurements were taken *in situ*, using three types of temperature-correcting hand held probes: a Hach sensION 156 meter; an Orion Model 290A probe; and an Oaktron Waterproof Series 300 meter. Different meters were needed because the spring environment rapidly degrades and destroys probes. The meters were cal-

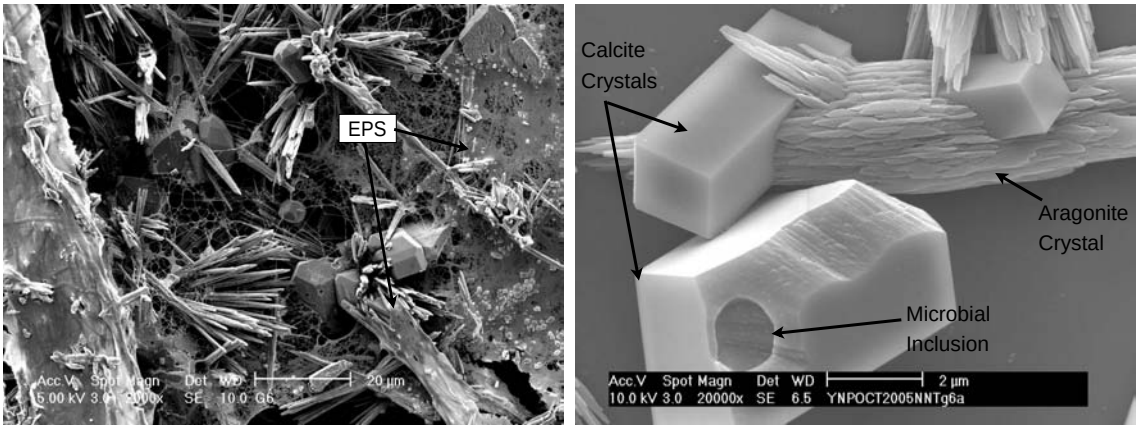

Figure 5.9: Two images of precipitated carbonate. The left picture, at 2000x magnification, shows calcite crystals, aragonite needles, microbially-produced extracellular polysaccharide (EPS), as well as some barely visible microbial inclusions. The right picture shows 20000x magnification, clearly showing the two kinds of crystals as well as a $1\mu m$ size hole which may have been formed by a single cell bacterium. Photos by Roy Johnson.

ibrated before, during, and after each transect using Fisher pH buffer solutions (4.0, 7.0, and 10.0) with an accuracy of ± 0.01 pH at 25^oC. pH measurements were complicated because of the rapid deposition of $CaCO_3$ on the probe's electrode, which causes instrument drift and slow convergence to a steady measurement. Carbonate buildup was assessed prior to each measurement, based on both convergence and a visual check for precipitation. Upon any signs of $CaCO_3$, the probe was cleaned by soaking it in a $0.1M\ HCl$ solution and then re-calibrated.

Temperature measurements were taken at the same times and locations as pH measurements, using the same probes. At Spring AT-3, temperature was simultaneously collected in situ at two locations (in a pond and in the proximal slope) using HOBO Temperature Data loggers, Model H20-001, which recorded a temperature measurement every 30 seconds.

Total flux in spring AT-3 was measured at the vent source using a propeller-based current meter, USGS Pygmy Meter Model 6205, and by measuring the area through which current was flowing. On the January 2005 trip (Table 5.1), we obtained an

Compilation of previous studies of Angel Terrace Hot Springs

Year	Spring	T/pH	Flux	Solid Chemistry	Aqueous Chemistry	Isotopic Analysis	Precipitation Collection	Microbial Samples
2005	AT-3	X	X	X	X	X	X	
2004	AT-3	X	X	X	X	X	X	X
2003	AT-3	X					X	X
2002	AT-1	X					X	X
1999	AT-1	X		X	X	X	X	X
1998	AT-1	X		X	X	X	X	X
Samples Collected	6	343	2	90	48	71	850	64

Table 5.1: A compilation of previous studies by our group at springs AT-1 and AT-3 during a 7-year period from 1998—2005. Each 'x' represents a sample or suite of samples collected during field activities for each of the parameters listed at the top of each column. The last row details the total number of collected samples or measurements of each parameter. A triplicate measurement is reported here as a single sample.

independent measure of flux using time of flight measurements inside a channel which had a fixed cross sectional area. This technique records the length of time which small, floating travertine flakes needed to travel a given distance. The pygmy meter was also used to characterize typical flow velocities in the spring system and, where possible, Pitot tubes were used to validate these measurements.

5.8 Definition of the primary flow path

Understanding which specific processes control deposition in each facies or making detailed comparisons between springs requires a more spatially-refined framework than the facies model. To connect macroscopic processes to those which control precipitation at a microscopic scale, we must follow the evolution of spring water as it progresses

along a single *flow path*. A flow path is the set of points traversed by a packet of water as it moves from the vent to the distal slope. Organizing water chemistry samples according to both the flow path and facies from which they are collected permits a consistent comparison between springs and thereby affords insight into travertine precipitation.

Unlike a stream, where it is easy to define a flow path, water at a hot spring spreads out over a broad area; streams divide and recombine. The water at a given point can depend in a complicated way on any number of upstream points, each of which could have different physical and chemical properties. But without knowing how two aqueous samples are related, it is impossible to know which differences are due to chemical, physical, and biological changes in the water and which are due to the mixing of different flow paths (e.g., the outflow from several ponds can mix as it enters the proximal slope).

This problem can be resolved by defining the *primary flow path*. Given a contiguous area covered by spring water, the primary flow path is the set of points at a given distance from the vent which are traversed by the largest volume of water. Where a large percentage of the spring water follows along a single trajectory, the chemical and physical properties at a point are overwhelmingly related to those of its upstream neighbor. This idea can be extended, and secondary or tertiary flow paths can be defined in regions of the spring which are disconnected by dry areas. This definition does not uniquely define a flow path in the sense of a streamline. When water spreads out evenly across an area, the primary flow path defined here can include a set of equivalent points.

While the primary flow path can sometimes be identified by visual inspection, this is not always the case, particularly in thin sheet flow further away from the vent. In this situation the primary flow path can be identified by considering temperature. It locally follows the trajectory along which temperature decreases most slowly as a function of

distance.

In most carbonate hot springs there is only a single vent, and the temperature of spring water monotonically decreases after it leaves the vent. Until it equilibrates with the atmosphere, the temperature of a packet of water is only a function of the length of time since it emerged from the vent. If T denotes temperature and t denotes time, then $T = f(t)$. Because $f(t)$ is monotonically increasing, we can write $t = f^{-1}(T)$.

Consider the contour of all points which are the same distance, $|\Delta r|$, from the last identified point on the flow path, which we denote $\vec{r}_o$. If we are looking nearby, $|\Delta r|$ must be small, and all these points will have the same water depth, H. In most of our system, this is a good approximation even for non-infinitesimal $|\Delta r|$. If $v(r)$ is the local velocity, and $\Delta\vec{s}$ a small displacement along the equidistant contour, then flux $Q(\vec{r}_0 + \Delta\vec{r})$ at each of these points may be estimated as follows:

$$Q(\vec{r}_0 + \Delta\vec{r}) \propto H\Delta\vec{s}v(\vec{r}_0 + \Delta\vec{r}) \approx \frac{\Delta\vec{r}}{\Delta t} \tag{5.7}$$

Here Δt denotes the time it takes a packet of water to go from $\vec{r}_0$ to $\vec{r}_0 + \Delta\vec{r}$. Because the points being considered are equidistant, $\Delta\vec{r}$ is the same, and $Q(\vec{r}_0 + \Delta\vec{r}) \propto 1/\Delta t$. Because $|\Delta r|$ is sufficiently small, Δt must also be small, and by expanding $f^{-1}(T+\Delta T)$ for small values of ΔT, we can write $\Delta t \propto \Delta T$. This implies that:

$$Q(\vec{r}_0 + \Delta\vec{r}) \propto \frac{1}{\Delta T} \tag{5.8}$$

Hence $Q(\vec{r}_0 + \Delta\vec{r})$ will be maximized where ΔT is minimized. The next point on the primary flow path will be the spot among these equidistant points with the smallest ΔT, or equivalently the spot with the highest temperature.

Thus we have shown that the primary flow path *locally* follows the trajectory along which temperature decreases most slowly as a function of distance. This theoretical discussion was applied to the data collected at spring AT-3 in 2004, and the primary flow path was determined by looking at the average temperatures at the 24 sample locations. The results in Figure 5.10 show that the points that comprise the primary

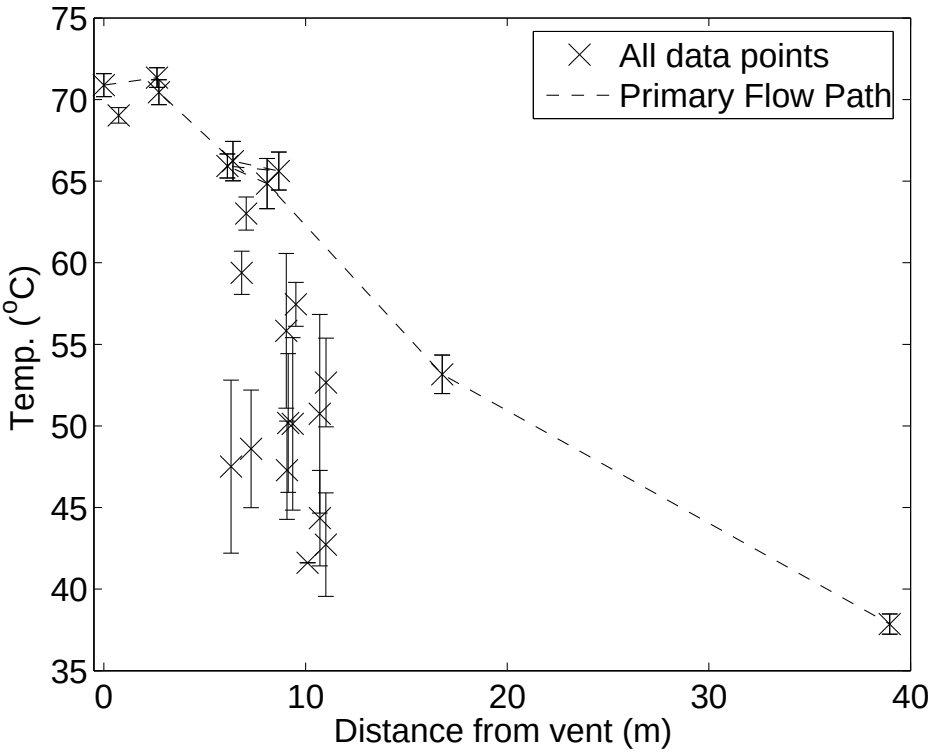

Figure 5.10: Average temperature values vs. distance for Spring AT-3 during 2004. The primary flow path is the line which locally minimized $\mathrm{d}T/\mathrm{d}r$.

flow path minimize $\mathrm{d}T/\mathrm{d}r$ as we step from point to point. The primary flow path locally minimizes the slope of the dotted line in Figure 5.10.

This analysis provides a useful recipe for analyzing aqueous measurements and for organizing experiments at hot springs. If one proceeds downstream with a meter stick and a thermometer, and draws arcs with the meter stick, the next point in the flow path will be the point along the arc with the highest temperature. When combined with standard qualitative observations, this approach allows sampling strategies which account for mixed flow paths, regardless of variations in water depth, velocity, or changes in underlying topography.

Figure 5.11 displays the results of the flow path analysis in two dimensions. The overhead view also shows a group of cooler points, which lie in the pond facies, off to one side of the primary flowpath. This is consistent with ponds forming in regions of lower

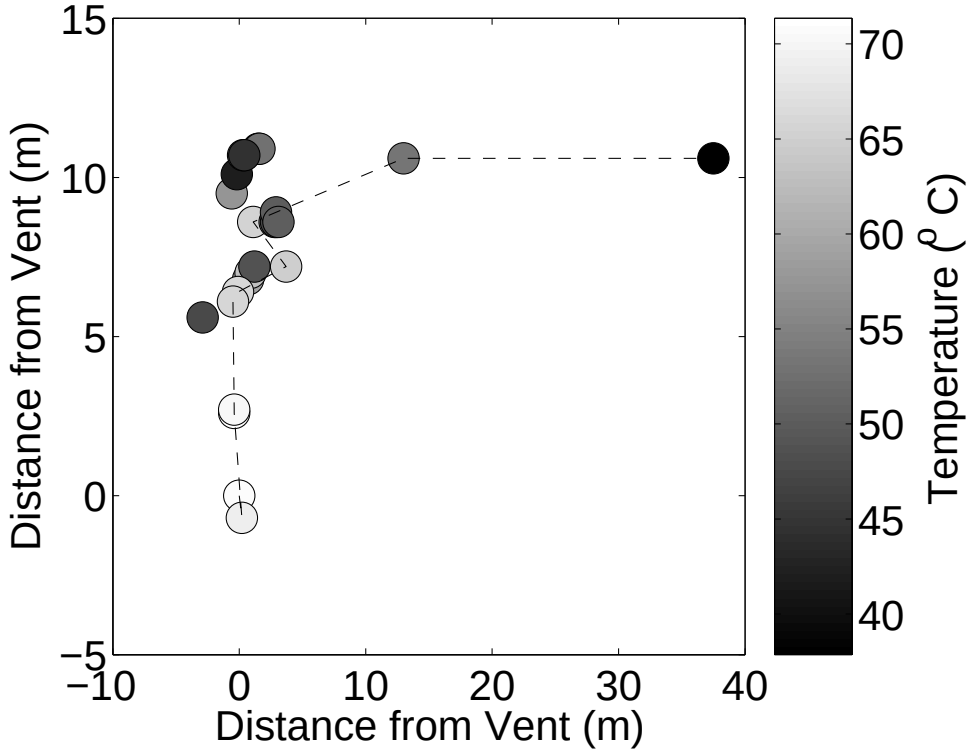

Figure 5.11: This plot shows the overhead view of the sample points from AT-3 in 2004. The primary flow path appears as a dotted line.

flux. This differs from AT-1, where the pond facies lay in between the apron channel and proximal slope facies[103], and explicitly demonstrates that the depositional facies are not strictly tied to the flow path.

5.9 Conclusions

This chapter has introduced our multidisciplinary collaborative work at Yellowstone National Park. We have explained why this system is interesting from the perspective of geologists, microbiologists, and physicists, and have posed our two fundamental questions: Why do terraces form at carbonate hot springs, and do microbes play a role? We have also broadly outlined the seven field expeditions during which we have collected experimental data.

To answer these questions, and to organize our data, we have introduced the conceptual tools needed to make sense of this complex and evolving system. Specifically, we explained the facies model, which enables us to understand the modern spring on the basis of the sedimentary rocks which are precipitating. This model also allows us to connect our work to other hot springs, to ancient rocks, and possibly to rocks from other planets.

To understand the process which give rise to these depositional facies we have introduced and defined a hierarchy of length scales, ranging from the microscopic to systems level. Although different processes can influence and control pattern formation at different length scales, these chapters will demonstrate that mesoscopic, macroscopic, and systems-level patterns (specifically terraces) are insensitive to the microscopic details, and can be described by minimal models. This approach is motivated by the apparent self-similarity of carbonate terraces at these larger length scales.

Finally, to connect microscopic processes to these larger patterns, we have defined and justified the primary flow path. Chapter 7 will use this tool to examine in detail data from AT-3, with the goal of identifying the processes which control macroscopic carbonate deposition.

Chapter 6

Toward a Predictive Model of Hot Spring Water from Modern and Ancient Travertine Depositional Facies

As mentioned in the previous chapter, we have collected an extensive data set of the physical, chemical, and microbiological attributes of the AT-1 and AT-3 hotsprings. In this chapter, we analyze this data, reporting a strong macroscopic correlation between travertine depositional facies and the microbial community structure, temperature, pH, and flux of hot spring water. This provides a basis for the use of facies models to reconstruct modern and ancient carbonate aqueous environments. Focusing on spring water temperature and pH, we quantify variability between and within springs. We also identify and explain two specific trends in microbial community diversity.

6.1 Introduction

In our investigation into the interaction between fluid and travertine precipitation, we implicitly assume that **specific** characteristics of the spring water control the shape and aggregate form of the travertine that precipitates. To dispense with the notion that this is obviously true, and implied by causality, consider two examples.

First, do the specific microbes present at the time of deposition, actually play an individualized role? Do the streamer fabrics shown in Figure 5.8 only form because of *Aquificales* bacteria, or will the presence of *any* bacteria suffice (e.g. to act as a surface for nucleation)? Do bacteria only influence precipitation rates (and not morphology), or do bacteria not matter at all? It might be reasonable to suppose that any bacteria present in the apron channel would be swept downstream, and that all bacteria would be found essentially everywhere in the hot spring, and therefore could not have anything to do with the differences between depositional facies. This turns out not to be true!

The same considerations apply to spring water chemistry. While there are obviously chemical differences at the *systems level* — the vent water is charged with CO_2, and the distal slope water is not — it is possible that the pond and distal slope water could arise from essentially the same spring water chemistry. Because the depositional facies often exhibit sharp boundaries, and most physical and chemical attributes of the spring water change smoothly and continuously over the course of the drainage outflow path, one might expect a loose relationship between aqueous chemistry and travertine facies.

The purpose of this chapter is to demonstrate that, on the macroscopic level, there are significant correlations between microbial communities, aqueous chemistry, and the underlying depositional facies. This implies that we can infer these properties of the depositional environment at carbonate hot springs simply by examining the travertine. To understand why this represents an important contribution to sedimentary geology, we now explain the history and usage of the facies concept in greater depth than before.

The concept of sedimentary depositional facies serves as a fundamental means for

reconstructing aqueous paleoenvironments from the geological record[138, 141]. However this basic facies concept is broadly defined, contains many assumptions about the physics, chemistry, and biology of the environment, and spans multiple temporal and spatial scales. The classic definition of facies is based solely on specific characteristics of a body of sedimentary rock[139, 140]. This original definition has been broadened and is now commonly used to describe the environmental processes *believed* to have formed the sedimentary deposit[142]. However, this secondary, interpretive sense of the word facies implicitly assumes knowledge of specific environmental conditions at the time of deposition. Yet despite the fundamental role of these assumptions, few studies have quantitatively examined the aqueous characteristics of modern depositional environments within a classic facies context. In this chapter, we approach this problem by studying terrestrial hot springs, where deposition is so rapid that the connection between the modern facies and the aqueous environment can be readily made.

Most previous studies have included a combination of field analyses of spring water chemistry, calcium carbonate geochemistry[100, 104, 105, 106, 133, 134, 143, 144, 145], and controlled lab experiments that quantify precipitation dynamics[127, 128, 132, 146, 147, 148, 149, 150]. However, these investigations have generally not simultaneously studied chemistry of spring water, microbial communities and solid $CaCO_3$ in a natural environment. Therefore, the results of these studies are not adequate for comparing $CaCO_3$ precipitated in different hot springs (travertine) or reconstructing paleoenvironments from ancient travertine.

In order to justify the environmental assumptions implicit in carbonate facies models, our research at Mammoth Hot Springs has analyzed the depositional environment within the context of a five-component facies model[103] (Section 5.5). In keeping with the original geological definition of facies, this model is based solely on the shape, structure, and chemistry of carbonate deposits on the floor of the spring drainage system[103].

The first part of this chapter examines macroscopic correlations between microbial communities and the underlying facies[99]. Culture-independent analyses of 16S rRNA gene sequences with universally conserved bacterial primers has identified over 553 unique partial and 104 complete gene sequences (derived from more than 14,000 clones) affiliated with 221 unique "species" that represent 21 bacterial divisions. We use the term "species" loosely here, but will define it precisely and operationally in section 6.2.2; whether or not the species concept is strictly applicable to microbes is questionable, but outside the scope of this thesis. These species exhibit a remarkable 87% partitioning between facies, implying that relatively little downstream bacterial transport and colonization take place despite the rapid and continuous flow of spring water from the high-temperature to low-temperature facies[18, 99]. Microbial fatty acids are also partitioned into facies and indicate that chemolithoautotrophic cycling by Aquificales bacteria occurs in the apron channel facies[151]. These analyses of the phylogenetic and metabolic diversity demonstrate that the facies model can accurately reproduce the macroscopic microbiological attributes of the depositional environment.

Previous studies of aqueous and isotopic chemistry have suggested that CO_2 degassing and spring water cooling are the dominant controls on systems-level changes in aqueous chemistry[103, 152]. The second part of this chapter examines the chemical and physical properties of the spring water at finer spatial resolution, the macroscopic level. Because $CaCO_3$ mineral precipitation occurs on time scales of days to months, the rock record inherently averages out more rapid fluctuations in the aqueous environment. By quantifying these fluctuations, we show that spring water pH, temperature, and flux are strongly correlated with the underlying facies on the macroscopic scale. We use these results to explain two trends in microbial diversity and community structure[18, 99].

Active $CaCO_3$ precipitation connects to bulk water chemistry through a microscopic layer at the interface between the flowing springwater and accreting travertine.

Although this process depends in detail on substrate mineralogy, crystal growth form, and microbial activity, we show that macroscopic water parameters (pH in particular) are not sensitive to this microscopic complexity. This result rigorously justifies using facies models to characterize depositional environments, and demonstrates that they may be used to reconstruct the macroscopic chemical and physical properties of modern and ancient aqueous environments.

The work reported here documents this macroscopic association, showing that in the modern environment there are significant and reproducible correlations between hot spring water and depositional facies. Other members of our collaboration will address in detail how microscopic aqueous chemistry and the kinetics of travertine nucleation and crystallization give rise to the correlations documented here. In this sense, my work is a necessary first step toward a fundamental predictive model of travertine precipitation in hot springs.

6.2 Brief introduction to microbial work at YNP

It is possible that common earth-surface geological features can arise as a result of microbes interacting with purely physical and chemical processes. The ability to distinguish ancient and modern mineral deposits that are biologically influenced from those that are purely abiotic in origin will advance our ability to interpret microbial evolution from the ancient rock record on Earth and potentially other planets. As a step toward deciphering biotic from abiotic processes, we have combined carbonate mineralogical and geochemical analyses together with community-based microbial genetic analyses in hot spring drainage systems at Mammoth Hot Springs in Yellowstone National Park. The unexpectedly sharp correlation between microbial phylogenetic diversity and travertine facies suggests that changes in bacterial community composition are a sensitive indicator of environmental conditions along the spring outflow. These

results provide an environmental context for constraining abiotic and biotic theories for the origin of distinct crystalline structures and chemistries formed during hot spring travertine precipitation.

In this section, we spatially map the ecological distribution of bacteria with respect to the travertine facies in which they live, using what have now become standard culture-independent molecular analyses of 16S rRNA gene sequences[18]. We describe a rigorous statistical evaluation of these bacterial 16S rRNA gene sequences, which is the first such analysis completed within a natural environmental setting. The results here provide the first quantitative estimation of microbial diversity and abundance in these hot springs. They also present evidence of significant correlations between microbial communities and travertine facies, which is the critical ecological relationship required to begin to identify the dynamics of microbial influence on carbonate mineral precipitation. These statistical approaches are universally applicable to other molecular studies of microbial ecology and will thus will be a valuable tool for studying biocomplex dynamics in the environment.

The vast majority of this work, including sampling, DNA extraction, PCR amplification, and sequencing was done by workers in the Fouke lab, especially George Bonheyo, Beth Sanzenbacher, and Jorge Frias-Lopez. George Bonheyo is also responsible for the laborious process of sequence alignment and comparison. My experimental work in this regard has consisted of technical help in taking samples at YNP during the 2002 and 2003 trips. I have, however, contributed substantially to the analysis of resulting sequence data. Specifically, I collated and organized several hundred different sequences, and then examined the relationship between these sequences and the facies from which they were sampled. Héctor García Martín, a former physics student in Nigel Goldenfeld's group, helped with this study of microbial partitioning, and is solely responsible for using accumulation curves to estimate the total microbial diversity and abundance. My work on microbial partitioning is meaningless without his rigorous

statistical justification.

Some of the the work in this section has been previously published[18], and the rest has been submitted for publication[99].

6.2.1 The 16s rRNA gene method

Identification of microorganisms in the YNP hot springs was done through the 16s rRNA gene method. This method relies on classifying microorganisms by their 16s rRNA gene, a gene which is present in a highly conserved form in virtually all cells. Although the coding part of this gene is unchanged between organisms, the non-coding ("junk") part changes as a result of random mutations[153]. The differences between these non-coding parts can be used to classify microorganisms. This was the insight that Carl Woese and coworkers used to create the phylogenetic tree of life on earth, leading to their discovery of *archaea*[154, 155].

To identify microorganisms via this method, we first take an environmental sample either by filtering the hot spring water, the "microbial mat" which sometimes grows on travertine, or even the steam which rises from the spring. We note the facies from which the sample came. We then preserve this sample for transport, usually by refrigeration, freezing, and $EtOH$.[1] Our work is based on multiple samples from each facies and computer-intensive resampling techniques that allow estimation of abundance distributions and even dominant microbial metabolism[156].

In the laboratory, the sampled cells are lysed using several different techniques, e.g., physically by "bead beating", chemically, or thermally. The resulting solution is purified to extract DNA. The 16s rRNA portion is exponentially amplified via PCR[153]. The resulting product is purified and then ligated into plasmids, which are then introduced into *E. coli* cells. These cells are isolated and grown in a culture, with care being taken to eliminate any which did not take up any of the 16s rRNA containing

[1] Different preservation techniques work for different microorganisms.

plasmids.

The resulting "clones" are screened using RFLP (Restriction Fragment Length Polymorphism)[153], which results in unique clones. We have the 16s rRNA fragments in these clones sequenced, and submit the result to GenBank[18], which allows identification of previous detections and (if they are known) the associated metabolisms.

6.2.2 What is an OTU?

One of the initially confusing aspects of microbial ecology, particularly when based on differences in 16s rRNA, is that the idea of a "species" quickly becomes meaningless and hopelessly obsolete. First, the 16s rRNA gene reflects only a portion of the total genome. There are examples of different species with very similar 16s rRNA sequences[157]. Secondly, measurements of genomic differences are continuous, while the definition of a species is discrete.

These problems are resolved by replacing the idea of a species with the *OTU*, or operational taxonomical unit. We define OTUs based on the percentage difference between 16s rRNA genes, a definition which does not consider any other attributes of the microorganisms. For instance, if we choose a 3% difference in 16s rRNA as our OTU definition, then two microorganisms which are (e.g.) 96% alike are considered to be different OTUs. 3% is a commonly used definition[157], but this choice is somewhat arbitrary. For our study, we considered three definitions: 3%, 1%, and 0.5%. We found that our conclusions are insensitive to this choice.

6.3 Microbial Sampling at YNP

We collected multiple samples from within each of the five travertine facies at Spring AT-1 for the purpose of conducting the first direct correlation of bacterial 16S rRNA gene sequence identifications with travertine mineral precipitation in the context of

sedimentary depositional facies[18]. Samples were collected from the interior of each of the five facies, with each sample occurring within the primary flow path. Each sample was collected from the middle of each facies and was thus laterally separated from the next sample by as much as a few meters.

The DNA extraction protocols and 16S rRNA gene sequence PCR amplification protocols employed have been optimized to avoid biases and are described in detail in the references[18, 158]. DNA amplified using universal bacterial primers was then cloned in order to isolate the individual 16S rRNA gene sequences. To maximize the number of unique sequences identified (thus better characterizing the total diversity of the spring system) we chose to avoid sequencing identical clones derived from a single PCR reaction. Of the greater than 14,000 clones generated, approximately 5,000 clones were screened by RFLP analyses and 1,050 potentially unique clones were selected for sequencing. Ultimately, 657 partial 16S rRNA gene sequences were obtained, and 108 of these were sequenced as contigs to completion. The GenBank accession numbers for the 16S rRNA gene sequences analyzed in this study have previously been reported[18]. Although archaea are known to be present in the YNP hot springs, we have not yet performed the corresponding analyses for archaeal primers due to personnel and budget constraints.

6.4 Results in Microbial Ecology

6.4.1 Statistical analyses

We analyzed our sequences using three Operational Taxonomic Unit (OTU) definitions, defined by sequence differences of 0.5%, 1%, and 3%, to determine whether our interpretations of environmental partitioning could be affected by such variation. The lower bound is due to our PCR and sequence derived error rate[159, 160] and the 3% difference is a typical OTU definition[161].

To quantitatively estimate how well each facies has been sampled, accumulation curves were fitted to analytical curves obtained by modeling the sampling process. This analysis is due to Héctor García Martín. We assume that in each environmental sample collected, there is a maximum of N possible bacterial cells that could be detected, and that each of these cells would be present and detected in the sample with a probability p, regardless of the cell's identity. The factor p includes the combined probability of the cell being captured and detected through the process of DNA extraction and amplification of the 16S rRNA gene sequences via PCR. Thus, we use multiple methods of DNA extraction to eliminate cell durability biases and amplify the 16S rRNA gene via PCR. Finally, we screen the resultant clone library in an attempt to sequence only unique clones within that sample, as opposed to repeatedly sequencing identical clones. In this manner we increase the likelihood that an OTU will be detected even if it is not numerically dominant in the clone library (which may be due to extraction, amplification, and cloning biases rather than environmental population abundance).

The likelihood that each sequence we analyze will represent a new OTU is approximated as $(1 - S/S_o)$, where S is the number of different OTUs already identified and S_o is the total number of different OTUs present in the environment. For each sequence, the probability that the number of different OTUs will increase is $p(1-S/S_o)$. This leads to an accumulation curve of the type $S = S_m(1 - e^{-Kt})$, where t is the maximum number of individuals that would be found if $p = 1$ and K is a constant related to the sampling procedure. This is not quite what was represented in the accumulation curves, since we only have information about samples rather than individuals, as explained above. Nonetheless, the number of samples n is simply $n = t/N$, so $S = S_m(1 - e^{-Kn})$. The parameters K and S_m were determined from a linear fit of $\log_{10} dS/dn$ versus $-n$. Estimates through other methods were also attempted: fits to hyperbolic accumulation[162] curves were not convincing and non-parametric methods[163, 164] yielded variances that were too large to be trustworthy.

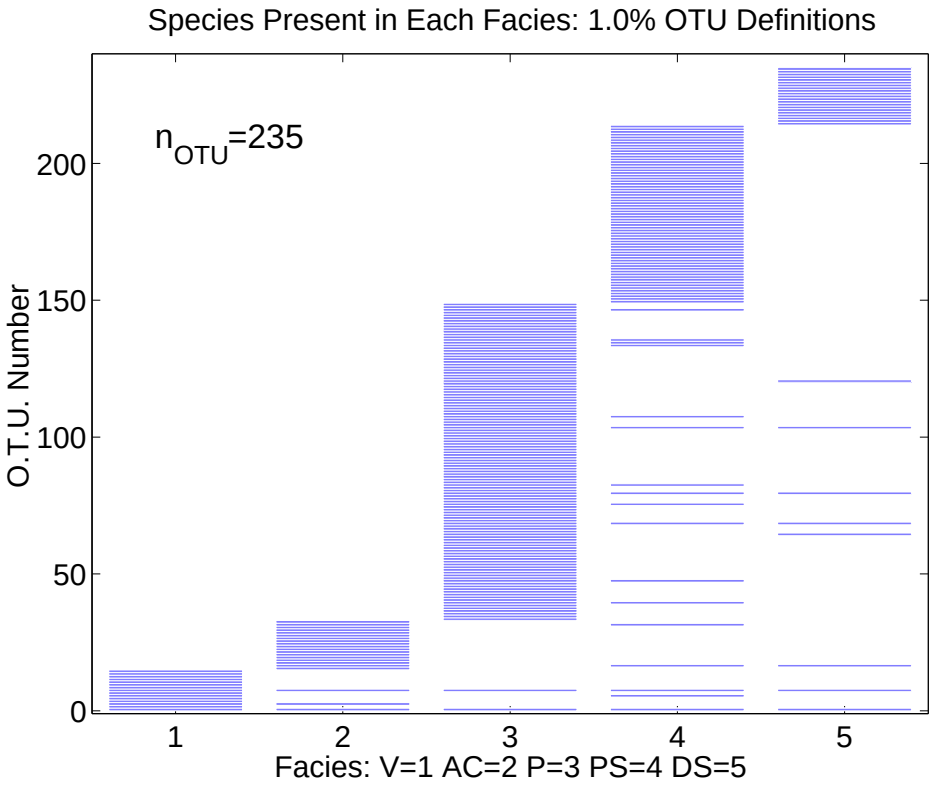

Figure 6.1: This plot provides a graphical representation of where each of 235 OTUs is found. It shows that microbial ecology is partitioned by facies. The X axis represents facies and the Y axis represents OTUs, with a bar appearing if an OTU is present.

6.4.2 Results in YNP microbial ecology

We identified 193 OTUs using the 3% cutoff and found that 90% of these could be found in only one of the facies (partitioned between facies). There were 237 OTUs using the 1% cutoff and 331 OTUs using the 0.5% cutoff with 91% and 93% (respectively) of the sequences partitioning to a single facies. Figure 6.1 graphically represents the distribution of sequences between the 5 facies using a 1% OTU definition. The plots for the 3% and 0.5% definitions are similar in appearance; however, under the 3% rule, 2 sequences may be found in all 5 facies.

The total number of sequences that can be found in more than one facies remains low under each OTU definition: 19 OTUs under the 3% definition, 20 OTUs under the 1% definition, and 24 OTUs under the 0.5% definition. Figure 6.2 summarizes the

number of overlapping OTUs in each facies as a function of the OTU definition. This demonstrates that the partitioning in Figure 6.2 is *not* an artifact of our particular definition of an OTU.

To validate our interpretation of facies partitioning, we need to determine what proportion of the total community in each facies we have identified. Severe undersampling might prevent us from identifying OTUs that actually do occur in multiple facies. Estimates for the total number of OTUs in each facies, made using an exponential fit to the accumulation curve, are shown in Figure 6.4. The accumulation curve plotted the number of different OTUs, S, found in a given number of samples versus this number of samples, n. Since all of the samples are assumed to be equivalent, this graph is an average over all possible permutations of these samples. Accumulation curves are traditionally made using the number of individuals instead of the number of samples[165]. However, our samples amalgamate large numbers of individuals: we have information regarding which OTUs are present in each sample, but not the OTU identity for every individual in the sample. The abundance of unique gene sequences in the clone libraries are not representative of the abundances in the environmental sample due to the inherent DNA extraction and PCR biases.

Accumulation curves like that shown in Fig. 6.3 were generated for the three different OTU definitions (3%, 1% and 0.5%) for each facies. The data from each OTU definition collapses into a single universal curve (e.g., the pond facies curves shown in Fig. 6.3), giving some confidence in the robustness of the sampling procedure and the validity of the assumption of random sampling used to derive the exponential accumulation curve. We see this pattern no matter which OTU definition is used. In the model above, all of the OTUs were assumed equally likely to appear (hence the factor $1 - S/S_o$). In a more realistic model the probability of finding each OTU should be proportional to its abundance. However, the approximations used above describe the data well and provide a tractable expression for the accumulation curve.

OTU overlap between facies for each OTU definition

0.5%	V	AC	P	PS	DS
V	**23**	2	3	5	2
AC		**24**	2	3	2
P			**167**	17	10
PS				**114**	8
DS					**40**

1.0%	V	AC	P	PS	DS
V	**32**	3	2	3	2
AC		**24**	2	4	3
P			**167**	14	7
PS				**114**	6
DS					**114**

3.0%	V	AC	P	PS	DS
V	**6**	3	2	2	2
AC		**20**	5	6	4
P			**99**	14	8
PS				**71**	6
DS					**28**

Figure 6.2: Number of OTUs shared between each facies, for OTU definitions 0.5%, 1% and 3%. Although the total number of OTUs increases with tighter overlap requirements, the number of OTUs observed in more than one facies increases only slightly. This indicates that the partitioning in Figure 6.1, is not an artifact of the OTU definition.

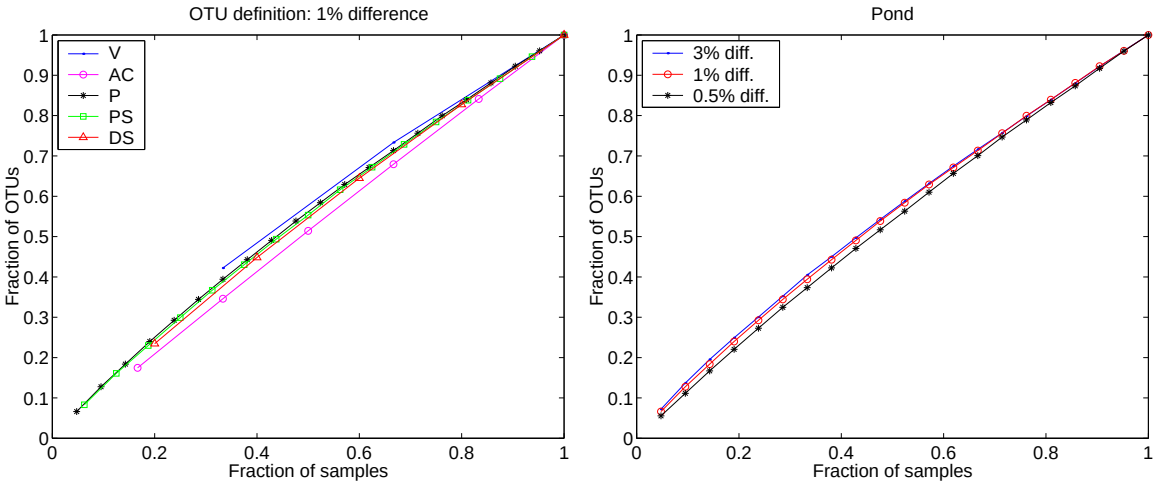

Figure 6.3: (left) Accumulation curves for the 1% OTU definition. Only the vent, proximal slope and pond facies show some curvature indicating that they are the better sampled than the rest. (right) Comparing pond accumulation curves for different OTU definitions. Although the curves become more linear with tighter definitions, but there are no substantial differences. The number of samples in each facies is: V: 3, AC: 4, P: 25, P: 14, DS: 6. Figure by Héctor García Martín.

For the 1% OTU definition, the result of fitting exponential accumulation curves for each facies is shown in Figure 6.4. This plot makes two key points. First, although we have sampled a significant fraction of the estimated total number of OTUs in each facies, we are a long way from sampling total biodiversity. This is the caveat which must kept in mind when interpreting Figure 6.1. Secondly, both our samples and total diversity estimates indicate that the *greatest* bacterial diversity is found in the pond facies, and that the *least* diversity is seen in the vent and distal slope facies. This result is somewhat surprising; while minimal diversity in the vent, with its high temperatures and low pH is not unexpected, the distal slope could be expected to have greater diversity because its relatively moderate physical and chemical environment can theoretically support a broad array of microbial metabolic activities (e.g., heterotrophy). This finding will be explained in the last part of this chapter.

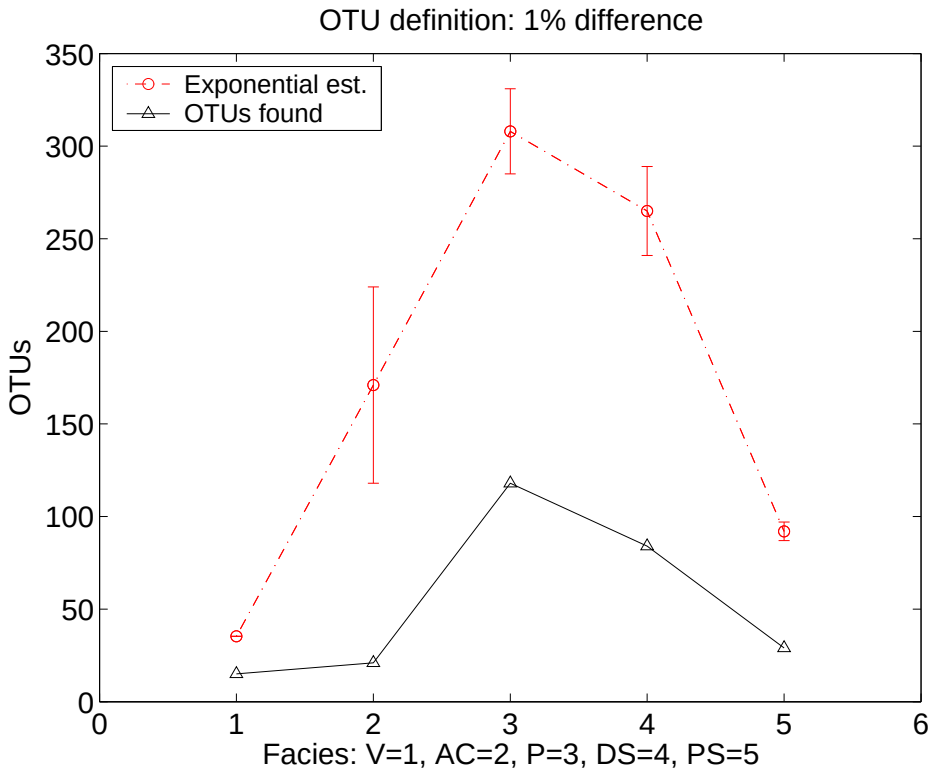

Figure 6.4: Comparing the observed number of OTUs in each facies with the exponential accumulation curve predictions for total facies diversity. Figure by Héctor García Martín.

6.4.3 Discussion

Although different microbial species have specific growth requirements and preferred temperature and pH ranges, the tight partitioning with respect to the travertine facies is nonetheless remarkable. First, it is surprising that very few of the upstream sequences were not also detected downstream. We initially expected that the rapid flow of the spring would result in downstream transport of microbial cells, and thus we thought that many sequences would also be identified downstream of their point of initial detection. Consequently, we performed most of our analyses on the first four facies extending from the vent. Surprisingly, the sequences detected in the water column of one facies, which are presumably most susceptible to being flushed downstream, were not typically detected downstream of their original facies, possibly due to low rela-

tive abundance or carbonate encrustation. Secondly, because bacterial species have a preferred range of environmental growth conditions, we expected that many sequences would be found across facies boundaries, coinciding with gradual temperature and pH changes. However, the facies boundaries proved to be nearly absolute boundaries with respect to detected bacterial 16S rRNA gene sequences. Although we observed particular sequences over a range of conditions within each travertine facies, with very few exceptions, OTUs were not found to traverse the facies boundaries.

Inferred metabolic activity of the identified bacteria, derived from comparison of our sequences to GenBank, indicates that the bacterial communities found in the spring drainage system change from primarily chemolithotrophic in the vent facies, to photoautotrophic and ultimately to heterotrophic in the distal slope facies. Associated with this transition is an observed increase in the total number of OTUs and their associated bacterial divisions from the vent to the pond facies. The number of OTUs decreases, however, with down flow progression into the proximal slope and distal slope facies. These trends in our data can be interpreted as follows: fewer OTUs and bacterial divisions would be expected at the upper temperature limits of the spring where little organic matter is available for heterotrophy and the temperature is at the upper limit for photosynthesis[166]. Although the pond through distal slope facies have temperature profiles that would support both autotrophic and heterotrophic lifestyles, we actually find a reduction in the number of OTUs represented in the proximal slope and distal slope. In the last part of this chapter, we demonstrate that this results from the spatial variations in spring water temperature and pH within each facies.

In conclusion, our sampled OTUs (and therefore most of the bacterial consortia) are macroscopically partitioned according to the travertine facies model; the microbial communities are highly correlated with the underlying depositional facies. This finding constrains abiotic theories for the origin of travertine terraces: either the origin is biotic, or else the microbial ecology is strongly coupled to the geochemistry through

other mechanisms.

6.5 Spring Water Chemistry

The simplest explanation for bacterial partitioning invokes springwater aqueous chemistry. If each facies is highly correlated with a unique chemistry, which in turn defines a specific bacterial habitat, then the correlations between microorganisms and depositional facies would follow transitively from the chemistry - facies correlations. This, in turn, would allow explanations of travertine deposition which did not depend on microbial activity.

We will now turn to our analysis of macroscopic spring water chemistry. Our results show that this explanation for microbial partitioning is plausible (e.g. that the microbial correlations follow from chemical ones). Moreover, we use our analysis of variability within each facies to explain why the curve in Figure 6.4 is peaked in the pond facies.

6.5.1 Variability and fluctuations in hot spring water

Table 6.1 compares the ranges of observed spring water temperatures and pH, as a function of facies, for multiple hot springs and times. There is consistent overlap between measurements from the same facies at different springs (AT1 and AT3), but there are also large variations within each facies. Although this variability does not obscure the overall down flow trends, identifying meaningful differences that exist between or within hot springs requires first quantifying the macroscopic fluctuations and variations that occur within a single spring.

Three kinds of variability are relevant to carbonate precipitation. First, there are temporal fluctuations on the time scale of our measurements (e.g. 10s of seconds). In our error model, these fluctuations are treated as measurement errors. Second, there

Comparative temperature and pH ranges for each facies and previous study

Year	Location	Value	Vent	Apron Channel	Pond	Proximal Slope	Distal Slope
2005	AT-3	T(^{o}C)	69.6	64.1	N/A	47.2	34.1
		pH	6.29	6.62		7.76	8.14
2004	AT-3	T(^{o}C)	68.0 - 72.2	60.5 - 68.6	35.6 - 61.7	50.6 - 56.4	34.2 - 39.4
		pH	6.21 - 6.57	6.59 - 7.26	6.84 - 8.04	6.99 - 7.77	7.32 - 8.07
2003	AT-3	T(^{o}C)	71.2	61.6 - 69.3	56.1 - 61.5	41.4 - 65.2	28.4 - 44.0
		pH	6.58 - 6.61	6.60 - 7.05	6.94 - 7.01	6.94 - 8.01	7.75 - 8.14
2002	AT-1	T(^{o}C)	67.9 - 69.3	60.5 - 64.0	59.9 - 60.3	46.7 - 50.8	24.0
		pH	6.59 - 6.76	7.00 - 7.26	7.29 - 7.30	7.95 - 8.05	8.12
1999	AT-1	T(^{o}C)	72.0 - 72.2	70.1 - 70.3	46.8 - 55.3	39.0 - 39.8	37.1 - 37.5
		pH	6.31 - 6.32	6.46 - 6.52	7.42 - 7.62	7.73 - 7.77	7.90 - 7.92
1998	AT-1	T(^{o}C)	73.2	N/A	4.53	54.2	30.2
		pH	6.00		7.43	7.40	8.00

Table 6.1: A compilation of previous studies by our group at springs AT-1 and AT-3 during a 7-year period from 1998—2005. Each 'x' represents a sample or suite of samples collected during field activities for each of the parameters listed at the top of each column. The last row details the total number of collected samples or measurements of each parameter. A triplicate measurement is reported here as a single sample.

are temporal fluctuations on the time scale of days, such as changes in spring discharge. These are relevant for comparing different measurements, but not for understanding travertine deposits thicker than a few millimeters. Finally, there are spatial variations on macroscopic scale. There are also spatial differences on both the microscopic and systems-level scale which are not relevant to the scope of this discussion.

Figure 6.5 shows these three kinds of variability for both temperature and pH as a function of facies. We quantified temporal fluctuations by considering the ensemble of measurements at a given point in space (taken over a period of three days), calculating the standard deviation of that ensemble, and then averaging those deviations over each facies. We quantified spatial variations by grouping all measurements collected within

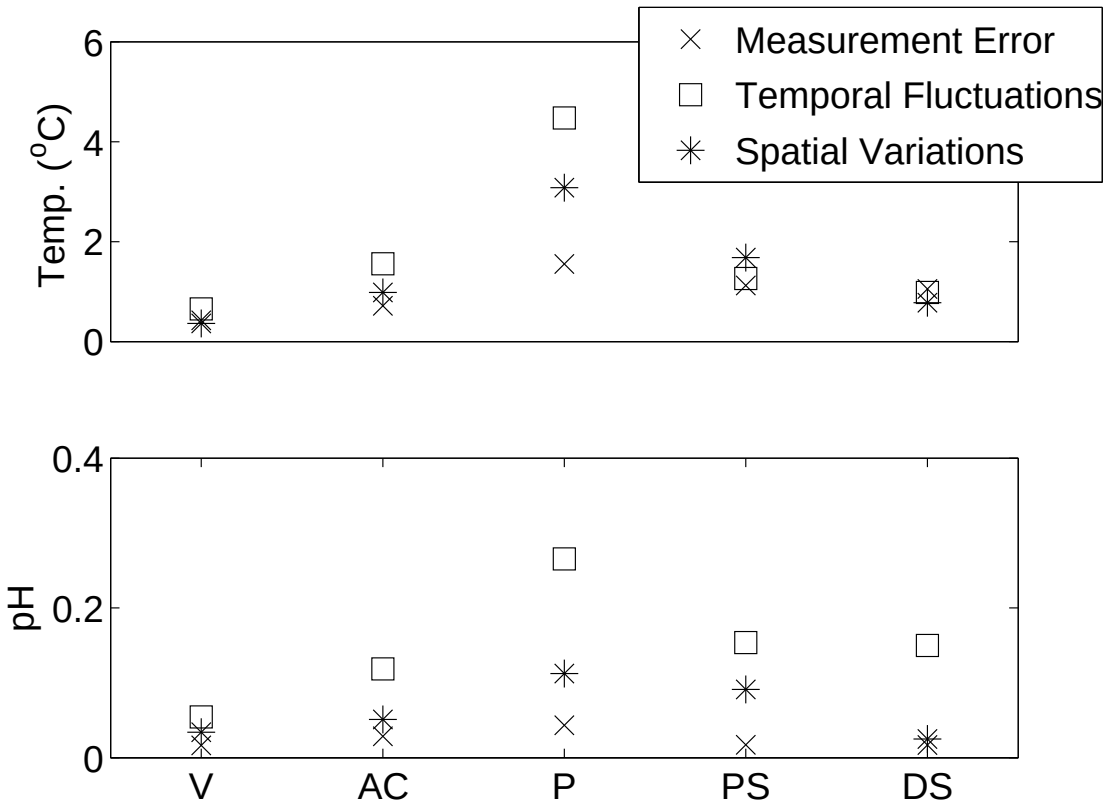

Figure 6.5: The three kinds of variability relevant to understanding travertine deposition in comparing hot springs, organized by facies. V—vent facies. AC—apron channel. P—pond. PS—proximal slope. DS—distal slope.

a facies at a given time, and then calculating the standard deviation of that ensemble. This was repeated for measurements taken at different times, and the results were averaged over each facies.

While the measurement errors are small (approximately 0.03 pH units and 1^oC), there are significant temporal fluctuations and spatial variations. These differences are not driven by large changes in the source water, as the vent exhibits the smallest changes. Both fluctuations and variations result from the interplay of myriad smaller factors that include: changes in the flow patterns upstream from a point; diurnal insolation[2]; changes in total spring flux; and atmospheric conditions, particularly wind.

[2] *Insolation* refers to solar flux per unit area of horizontal surface.

We observed wind driving water over the pond lips, dramatically changing the pH and temperature of downstream points, particularly in areas of low flux.

During a period of three days a HOBO temperature data logger in the proximal slope recorded a maximum change of 9^oC, with a standard deviation of 2^oC. These measurements exhibit a clear diurnal signal, driven by differences in daytime and nighttime air temperatures as large as 20^oC. The fluctuations seen in the HOBO data are consistent with the results shown in Figure 6.5.

Total spring flux was considerably more difficult to measure than either temperature or pH. In June 2004, using a pygmy current meter, we estimated Spring AT-3 discharge at 59.0 L/s. In January 2005, using time of flight techniques, we measured 12.0 L/s (see Appendix 6.8 for details). Both measurements are accurate to within 10%. These numbers indicate significant variation in total spring flux, which will result in changes in downstream aqueous chemistry, temperature, and flux — even if the source water remains otherwise unchanged. Our flux measurements do not agree with the report of Sorey[123], who reported the total discharge of *all* hot springs in the Mammoth complex as 59.0 L/s. We observed this much fluid coming out of a single spring. Soreys result was based on the assumption that approximately 10% of the total groundwater erupts through hot springs. This assumption seems to be incorrect, although the large variations in flux which we have observed would make it difficult to confidently estimate total discharge.

Figure 6.5 shows a noteworthy trend in spatial variability. The largest heterogeneities in temperature and pH are seen in the pond facies, and the most homogeneous regions are the vent and distal slope facies. This occurs because the spring system is essentially held fixed at the beginning and the end of the primary flow path. At the vent, pH and temperature are held constant by the steady influx of homogeneous source water, which has relatively constant temperature and chemistry (Table 6.1). In the distal slope, CO_2 fugacity asymptotically nears what would be expected for spring wa-

ter in equilibrium with both the atmosphere and solid $CaCO_3$ in the substrate. This largely determines pH and temperature in the distal slope. The temporal fluctuations in temperature mirror the trend seen in spatial variations.

Significantly, this trend in spatial variability could explain why the greatest bacterial diversity is seen in the pond facies, as shown in Figure 6.4[18, 99]. Our results show that the pond facies offers the broadest array of mesoscopic physiochemical conditions. As a result, it can support the greatest microbial diversity[153]. Conversely, the most homogeneous environments, at the beginning and end of the primary flow path, have been shown to support the least diverse bacterial communities[99].

6.5.2 Relating the rock record to the depositional environment

Despite large variations and fluctuations in the aqueous environment, we see statistically significant correlations between physical and chemical attributes of spring water and underlying depositional facies. These correlations exist because macroscopic $CaCO_3$ mineral precipitation occurs on time scales of days to months, and the rock record inherently averages out more rapid fluctuations in the aqueous environment.

Figure 6.6 shows the distribution of all pH and temperature measurements arranged by facies. It illustrates that the vent, apron channel, and distal slope facies can be identified by pH and temperature, implying that these facies are associated with distinct depositional environments. The transition from the vent to the apron channel facies is associated with the pH increasing beyond 6.6 while temperature is relatively unchanged. This suggests that this transition is controlled by CO_2 exsolution and the onset of carbonate precipitation.

The pond and proximal slope facies cannot be differentiated from each other on the basis of spring water temperature and pH. While petrography documents some similarity on a microscopic level[18, 103], clear distinctions (such as travertine dams

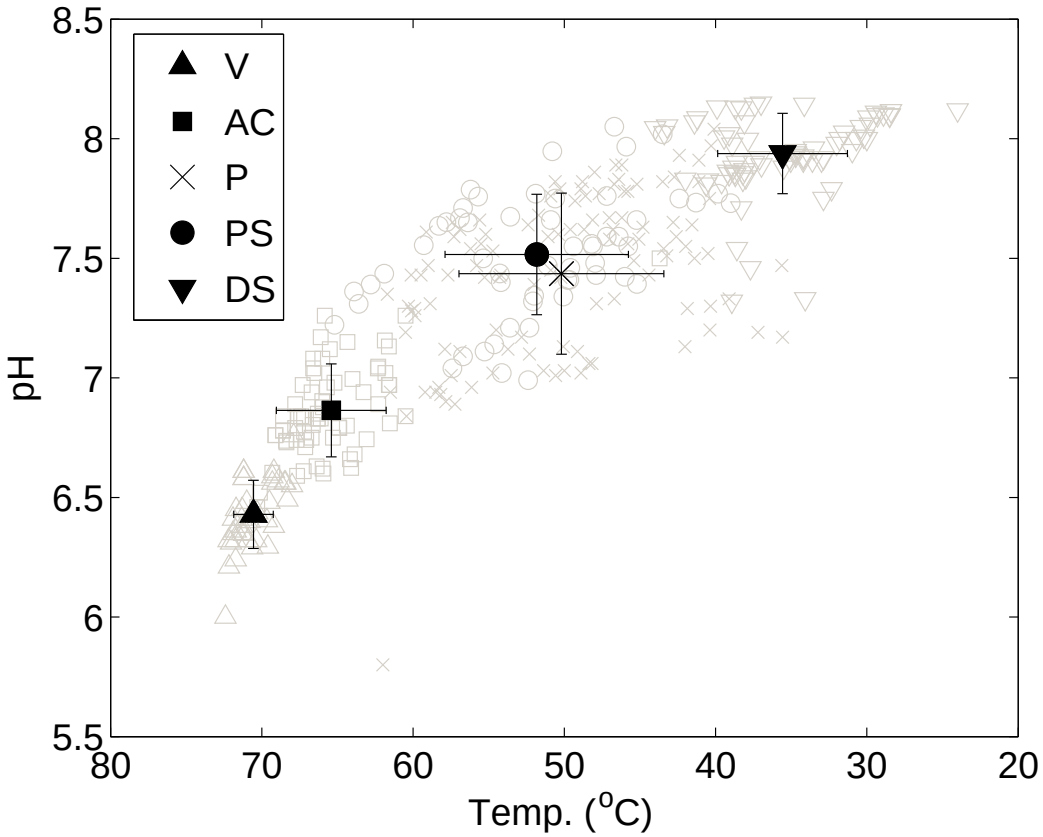

Figure 6.6: The distribution of 343 triplicate pH and temperature measurements taken from Springs AT-1 and AT-3 at different times. The black symbols show the facies averages, with the error bars denoting 1 standard deviation.

and a terraced architecture) emerge on a macroscopic level. We therefore considered additional physical parameters to understand how the same spring water can give rise to two distinct aggregate morphologies. These two facies are differentiated by average fluid velocity.

Velocities in these thin sheet flows are difficult to measure, particularly because techniques like particle image velocimetry could negatively impact the natural hot spring environment. We have used several independent techniques, including the pygmy current meter, Pitot tubes, and time of flight measurements. We observed velocities less than 20 cm/s in the pond, but over 35 cm/s in the proximal slope. The average velocities in these facies are a function of slope, total flux, fluid depth, and facies

area. Using our measurements of area, depth, and total flux, we calculate velocities consistent with these experimental findings.

These results imply that, if other factors remain unchanged, ponds will form in flatter areas with lower flux and the proximal slope facies will form on steeper underlying topographies with higher flux. This is consistent with qualitative observations at the springs AT-1 and AT-3, and also agrees with computational models of these hot spring systems[20].

Organizing the physical and chemical parameters of the hot springs according to the facies model leads to clear partitioning. These macroscopic partitions in spring water explain the previous observations (Fig. 6.1) which showed that microbial communities are also partitioned according to facies[18, 103]. Because different microorganisms thrive in different niches, microbial community structure must be correlated with the partitions seen in the aqueous environment.

6.6 Reconstructing Modern and ancient Environments

The strong correlations between travertine facies and the macroscopic physical and chemical parameters of the depositional environment allow each facies to be uniquely identified solely by pH, temperature, and flow velocity. While fluctuations and variations complicate comparisons between springs, we have shown that when viewed statistically (Fig. 6.6), measurements taken from the same facies in different springs are equivalent. When combined with previous work[18, 99, 103], these results demonstrate that all travertine depositing hot springs exhibit the same macroscopic partitions chemically, physically, petrographically, and microbially. The only other study which reports hot spring temperature and pH in the context of aggregate morphology[167] is consistent with the partitions shown in Figure 6.6.

This simple linkage is a powerful predictive tool, implying that the facies model, which inherently averages fluctuations and variations, can be directly used for paleoenvironmental reconstructions. Our results also put precise quantitative bounds on the temperature and pH of the spring water from which ancient travertine facies originally precipitated. As an example, Chafetz et al.[106, 107] documented streamers in Pleistocene travertine. Their description suggests that these deposits were formed in the apron channel facies. Utilizing Figure 6.5, we conclude that these deposits precipitated from spring waters with a pH of 6.75 ± 0.13 and a temperature of $65.5 \pm 3.7^{o}C$. Figure 6.1, allows us to predict the microorganisms which were present at the time these fabrics were deposited, and to do so with better than systems-level specificity.

6.7 Conclusions

During six trips spanning seven years, we collected a large, comprehensive set of measurements which characterize the spring water, travertine, and microbial ecology at two carbonate hot springs in Yellowstone National Park. Although the aqueous environment of these springs exhibits large spatial variations and temporal fluctuations on the macroscopic scale, we have shown that pH, temperature, and flux are sufficient to tie the spring water to the underlying depositional facies. The observed strong correlations between travertine facies and the spring water remove any assumptions about the aqueous environment during deposition. This validates the facies concept as a macroscopic framework for comparing modern hot springs. These results allow us to put more accurate quantitative bounds on paleoenvironmental reconstructions from ancient travertine.

We have quantified both fluctuations and spatial variations which, while large, do not preclude comparison of aqueous samples from different springs. When compared statistically, these samples demonstrate that there are no substantial differences be-

tween the springs on the macroscopic level. All travertine depositing hot springs exhibit the same partitions (Fig. 6.5). This implies that the distinct processes which control deposition in each facies are the same in all carbonate hot springs, and explains why the facies model has universal applicability to the solid phase, aqueous environment, and microbial communities.

The observed fluctuations and correlations in spring water temperature and pH explain our two primary results in microbial ecology. The greatest microbial diversity has been observed in the pond facies, and the least in the vent and distal slope. We quantitatively demonstrate that spring water in the pond facies has the largest spatial heterogeneities, whereas the vent and distal slope facies are the most homogeneous. We hypothesize that trends in microbial diversity are driven by the observed variations in the aqueous environment. Habitat considerations also explain why microbial communities exhibit partitions between facies[99]. Because microbial niches are defined by a unique set of physical and chemical parameters, microbial communities must follow the partitions seen in the aqueous environment.

6.8 Appendix A: Details of flux measurement

This appendix details how we calculated total spring flux using a pygmy current meter (USGS Pygmy Meter Model 6205). Using this meter, we measured total current through a surface that enveloped the vent aperture. We also measured the geometry needed to calculate the area of said surface. The geometry of our configuration is shown in Fig. 6.7. Measurements of flow velocity were taken at two different depths on each of the four faces of the cube shown. The area of the top surface was also measured. Using the area of each face through which current is flowing (nothing flows through the top or bottom surfaces), we calculated the total flux through the cube, which is equivalent to the total flux out of the vent.

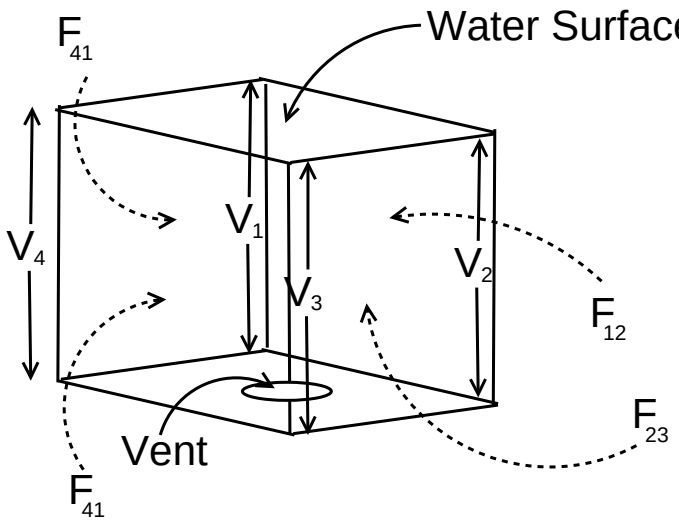

Figure 6.7: Schematic of pygmy based flux measurement.

The pygmy meter counts (via audible clicks monitored by headphone) the number of revolutions of the propeller. By measuring the time period, and using the meter's conversion table, we can calculate the flow velocity at each point. The conversion table specifies that, given R in revolutions per second, the velocity given in feet per second is obtained by: V=0.9604R+0.0312. These measurements, along with the depth of the fluid at the current measurement location, are given in Table 6.2. The other aspects of the geometry of the cube in Figure 6.7 are summarized in Table 6.3.

The flux through each face was estimated using a crude numerical integration. Each face was broken into two regions. The top one is a rectangular region that extends to 15.25 cm deep. The bottom region is a quadrilateral, the sides of which extend from 15.25 cm to the bottoms of the vertices, as reported in Table 6.3. Each of our measurements are paired for each face. The higher measurement (at a depth of 12.2 cm) was considered to be the flux across the entire top area of the face. The bottom measurement was similarly used for the bottom area. Multiplying each velocity times the appropriate area gives the flux through that area. Summing the flux for all of the areas gives the total flux of the spring: 59.47 l/s (59,470 cm^3/s).

Face	Revolutions	Time (sec)	Velocity (cm/s)	Depth (cm)
F12	40	60	20.5	12.2
F12	25.5	60	13.4	18.29
F23	31.5	60	16.32	12.2
F23	30	60	15.59	18.9
F43	29	60	15.1	12.2
F43	84.5	40	62.79	18.9
F41	92	40	59.5	12.2
F41	84	60	55.1	18.9

Table 6.2: Raw measurements involved in flux calculation.

Edge	Depth (cm)
V1	24.4
V2	19.8
V3	21.3
V4	24.4

Table 6.3: Edges of cube through which flux was calculated.

Chapter 7

Processes which control $CaCO_3$ precipitation

7.1 Introduction

The previous chapter demonstrated that there are macroscopic correlations between the depositional environment and precipitated travertine. However, explaining terrace formation, building minimal models, or proving that specific carbonate patterns result from microbial influence requires more than correlations: we need to establish causal connections between chemical changes or biological agents and the resulting precipitation. We wish to understand what processes control macroscopic pattern formation.

This chapter presents experiments and data analysis which shed light on the processes that control deposition. We do not address microscopic processes, where bacteria certainly impact precipitation (e.g. Figs. 5.9 and 5.8), and where chemistry is controlled by detailed nucleation kinetics, diffusion, and local concentration gradients. That is a subject for future work. Instead, we continue to work on a macroscopic scale, with the goal of understanding and modelling terrace formation. As mentioned before, we expect that the processes controlling the formation of these features, and consequently the terraces themselves, will be insensitive to the microscopic details.

To understand how macroscopic changes in chemistry impact deposition rates (and thereby travertine patterns), we use the primary flow path defined previously in Section 5.8 to analyze our experimental data. By tracking changes in bulk water chemistry along this flow path, we connect specific changes in aqueous chemistry with changes in both topography and measured precipitation rates. We find that rapid drops in height (e.g. a "waterfall") cause CO_2 degassing and increased precipitation rates.

To determine whether microbial metabolic activities have any *macroscopic* impact on travertine precipitation, we look for (but do not find) indications of photosynthesis during 24 hour periods. We also present a theoretical calculation which bounds the potential impact of microbial metabolisms on water chemistry. This analysis shows that any possible macroscopic impact of active microbial activies is negligibly small in comparison to inorganic chemical changes. This does *not* mean that microbes cannot play an important passive role (e.g., by serving as nucleation sites), or that they do not have important microscopic effects in localized regions where other sources of degassing are suppressed; for example, at the bottom of the deeper ponds, turbulent degassing may not be as important as it is at pond lips or waterfalls. But our findings imply that terrace formation does not require the presence of any specific bacteria

These experimental studies took place at spring AT-3 in 2003 and 2004 (Table 5.1). Before presenting the results, we explain our experimental procedure in detail. We do this because we are presenting additional chemistry, aspects of which are difficult to measure, and because our sample points (within the primary flow path) must be more precisely located than in the previous chapter.

7.2 Experimental Methodology at AT-3

Our most comprehensive set of measurements came during the June 2004 field trip, at spring AT-3, in which I participated, along with Nigel Goldenfeld, Tom Schickel, Bruce

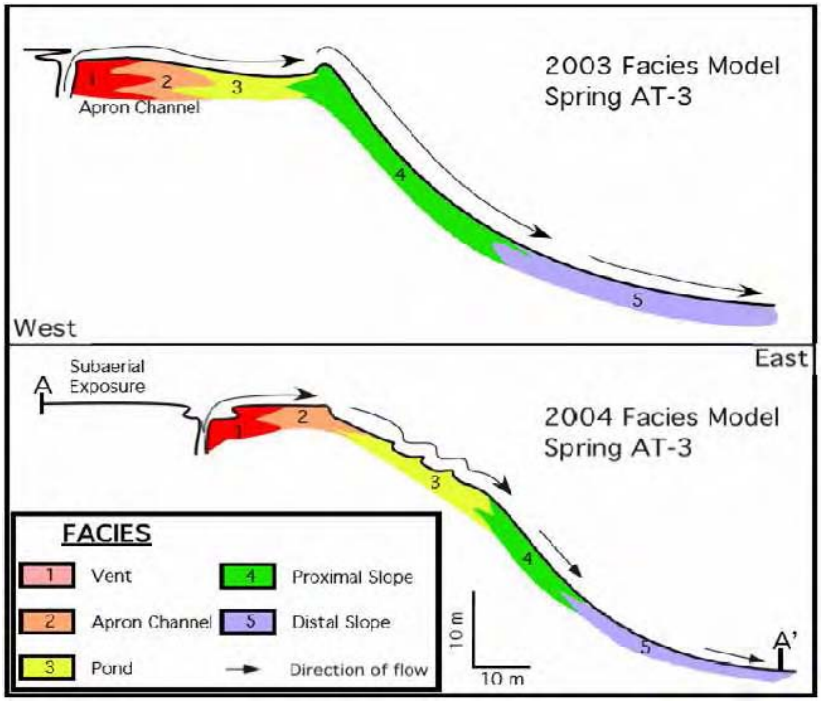

Figure 7.1: This schematic depicts cross-sectional profiles of spring AT-3 in 2003 and 2004, including facies locations. Between the two trips, the spring vent moved, resulting in changes in downstream patterns. Specifically, there are a number of small ponds developing, rather than one large one at the end of the apron channel. Figure by Tom Schickel.

Fouke, and Kelly Zimmerman. The previous summer, we had collected two spatially detailed transects, but had not collected a full suite of water chemistry (including DIC) or precipitation measurements. That field trip had been focused on observations of microbial ecology. The spring had changed substantially during the intervening year. Figure 7.1 depicts how the largest changes resulted from the the lateral movement of the spring vent. These changes offered us an opportunity to understand how the original topography at a hot spring influences initial precipitation patterns after a new vent opens.

During June 2004, we identified 25 different sample locations, distributed throughout the five facies. These were located with respect to the vent, and are depicted schematically in Figure 7.2. Between May 31 and June 3, 2004, we collected an enormous set of measurements at these data points. Operating continuously during the day

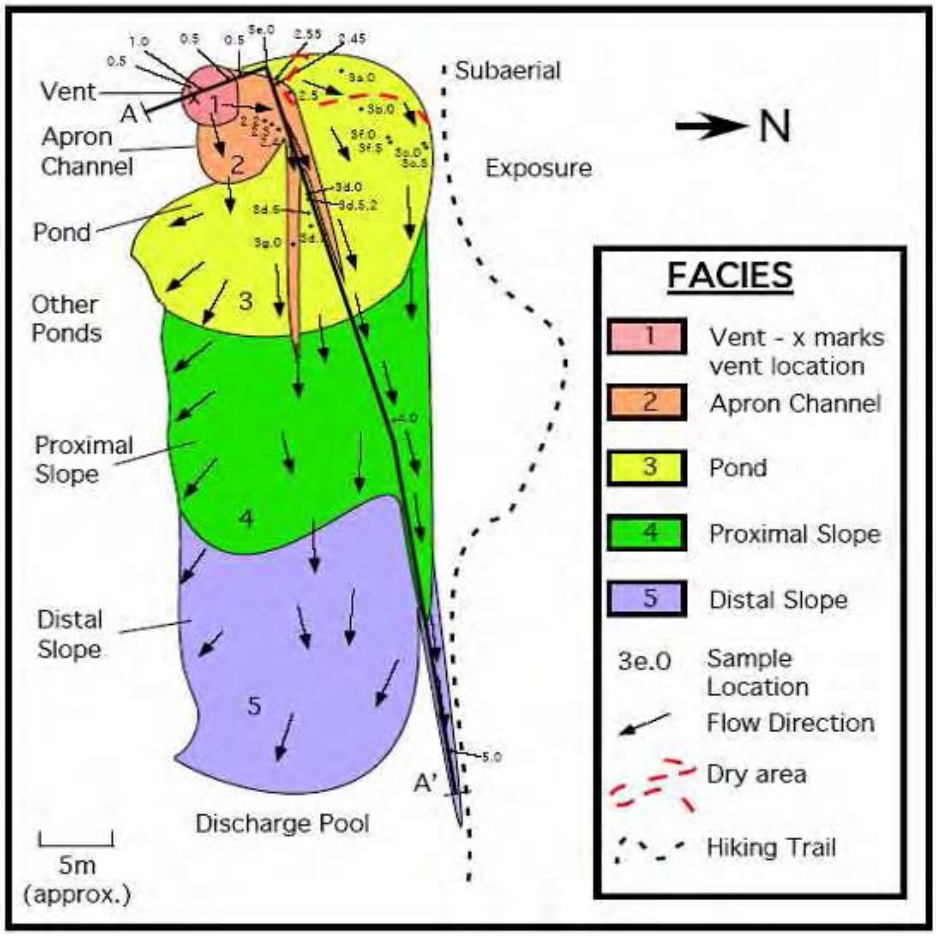

Figure 7.2: This schematic depicts sample locations at AT-3 in 2004, including facies, the primary flow path, a foot path, and an area of the spring which dried up during our experiments. Figure by Tom Schickel.

and the night (so as to look for changes due to the presence or absence of photosynthesis), we repeatedly measured precipitation, temperature, pH, and DIC at each of these points. At six of the points, we also collected water samples for chemical and isotopic analysis. Using two teams, with shifts of up to 36 hours, we intensively monitored all changes in spring chemistry and precipitation for three full day and night diurnal cycles.

To quantify precipitation, we placed over 750 stainless steel washers on the bottom of the spring, as shown in Figure 7.3. Washers were left in the spring water for 4, 12, and 48 hour periods, so that we could determine the extent of any initial transitory effects

associated with $CaCO_3$ nucleating on the steel washers (as opposed to nucleation on calcite or aragonite). These time periods also allowed us to look for diurnal differences in precipitation rates.

These washers had been carefully weighed beforehand, and were later transported, dried, and meticulously re-weighed, so as to determine the mass of precipitated travertine during a fixed time period. Combining this result with the usual mensuration formula for an annulus, we calculated the precipitation rate for each sample location. This latter painstaking set of measurements and subsequent analysis was performed by Tom Schickel. After being weighed, some of the "48 hour" washers were used for petrography, including both optical work and environmental scanning electron microscopy (ESEM).

The precipitation measurements were carefully tied to measurements of DIC, Ca^{2+}, temperature, pH, and total dissolved solids (TDS) at the same locations. As they can be relatively easily taken with a hand-held probe, we collected 2,104 measurements of pH, temperature, and TDS, by collecting transects approximately every 4 hours at each of our sample locations (Fig. 7.4) The first four transects were not completed in triplicate, but the last five were. At locations where the spring water was less than a few millimeters thick, it was collected *in situ* using a rinsed, empty 35 mm film canister. The probe was then inserted into the water which accumulated in the canister. The two HOBO data loggers mentioned in the previous chapter were also deployed at two sample locations in the

Figure 7.3: Photograph showing washers placed in the hot spring to measure precipitation. The color coded electrical tape denotes how long the washers were left in the spring (either 4, 12, or 48 hours).

spring. They provided a useful check on our probe-based measurements.

Six sets of DIC measurements were obtained using two kinds of titrations. In the first, we titrated (Fig. 7.5) using HCl, and in the second $NaOH$, with reagent concentrations selected according to facies. Strong reagents are required in the vent facies (which has high concentrations of DIC), and weaker ones in the distal slope (which has relatively little). Twenty five water samples were collected at six locations (0.5, 3e.0, 2.3, 3g.0, 4.0, 5.0). There were 12 day samples, 12 night samples, and one duplicate sample (which was collected for quality assurance). These samples were filtered through a 0.22 μm Millipore filter, and then split for separate analysis of cations and anions. The cation samples were preserved with 1.25 mL of HNO_3 (1% by volume). All samples were put on ice and taken to the Illinois State Water Survey (ISWS) for analysis.

There were two serious problems encountered during the above measurements. During the four days and nights when we were working at the spring, an area (shown in Fig. 7.2) dried up. This may have resulted from changes in total spring flux or from our activities. Much like certain quantum mechanical systems, it is impossible to make measurements on carbonate hot springs without changing the results. As a consequence of these changes in spring water flow, two of our sample locations (2.5 and 3a) disappeared, and had to be eliminated from all subsequent measurements and analysis. Secondly, we had our water samples analyzed by the ISWS. Unfortunately and initially unbeknown to

Figure 7.4: Photograph showing collection of pH, temperature, and TDS data near the washers located at sample location 2.4.5. John Veysey and Tom Schickel are shown for scale. They stood on boards to protect carbonate features.

us, this laboratory only analyzes their samples subject to EPA requirements, meaning that their cation and anion concentration measurements are accurate to 10-15%. These shortcomings were identified because of our quality control sample. This is not sufficient for anything other than identification of broad systems level changes in chemistry. By the time we identified this problem, our samples were out of date. An unfortunate side effect of this problem was that we could not complete our DIC measurements. The titration technique we used required accurate water chemistry measurements in order to reconstruct the carbon containing chemical species, including the constituents of DIC. Nonetheless, the precipitation, pH, and temperature measurements are valid and form the basis of the subsequent analysis.[1]

7.2.1 Measuring DIC

When I joined the Yellowstone collaboration, Prof. Bruce Fouke and others had already obtained reliable water chemistry from spring AT-1. In each of the five facies, he had measured: temperature, pH, Ca^{2+}, Na^{+}, Cl^{-}, SO_4^{-}, Mg^{2+}, As, $B(OH)_3$, Ba^{2+}, K^{+}, Li^{+}, Mn^{2},$SiO_2(aq)$, Sr^{2+}, F^{-}, Br^{-}, Al^{3+}, Zn^{+}, Fe^{2+}, and TDS.

Figure 7.5: Photograph showing DIC measurement via titration. The nighttime crew shown here consisted of Nigel Goldenfeld and John Veysey (shown performing the titration) as well as Tom Schickel (shown supervising).

Unfortunately, the titrations which had been used to measure DIC were corrupted by poor reagents. Moreover, while a standard practice[168], titrations are not a good way to measure DIC in these highly charged spring waters. Many other workers have relied on them (e.g., [145]), but primarily in cooler streams which contain much less CO_2 than our system. In the time we need

[1] They were also used in the previous chapter.

to take a water sample from the spring to the titration station in the field[2], a significant quantity (e.g., 10%) of CO_2 can exsolve. In the 2004 trip, we took a number of precautions against this, including: a large volume of sample water, large volume to surface area ratios for sample containers, rapid titrations, and repeated re-sampling if the sample pH was observed to drop significantly below the *in situ* pH measurement. However none of these precautions were in place for earlier DIC measurements.

To avoid this problem, the Fouke group had used a program known as *the Geochemist's Workbench* (GWB). This program contains an extensive database of possible aqueous species, reaction rates, minerals, and has the ability to calculate activity coefficients using Debye-Huckel theory. By inputing the measured concentrations of most chemical species, as well as the $[H^+]$ implied by the field pH, we used GWB to iteratively calculate expected concentrations of other molecules. The only significant ion which was not measured was HCO_3^- (and other components of DIC). But because the water is electrically neutral we can calculate the concentrations of the missing ions by imposing charge balance. Thus the GWB can calculate HCO_3^-, DIC, as well as other derivative quantities of interest, such as $\Omega_{\text{aragonite}}$, Ω_{calcite}, and CO_2 fugacity.

This hybrid numerical and experimental approach looks wonderful in theory; it offers a way to measure DIC without worrying about outgassing during sample collection![3] However, charge balance can be unreliable because it relies on very precise measurements of hydronium ion concentration, as determined via pH measurements. Because pH is measured on a log scale, it is insensitive to small changes in $[H^+]$. To address this concern, we developed a technique to calculate the error in DIC (and all other derived quantities) using Monte Carlo simulations.

This approach is conceptually simple. Each of the reported concentrations has an associated analytical uncertainty. For instance, Ca^{2+} concentrations are measured to

[2]To say nothing of transport to a laboratory, as done by some other workers[145].

[3]It avoids the problem by relying on *in situ* measured pH. The cation and anion measurements are not affected by outgassing.

± 0.038 mg/kg. pH measurements have an uncertainty derived from triplicate field measurements. We assume that our measurements are drawn from a random normal distribution with a mean given by our true measurement, and a standard deviation given by analytical or experimental uncertainty. Assuming that the measured quantities are not correlated[4], we construct a synthetic data set by generating separate random numbers for *each* of the 20 experimentally measured quantities. We put this synthetic data set back into the Workbench, and recorded the resulting values for HCO_3^-, SI, etc. By repeating this process 15,000 times, we assembled a distribution of measurements for the calculated quantities.

By examining these distributions, we can calculate error, look for correlations, and even back out the specific reactions which GWB determined to be the most important in our system.[5] The results depend on the specific facies and quantity of interest, but proved to be surprisingly accurate; typical errors were $\sim 5\%$. Two examples of our results (Ω and CO_2 fugacity) are shown in Figure 7.6.

We had planned on a comparative analysis amongst the two kinds of field titrations and the charge balance / Monte Carlo approach to calculating DIC. The goal was to understand the primary sources of DIC measurement inaccuracy, and to quantify the effect of outgassing on field titrations. Unfortunately, the limitations of ISWS water chemistry made this impossible. Another member of the Fouke group, Michael Kandianis, has recently made excellent field DIC measurements using syringes and tightly sealed test tubes. At the time of this writing, these data were still being analyzed, and will be reported in his Ph.D. thesis.

[4]The derived / calculated quantities certainly *are*, but this is not a bad approximation for the measured inputs.

[5]We did this using principal components analysis (PCA) and non-metric multidimensional scaling (nMDS). However, it yielded nothing surprising — we learned that our system was dominated by carbonate chemistry! Essentially we used a sensitivity analysis of the Monte Carlo data set to deduce GWB's internal models.

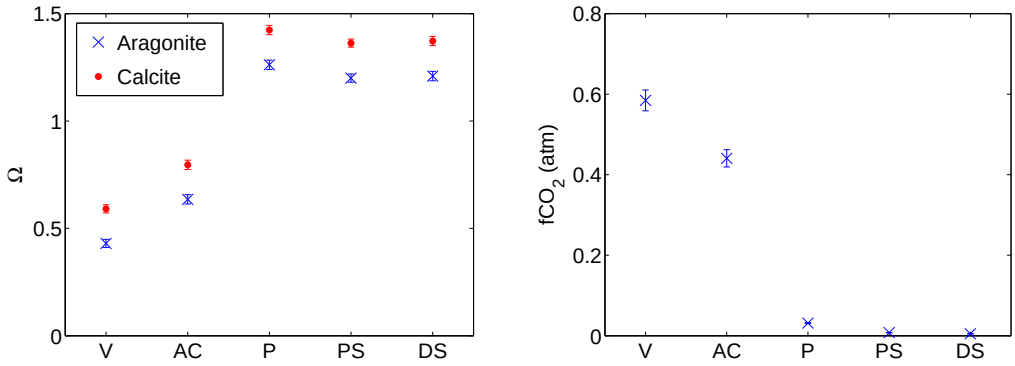

Figure 7.6: Results of GWB calculations for each of the five facies, based on data collected at AT-1 in 1999. Left figure shows Ω for calcite and aragonite. Right shows CO_2 fugacity. In both cases the error bars were determined by Monte Carlo.

7.3 Analyzing aqueous chemistry along the primary flow path

In chapter 5, we defined the primary flow path. Here we use it to analyze a portion of our measurements at AT-3. By studying how chemistry and precipitation change along a flow path, we can understand how macroscopic changes in chemistry impact deposition rates (and thereby travertine patterns). We demonstrate that rapid changes in height along the flow path result in increased outgassing and precipitation.

In section 5.8, we demonstrated that eight of our sample points from AT-3 (2004) lay along the primary flow path. Figure 7.7 shows the our $[Ca^{2+}]$, temperature, pH, and precipitation changes along the primary flow path, as a function of the distance from the vent. The precipitation measurements are averages based on the 48 hour washers, which represent the best long term average rates, and are not impacted by the nucleation kinetics of $CaCO_3$ crystallizing on steel washers.

This figure shows a number of interesting results. First, we see that $[Ca^{2+}]$ decreases as we progress along the flow path. This is expected, and is attributable to carbonate precipitation. Temperature decreases, owing to evaporative and radiative

cooling. Overall pH increases, an expected consequence of CO_2 exsolution.[6] There is, however, a sharp jump in pH beginning with the fifth data point (counting down the flow path from the vent). This corresponds with the largest precipitation rate seen along the flow path. These changes in pH and precipitation rate, occur at a sharp break in the flow path — a "waterfall". This feature can be seen in middle of the lower part of Fig. 5.5. The precipitation rate along the front of the waterfall is the largest seen in the system. The rates immediately downstream are also relatively large.

Our finding that hydraulic jumps result in increased outgassing, and therefore precipitation, is intuitive and has been reported by several workers studying *tufa* (travertine which precipitates from cool water)[133, 145]. While not unexpected, our results may be the first for carbonate hot springs.

Zhang advances a tripartite explanation for the increased degassing and precipitation seen at waterfalls[133]. His experiments and field observations argue that tufa formation at waterfalls is influenced by three factors. The first two may be grouped as turbulence effects. First, water going over a waterfall is aerated when it plunges over the edge. The increased surface area in contact with the water leads to increased outgassing. Secondly, Zhang identifies "jet flow" effects, in which rapidly moving jets of water break up into smaller droplets. This also increases the surface area available for outgassing. Finally, the water pressure at high velocities is reduced according to the *Bernoulli effect*. The lower water pressure enables more CO_2 to exsolve. This latter effect, which Zhang terms the "low pressure effect," occurs in both laminar and turbulent flows[133, 169]. It is not obvious which of these effects control precipitation in our system, but Figure 7.7 makes it clear that, if the water is charged with Ca^{2+} and DIC, then hydraulic jumps cause outgassing and precipitation.

Our data from the 2003 trip to AT-3 allow us to broaden our conclusions. During

[6] A large control on aqueous pH is carbonic acid. Exsolution of CO_2 reduces the concentration of carbonic acid, which in turn results in higher pH.

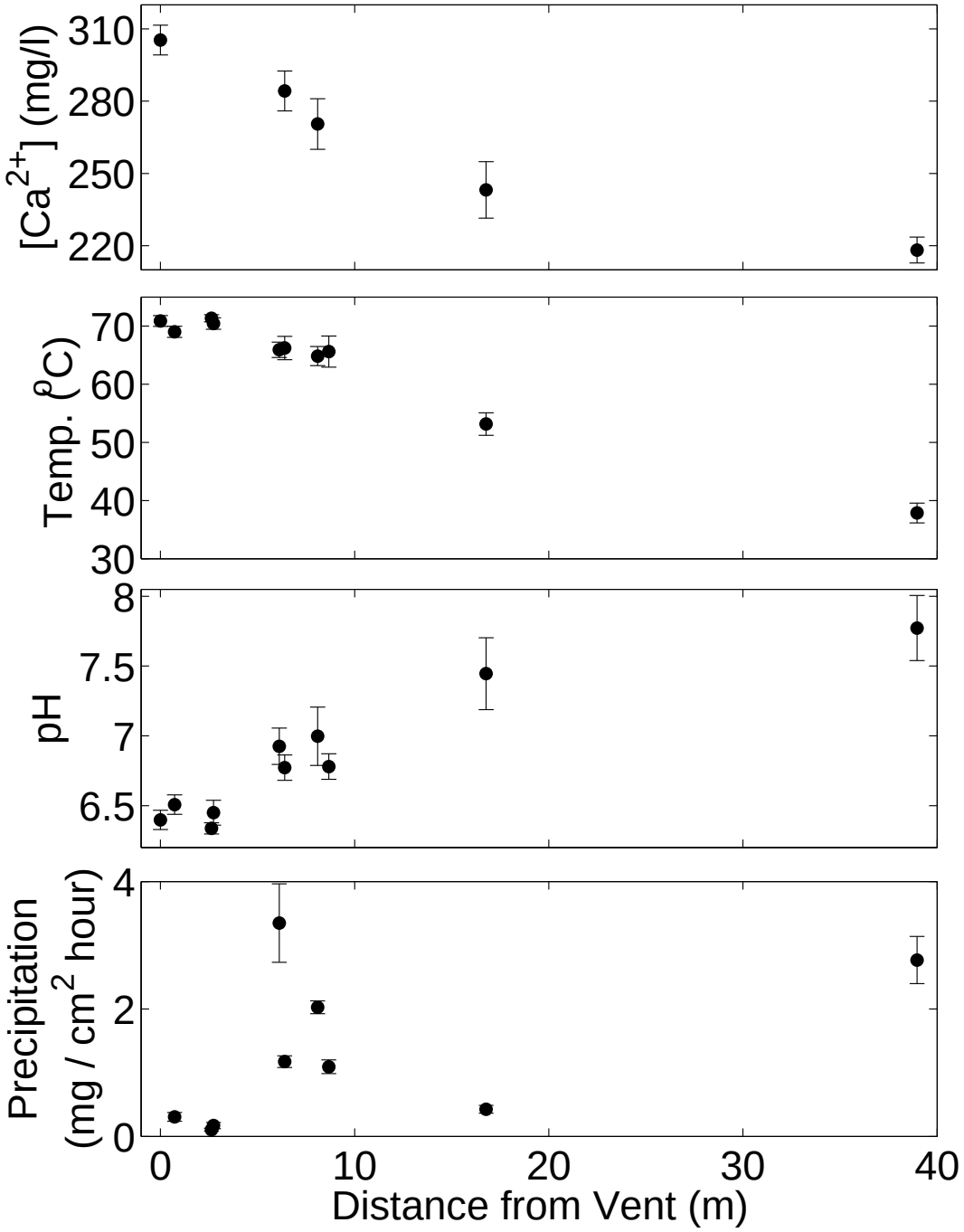

Figure 7.7: Chemical changes and precipitation rates along the primary flow path at AT-3 in 2004.

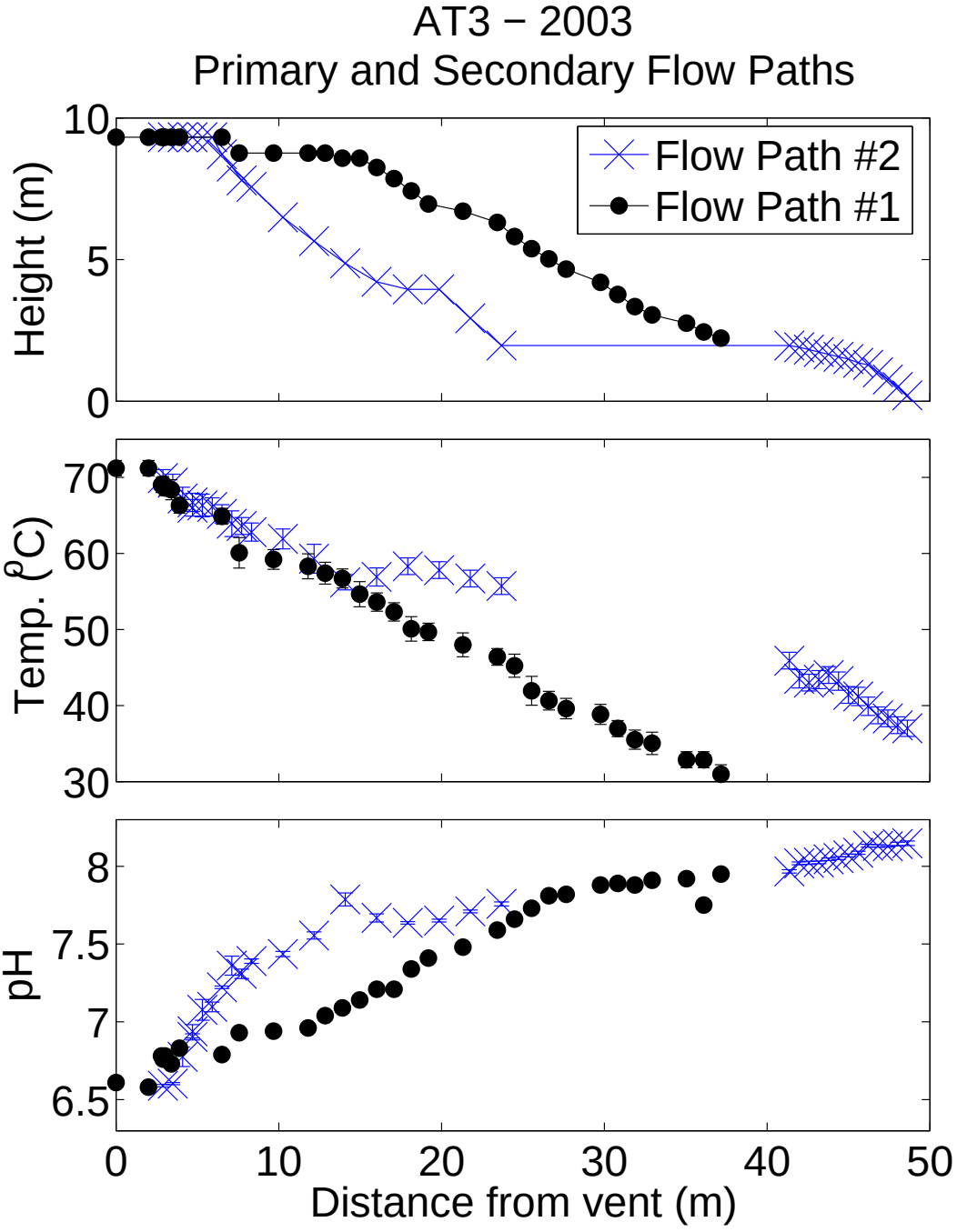

Figure 7.8: Changes in altitude, temperature, and pH along the primary and secondary flow paths at AT-3 in 2003. Note that one flow path is much steeper than the other. This results in more rapid outgassing, as can be seen from the sharper increases in pH.

our first trip to the spring, there were two distinct and physically separated[7] flow paths. Figure 7.8 compares changes in pH, temperature, and altitude along each of these two flow paths, which shared identical source water. There are large differences in $\mathrm{d}(pH)/\mathrm{d}s$ between the two flow paths, a result which we attribute to their very different slopes ($\mathrm{d}H/\mathrm{d}s$). The primary flow path descends much more gradually than the secondary one. While not a true waterfall, this steep slope nonetheless corresponds with rapid increases in pH. This implies that Bernoulli effects, which are associated only with increased flow velocities, impact precipitation rates in our system regardless of aeration or droplet formation.

Figure 7.8 also shows that flow path #2 is warmer at distal points. This again is a consequence of the increased flow velocity along the steeper slope; the water travels further faster, without the time needed to cool.

We can make one other inference using our flow path data. Figure 7.9 compares the percentage changes in hydronium ion concentration with changes in temperature for both of our primary flow paths at AT-3. It shows that $[H^+]$ initially decreases quite quickly, but then begins to decrease at a much slower (and nearly constant) rate after the first 15 m. We suggest that this indicates rapid outgassing of CO_2 early in the spring water's evolution. The difference between rapid upstream exsolution and slower downstream changes may indicate a change in the process which controls travertine deposition. In contrast to $[H^+]$, temperature decreases at a constant rate throughout the system, suggesting that evaporation alone cannot account for the observed differences between facies.

[7]They only spatial point in common was the vent itself.

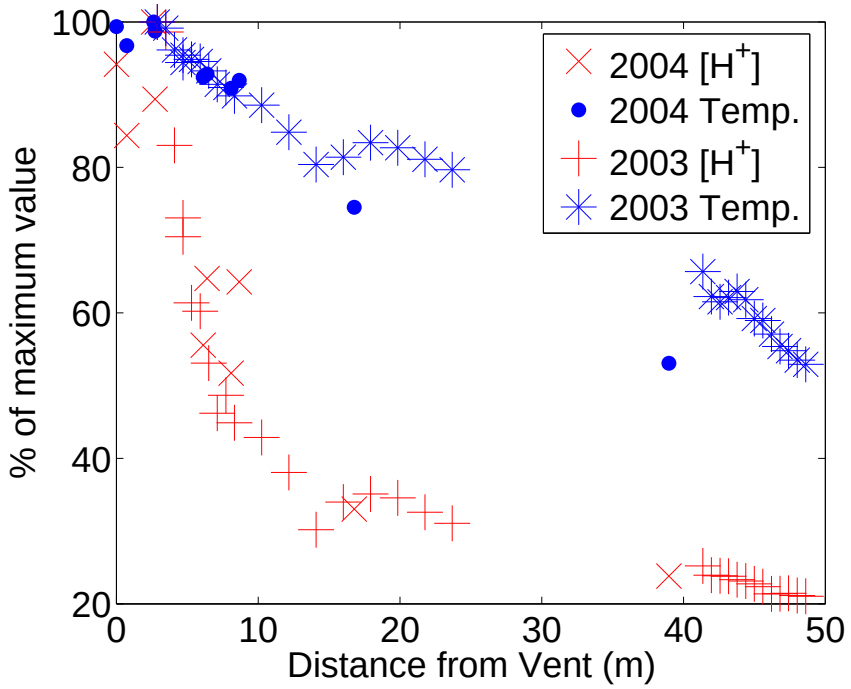

Figure 7.9: This shows the percent changes for both $[H^+]$ and temperature along the primary flow paths at AT-3 in both 2003 and 2004. Note that we have removed the $\log_{10}$ scale with which pH is defined. This shows that $[H^+]$ changes very quickly in the early parts of the spring, but flattens off. In contrast, temperature changes at an essentially constant rate throughout.

7.4 Looking for photosynthesis

We designed our 2004 trip to AT-3 to include three arduous 7pm - 7am shifts because we wanted to look for diurnal patterns in spring water chemistry and precipitation. Differences between day and night time measurements could indicate that photosynthesizing microorganisms influenced macroscopic chemistry or travertine precipitation. However, we saw no statistically significant differences between daytime and nighttime precipitation rates. Nor were any seen in springwater calcium concentrations.

Figure 7.10 shows each measurement of pH at a sample location normalized by the average value at that location. The results are color coded according to whether a measurement was collected during the day, the night, or in between. There is a slight, but not statistically significant difference between daytime and nighttime measurements:

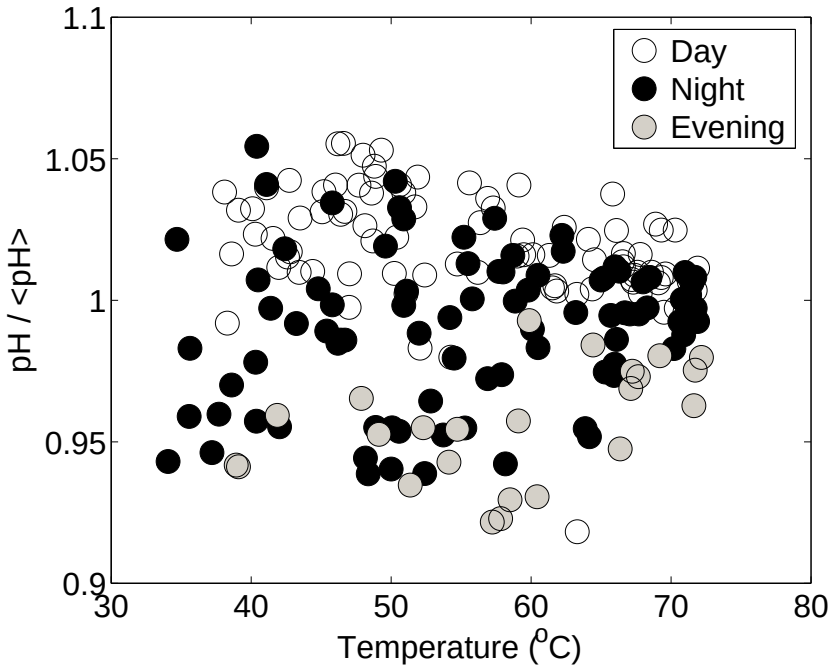

Figure 7.10: pH is normalized by the mean pH at its sample location (AT-3, 2004). The points are color coded to indicate whether the sample was taken by night, day, or twilight. Because not all these points lie on a flow path, they were plotted against the average temperature at their sample location. This permits consistent comparisons.

pH values measured during the day are higher than those measured at night. In general, this trend is not statistically significant. The only exception is in relatively cool ponds with low fluxes of water, where there is slight separation between daytime and nighttime measurements. But this effect need not be due to photosynthesis. It could be accounted for by differences in atmospheric temperature, meteorological conditions, or probe usage (different teams took the daytime and nighttime measurements).

On the basis of Figure 7.11, which shows differences in the day-night spring water temperature, we hypothesize that the pH changes in cool, slowly moving ponds are driven by changes in insolation and atmospheric temperature. In the thin sheet flow where the proximal slope data logger was placed, there are significant (almost 10^{o}C) changes between day and night spring water temperatures. The location had a similar total flux to the ponds which exhibited day-night changes in pH.

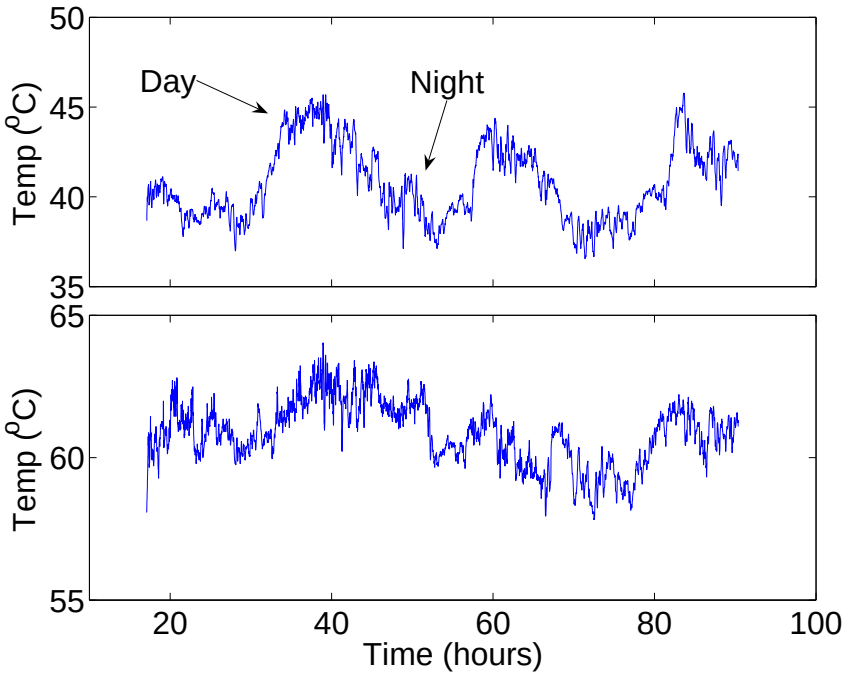

Figure 7.11: Plot of HOBO temperature measurements from AT-3, 2004. The data loggers recorded a temperature measurement every 30s. The bottom plot shows the results from a high temperature pond at sample location 3e.0. The top graph shows measurements from a relatively cool point in the proximal slope facies.

Even if our hypothesis is not correct, we have demonstrated that putative photosynthetic effects on water chemistry are restricted to very low flux ponds, and that they are small. Moreover, there is no evidence of a photosynthetic effect on precipitation rates in these ponds. Consequently, we conclude that photosynthetic microorganisms do not have a significant effect on macroscopic travertine patterns.

7.5 Comparing microorganism metabolic activities with inorganic processes

In light of our failure to find any evidence of photosynthesis, Michael Kandianis and I undertook a theoretical calculation, with the goal of comparing the magnitude of inorganic chemical processes with bacterial metabolic activities. Specifically, we wanted

to calculate an *upper limit* on the rate at which bacterial activities can remove dissolved inorganic carbon (e.g., CO_2), and a *lower limit* on the amount of DIC being removed from solution by chemical and physical processes (e.g., $CaCO_3$ deposition or CO_2 exsolution). We show that the *maximal* contribution of bacterial activities is much less than the *minimal* impact of abiotic processes. This demonstrates that the experimental results presented in the previous section were inevitable.

This section presents these calculations, which necessitate a large number of order-of-magnitude estimations. In order to track these estimations sensibly, all of our results here are expressed in terms of upper and lower bounds. Formulating our calculations in this manner results in rigorous estimations, which are suitable for comparing abiotic and biotic processes.

This calculation begins with DIC measurements, which were based on data collected at AT-1 in 1999. The actual DIC data, and associated error estimates, are the result of the GWB calculations detailed in Section 7.2.1. The AT-1 (1999) data set measured chemical concentrations during the day and night (albeit at a single point). For the purposes of this calculation, we average the two[8], with the relevant quantities shown in Table 7.1. These five sample points were located along the primary flow path, one in each facies. Note that we measure DIC in terms of mg of carbon. This is standard practice, as all components of DIC have one atom of carbon per molecule.

We first determine how much DIC leaves the system over the course of the flow path.[9] After converting the quantities in Table 7.1, we simply take the difference between DIC measurements at the end points:

$$\Delta \mathrm{DIC}_{\mathrm{total}} = (1.40 \pm 0.21)\mathrm{x}10^{-2} \frac{\mathrm{mol\ C}}{\mathrm{L}}$$

We want to determine how much of this change could be due to biotic and abiotic

[8]We performed the calculation in full generality, but observed no significant differences between the two sets of data.

[9]We performed this, and all subsequent calculations on a facies by facies basis, but present only the systems-level analysis here.

Facies	DIC (mg C/L)	Ca^{2+} (mg/L)
V	246.4 ± 17.2	303.5 ± 15.2
AC	194.8 ± 13.6	304.0 ± 15.2
P	115.3 ± 8.1	266.5 ± 13.3
PS	88.9 ± 6.2	229.5 ± 11.5
DS	78.4 ± 5.5	210.5 ± 10.5

Table 7.1: Measurements from AT-1 (1999) needed to estimated the relative contributions of abiotic and biotic processes. The DIC measurement is the result of a calculation with GWB, and its error the result of our Monte Carlo analysis.

processes. The most important abiotic processes are CO_2 exsolution and $CaCO_3$ precipitation. The former is difficult to measure (and is likely the largest contribution). The latter can be calculated from the second column of Table 7.1, by looking at Ca^{2+} changes, and assuming that carbon is removed from solution only by complexation with CO_3^{2-}. We obtain:

$$\Delta \text{DIC}_{CaCO_3} = (2.32 \pm 0.91)\text{x}10^{-3} \frac{\text{mol C}}{\text{L}}$$

This tells us that about 17% of dissolved carbon is incorporated into carbonate. This gives us a *lower* bound on the amount of DIC removed through inorganic processes.

So far this calculation has been unambiguous and straightforward. However, in order to compare the DIC flux due to precipitation with that due to biotic processes, we need to determine the DIC flux per unit area per unit time. This can be estimated by multiplying our previous result by Q_0/A, where Q_0 is the total spring flux, and A is the total area covered by all flow. Unfortunately, we know neither of these quantities reliably.

We have two measurements of the total flux, Q_0: $12,415 \pm 2244$ and $59,470 \pm 5947$ cm^3/s, and assume that these extremes bracket the actual flux at AT-1 when the chemistry samples were taken. This is supported by qualitative comparisons of the springs. Area can only be estimated crudely, based on photography and tape measure

Upper bound	285.0 μmol C/(m^{2}s)
Lower bound	42.0 μmol C/(m^{2}s)
"Best Estimate "	141.0 μmol C/(m^{2}s)

Table 7.2: Estimates of DIC flux due to precipitating $CaCO_3$. The upper and lower bounds bracket what is possible in our system.

surveys of the springs. But, for the purposes of upper and lower bounds, a crude estimate is usable. We estimated the area to be between 166 and 344 m^2, with our best estimate being 204 m^2.

Using these estimates, we computed upper and lower bounds for the DIC flux due to precipitation. These are summarized in Table 7.2, along with our "best estimate". To compute the lower bound, we used the *maximal* and the *minimum* flux — with the maximum error subtracted therefrom. The converse procedure was used for determining the upper bound.

Now we need to compare this to the possible impact of microbial activities. Although our sequencing work gives us some idea of possible metabolisms and species, there are few quantitative results available for bacterial metabolic rates, and they tend to be heavily focused on specific environments and bacteria. Nonetheless, we proceed forward, simply looking for the *maximum* possible metabolic effect. We found two useful results. First, Williams et al. reported a CO_2 *maximal* uptake rate of 0.3 μmol C/(m^{2}s) in dense algal blooms[170]. This is less than 1% of the minimal flux due to precipitation.

There are significant differences between our microbial ecology and that of snow algae. In particular, our communities live in the water column and in mats which line the floor of the hot spring. Moreover, our water is charged with CO_2, which could increase metabolic activity. Other workers have reported high CO_2 uptake rates in bacterial bioreactors which have been optimized (including CO_2 injection) to maximize growth rates[171]. They report maximal CO_2 conversion rates of 0.004 mol/(m^{3}s). To

compare this to our system, we need to estimate the height of the active bacterial communities. Allison Murray's optical and FITC (fluorescein isothiocyanate) cell counts indicate that the greatest cell densities (10^5-10^6 cells/ml) in our system are found in the microbial mats which line the system. These cell densities are at least an order of magnitude larger than the cell densities in the water column. To evaluate the metabolic activity due to the mats (assuming the water column is negligible), we multiply the mat thickness (less than 0.5 cm) by the reported CO_2 conversion rate. We obtain 20 μmol C/(m^2s).

Even if we assume that CO_2 uptake in our system is comparable to a highly optimized bio-reactor, it is still less than *half* of the *minimal* CO_2 flux due to precipitation. Furthermore, *that* flux is less than 20% of the total DIC flux in the system! Assuming the rest is due to CO_2 outgassing, the *greatest* possible biotic effect accounts for less than 10% of the total change in DIC. If one breaks this down on a facies by facies basis, it does not change the conclusions. Consequently, it is not a surprise that we did not find a diurnal signal due to photosynthetic activities.

7.6 Conclusions

In this chapter, we have described in detail our field work at Yellowstone National Park. We have explained why it is difficult to measure DIC, and offered a solution to this problem based on the software models of aqueous geochemistry. By combining this software (GWB) with Monte Carlo techniques, we can estimate the error in our computations, thereby estimating the reliability of our DIC calculations.

By analyzing our large data set along the primary flow path, we have identified a key chemical process which influences macroscopic carbonate precipitation. Waterfalls (hydraulic jumps) result in greatly increased CO_2 outgassing and precipitation rates. The same effect is seen on steeper (but not discontinuous) slopes; increased outgassing

and larger precipitation rates are seen anywhere flow velocity increases. This finding is a key component of the numerical models presented in the following chapters.

Finally, we demonstrated, through both experimental results and a theoretical calculation, that microbial metabolisms do not influence precipitation on a macroscopic scale through photosynthesis or CO_2 fixation.

Figure 7.12: John Veysey, Tom Schickel, and Nigel Goldenfeld enjoying breakfast for dinner after 36 hours of unbroken data collection on AT-3.

Chapter 8

Minimal models of terrace formation

8.1 Introduction

The previous two chapters described experimental results from YNP. We identified and explained macroscopic correlations between the microbial, chemical, and physical attributes of the spring water and resulting travertine depositional patterns. We have demonstrated that the metabolic activities of microorganisms cannot influence macroscopic chemistry or pattern formation, although the bacteria themselves may be important as nucleation sites, and may influence microscopic processes. We have also identified steep slopes (and concommitant high fluid velocities) as a cause of increased outgassing and precipitation. We have not, however, explained why terraces form at carbonate hot springs worldwide, as well as in low temperature speleothem rimstone dams.

The ubiquity of these large scale structures suggest that they may have a generic explanation, one which does not depend on detailed microscopic processes. In this and the next chapter, we explain the terrace morphology, its apparent scale-invariance, and the formation of circularly symmetric travertine domes (Fig. 5.6) via a minimal

model based on fluid dynamics, equilibrium chemistry, and the fact that increased precipitation rates are seen on steeper slopes. Although these travertine structures exhibit complex spatial patterns and temporal dynamics, our simple physical models explain their space-time correlations as arising from a depositional instability created by turbulent flow of supersaturated fluid over a surface.

By identifying the dynamical processes which control the formation of the large-scale patterns at carbonate hot springs, we have built two kinds of physical models. First, analytic models are used to predict specific morphological motifs, such as the profile of a pond lip and the shape of a circularly symmetric dome. Secondly, stochastic computer models are used to calculate the statistical properties of travertine patterns, such as the distribution of pond sizes. The deterministic models are built from differential equations which describe fluid mechanics and precipitation kinetics. These equations comprise a moving boundary problem, whos exact solutions describe the shape and evolution of specific carbonate patterns. These models also explain how instabilities grow to form the dams and ponds which characterize the terrace morphology on all length scales.

The stochastic models are cellular simulations governed by simple rules that caricature the precipitation and fluid flow dynamics on large scales. These models are solved numerically, and can reproduce the main instabilities which give rise to the terrace morphology. Cellular models are also used to calculate the statistical properties of carbonate patterns, to study the interaction and nonlinear dynamics of instabilities, and to explore the effect of surface tension in our model of the travertine dome. The work presented here has been submitted for publication[20], and represents the combined efforts of Nigel Goldenfeld, Pak Yuen Chan, and myself. Pak Yuen Chan is primarily responsible for the analytical models and solutions, and I for the numerical work.

In this chapter, we describe how these kinds of minimal models have been previously used in geomorphological problems. We present our cellular models, in both two and

three dimensions, and explain both our algorithms and their relationship to the physical processes which they characterize. In the next chapter, we describe the quantitative and qualitative results of our modelling. As our approach may at first seem too simplistic to describe our complex system, we begin with three examples of similar models which have been quite successful at explaining precipitative and geomorphological problems.

8.1.1 Examples of minimal models in geomorphology

Four other projects provided guiding examples for our work. First, Grotzinger and Rothman showed that layer fluctuations in ancient stromatolite rocks exhibit power-law correlations accurately captured by a simple generic model[172]. Stromatolites are layered rocks which are often said to be biotic in origin, and are therefore considered to be a proxy for early life on earth[173]. Grotzinger and Rothman showed that the layers in two photographed stromatolites have a self-affine fractal geometry, and that their power spectra obey the same power law ($\langle S(k) \rangle = Ak^{-\beta}$)[172]. They determined the fractal dimension, and treated that as a one parameter characterization of stromatolite growth dynamics. They then showed that this fractal dimension, as well as the power law power spectra, could be explained using a minimal model based on the KPZ equation[174]. This result unequivocally demonstrated that stromatolitic patterns could arise by purely abiotic processes. Our work at Yellowstone closely parallels their project. We address a similar question (i.e., Can carbonate terraces form abiotically?), characterize real carbonate patterns statistically, and show that abiotic minimal models can produce patterns with the same statistical properties.

Secondly, Murray and Paola used a cellular model to explain the statistical properties of braided river beds[137, 175]. They modeled networks of channels, known as *braided rivers*, that form in gravel river beds and exhibit complicated dynamics as the channels move, split, and rejoin. They developed models based on simple, approximate descriptions of turbulent fluid flow, sediment transport, and mass conservation. By ap-

plying these models to a uniform slope with small random noise, they reproduced the main temporal dynamics and spatial features of braided rivers. Although their model never reaches a steady state, it is *statistically* stable, and produces patterns with similar statistical properties as real braided rivers.

However, as noted by Doeschl et al., the quantitative comparisons undertaken by Murray and Paola are not sufficient to demonstrate that their model actually characterizes natural braided rivers[176].[1] Because their CA models cannot be directly compared with particular natural rivers, it is vital to separate universal from non-universal questions; meaningful comparison requires "the identification of characteristic properties common to all braided rivers."[176]. In spite of these weaknesses in comparing the predictions of their model to real braided rivers, Murray and Paola's success at reducing a complex problem involving both fluid flow and depositional patterns to simple cellular model helped motivate the work described in this chapter.

Pastor-Satorras and Rothman modeled anisotropic erosion using a stochastic differential equation[11]. Using renormalization group techniques, they analyze this equation, predicting the statistical properties of landscape height variations. Their simplified model gave rise to a realistic topography, which agrees fairly well (in a statistical sense) with real landscapes. We would like to apply their methodology to our analytical models (with an appropriate stochastic term), with the goal of obtaining statistical descriptions of terrace patterns, such as the distribution of pond areas. These statistical characterizations would ideally match both the cellular simulations and real hot springs. To date our only such results have come from simulation, as the governing equations are formidable to solve even without stochastic terms. However, the work of Rothman et al. is an example of how RG techniques can formally extract macroscopic patterns and statistical descriptions from minimal models of topography. This

[1] Doeschl et al. also warn more generally of a common problem with CA models: "The assessment of these models is often restricted to visualization of the results and lacks rigorous analysis"[176]. We have avoided this pitfall in our work.

paradigm underlies our approach to modelling throughout this thesis and explains how our analytical and numerical approaches could be unified. Although we have not accomplished this specific calculation, we will show how our analytic theory predicts data collapse which is seen in the results from simulations.

Finally, Short et al. recently used a deterministic model to explain the characteristic shape of stalactites[177, 178]. Although this work was published after the start of our project, it provides an excellent example, and is the only work comparable to ours which quantitatively models carbonate precipitation pattern formation. They coupled an analytic model of laminar thin film fluid flow down a constant slope to a solution for the reaction-diffusion equations which describe carbonate precipitation, and obtained a self-similar solution for the shape and velocity of the tip stalactites. This solution depends only on gravity, total fluid flux, viscosity, and the rate of precipitation.[2] This two parameter model is an example of free boundary dynamics in precipitative pattern formation, explains the shape of the tip of *all* stalactites, and exemplifies the analytical approach we followed to obtain solutions which describe the shape of pond lips and travertine domes.

As my contributions to this modelling program have focused on numerical simulations, I will now present the cellular models in detail. These results, together with the analytical solutions of Pak Yuen Chan, will be presented in the next chapter.

8.1.2 What are cellular models?

The concept of *cellular autotmata* (CA) has been widely used[179], over a number of years, in fields ranging from fluid mechanics[180] to HIV infections[181] to biology and ecology[182]. Books have been written on the subject[183]. As a result of this diversity of application, the idea of a CA has become quite broad and been saddled with a mixed reputation. We instead use the term *cellular* to describe our approach because, while it

[2]This is expressed though a mass transfer coefficient, R.

is discrete in space and time, the dynamical variables may be continuous. Such models are termed *cell dynamical systems* (CDS)[184]. CDS models have been previously used to explain physical problems, as well as geomorphological ones. What is important is the notion that we are only seeking to understand universal features, insensitive to discretization, such as dynamical correlation functions. For instance, in the phase ordering of two-dimensional aligned nematics, the measured disclination correlation function[6] obeyed with no adjustable parameter the scaling predicted on the basis of minimal model cell dynamical system simulations[5].

In these models, we represent the quantities of interest (e.g. the height of the landscape) as *cells* on a lattice. A cell interacts with neighboring cells through *rules* which are abstractions of the real physics which governs the system. These rules do not try to describe in detail the real processes, but instead aim to capture only the essential, minimal components needed to reproduce the system of interest. In our case, we aim to describe the formation of terraces and resulting statistical patterns, but will not try to model microscopic crystallography or mesoscopic carbonate fabrics.

8.1.3 Why CDS models?

We use CDS models to simulate terraced landscapes in the fully non-linear regime, which cannot be readily described by analytic solutions. Because the statistical properties of such landscapes (such as height - height correlation functions) often have a universal character, they can be computed essentially exactly, with no adjustable parameters by minimal CDS models. Such models are easier to implement than numerical solutions of differential equations, and allow comparason of different governing equations and dynamics. Furthermore, their simplicity results in computational speed, allowing us to examine many different initial conditions, and to accumulate the repeated trials necessary for computing the statistical properties of the evolved landscape.

The CDS models have allowed us to identify, via numerical experiments, the min-

imal set of rules needed to reproduce travertine patterns, including both domes and terraces. For instance, the analytical solution describing the shape of travertine domes agrees well with only the top half of observed domes. Without surface tension, our CDS model agrees with the analytical solution. However, when capillary effects are added, the simulated domes are virtually indistinguishable from the real ones. These effects are not so easy to add to the analytic governing equations.

Moreover, the CDS model which produces the domes is exactly the same as the one which models terrace formation! The only differences are in parameters (the flux of spring water and initial chemistry) and the initial conditions (a sloped surface instead of flat one). This demonstrates that the physics which the model captures is the same in both patterns. This comparison is far easier to do numerically than analytically.

8.2 2D cellular models

We introduce our CDS models with a very brief treatment of 2D models. This is primarily done because it will help in understanding the more complicated 3D models, but it also has some intrinsic value.

Figure 8.1 shows the components of a simple 2D CDS model. Each lattice site (cell) contains two fields, one of which contains the height of the underlying surface, and the second, which records the height of fluid at that location. These are the only components of this model, but we do need to specify the rules by which they interact. During each time step, the cells in the simulation are visited in a random order, and two things happen. First, the water moves, and second the surface grows according to the water flowing above each point.

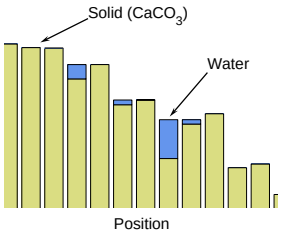

Figure 8.1: Schematic for 2D CDS models.

In our model of fluid transport, we model thin gravity driven flows. The water at

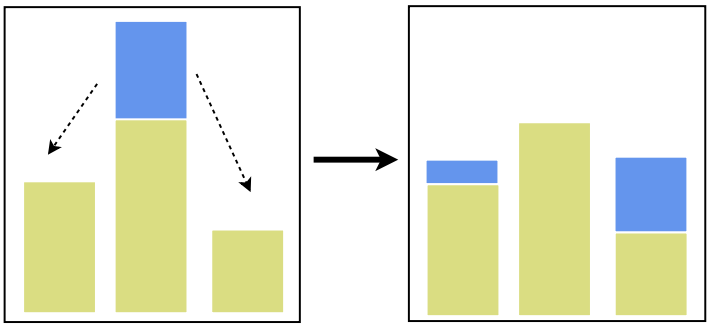

Figure 8.2: Illustration of 2D CDS water transport rule. The initial configuration on the left turns into the final configuration on the right, as the water (blue) initially in the center cell is moved to the lowest available configuration.

each cell is put into the lowest local configuration. This is illustrated in Fig. 8.2. Note that water can only flow between neighboring cells. One of the important facets of this model is that it allows *ponding* to occur. This means that water at a cell can stay stuck in a local minimum indefinitely.

After every cell has been visited, and its fluid transported according to the first rule, we apply the second rule to determine how the underlying surface evolves. Implicit in this sequence is a separation of time scales. Because the surface growth rate is of order 1-5 mm/day and the fluid flow rate is of order 1 mm/s, these two processes are independent, and can be implemented sequentially. For our growth rule, we determined the growth at each cell according to: $\Delta H \propto \vec{F} \cdot \hat{n}$, where $\vec{F}$ is the local flux of fluid, and $\hat{n}$ the direction of the surface normal, and ΔH the change in height. This rule describes growth due to transport of solute to the surface. To yield a nontrivial result in 2D, $\vec{F}$ was calculated using a forward derivative, and $\hat{n}$ using a backward one.

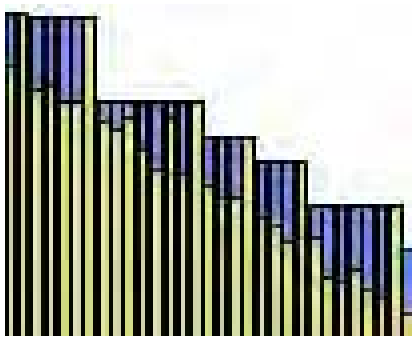

Figure 8.3: Result from 2D CDS models.

The CDS simulations were run on initial conditions which consisted of a constant slope plus small random normal noise. In addition to the two rules, we also added

some "vent" water at the top of the slope during each time step. An example of a typical result is shown in Fig. 8.3.

Although this 2D model is almost absurdly simple, it leads to a couple of key insights for the more complicated 3D case. First, this model produces a dynamical effect which we call *pond drowning*, in which a faster growing downflow pond causes its more slowly growing upstream neighbor to be flooded. The remains of ponds which have been innundated are visible in Figure 8.3. This suggests a mechanism by which smaller ponds can aggregate to form larger ones. This phenomenon turns out to be real, is seen in both YNP hot springs and our 3D models, and will be discussed further in the next chapter. The 2D model also indicates the importance of ponding. When an area fills with water, it will have — by construction — a locally flat surface. This means that it will cease to grow, because the dot product in our growth rule will always vanish. This is consistent with the observation in the real system that precipitation is fastest on steep slopes. Water flow models which do not allow ponding cannot capture this important behavior.

However, there are problems with this model. Most significantly, the continuity equation (Eqn. 2.1) (e.g. conservation of water), has a much more important role in 3D. In 2D, water flows down the same path regardless of any obstacles in front of it; if a cell grows to be higher than its upstream neighbor, a pond necessarily forms. In 3D, water will simply flow around obstacles. This means that a growth mechanism which leads to 2D ponds may not do so in 3D. Secondly, $|\vec{F}|$ becomes meaningless; all points on the slope will have the same volume of water flowing over them. Third, we cannot implement a sensible growth law which depends directly on local slope. The normal transport rule adopted here does involve slope, but only indirectly. In 3D we can have a more realistic model of precipitation rates. Finally, this model of fluid flow does not account for any turbulent effects. All of these shortcomings are remedied by our 3D model.

8.3 The 3D CDS model

The 3D model is similar in spirit to its 2D counterpart. However, we add one important feature: chemistry. The 3D model contains a caricature of the overall carbonate precipitation reaction (Eqn. 5.1). To implement this, we introduce variables which describe chemical concentrations, and which are advected by the moving fluid. However, the chemical modelling is *not* necessary to reproduce the basic phenomena of water ponding and terrace formation. It is necessary for the modelling of more sophisticated features, such as travertine domes, and the vent facies.

As in the 2D case, we represent the landscape as composed of stacked "bricks" or cells and represented by its height $H(i,j)$, above a horizontal reference plane, where i and j are $x-y$ coordinates in the reference plane. The water column is situated above the height field and represented by the variable $W(i,j)$ describing the volume of water above each coordinate element.

Each packet of water can also contain calcium ions, $C(i,j)$, and dissolved carbon dioxide vapor, $V(i,j)$, which may potentially cause precipitation through a caricature of the overall reaction pathway ($Ca^{2+} + 2HCO_3^- \rightleftharpoons CaCO_3(s) + H_2O + CO_2(g) \uparrow$). We model this reaction as controlled only by two species ($C(i,j)$[3] and $V(i,j)$). Crudely, precipitation cannot occur unless there is CO_2 exsolution, and will not occur if there is insufficient Ca^{2+}.

An important underlying assumption in our model is that local minima in the surface (e.g., a depression) fill up completely with water. We refer to those local minima as *ponds*. The presence of ponds influences water movement and precipitation. To expedite our calculations, we also keep track of the water that is ponded, $W_p(i,j)$. Finally, we record the temperature of the water $T(i,j)$. Neither of these last two fields is fundamental to the model; $W_p(i,j)$ can always be determined from the geometry of

[3]More precisely, this variable describes the total concentration of either Ca^{2+} or HCO_3^-, which ever is rate limiting. We assume that only one of them limits the overall reaction.

the surface, and $T(i,j)$ is introduced solely for comparing the CDS models with the real system; temperature has no effect on the simulation.

A complete lattice update consists of four steps, which implement the rules which caricature fluid flow, precipitation, and carbonate chemistry. These steps are:

1. Add water to the system, at the spring source, which is taken to be the origin of our coordinate system.

2. Propagate all the water in the system by moving packets to nearest and next nearest neighbor grid points according to the fluid flow rule.

3. Update the water chemistry (e.g. $C(i,j)$ and $V(i,j)$) to take into account outgassing due to fluid motion and depletion of Ca^{2+} ions due to precipitation.

4. Evolve the height field in response to the precipitation of $CaCO_3$.

We now describe each of these steps in detail.

8.3.1 Step 1: Adding Water

First, a quantity of water δW is added at the source, so that $W'(0)+H(0) = W_0+H_0(0)$. Here and below, primed quantities denote the updated variables. W_0 is a constant value appropriate for a constant pressure head. $H_0(0)$ is the initial landscape height at the origin. This expression accounts for the change in pressure due to the vertical growth of the landscape. It implies that, as the landscape height increases, the total flux from the spring will drop, and eventually cease. This means that the simulations will have a definite end.

The new water added to the system contains initial concentrations of calcium, C_0, carbon dioxide, V_0, and is at a temperature T_0; the fact that the source water is undersaturated is represented by $C_0 < V_0$. The values of these fields at the source

point are updated based on the volumetric ratio of the amount of water added to the existing water:

$$C'(0) = \frac{C(0)W(0) + \delta W C_0}{W_0} \tag{8.1}$$

$$V'(0) = \frac{V(0)W(0) + \delta W V_0}{W_0} \tag{8.2}$$

$$T'(0) = \frac{T(0)W(0) + \delta W T_0}{W_0} \tag{8.3}$$

8.3.2 Step 2: Fluid Transport

The transport of water is carried out by a variation of an algorithm used for braided river flow[137], in which the flux along bonds connecting a given lattice point to one of its eight closest neighbors is determined by the landscape gradient along that direction, while conserving the total volume of water.

An important part of our model is that the relevant surface over which water flows is not simply H. Instead, water flows in response to the combination of filled ponds and the underlying surface — the total surface $H_S = H + W_P$. This means that the water in ponds is modelled as flowing along the surface of the pond. For this to work, we require all ponds to be completely filled. The algorithm by which this is done is detailed in Appendix 8.5, and assumes that pond filling is independent of W. This is reasonable since the changes in pond volume are much less than the volume of water flowing over the surface: $\delta W_P \ll \delta W$.

For the purposes of water propagation, a pond is simply part of the landscape, and by requiring ponds to be completely filled, we gain an important algorithmic advantage: there are no "sinks" where water will accumulate. When flow from the source stops ($W_0 = 0$), all of the water will flow out of the simulation: $\lim_{t\to\infty} W(x, y) = 0 \,\forall (x, y)$

Consider a point on the lattice, $\vec{r}_0$, and its eight nearest neighbors. $W(\vec{r}_0)$, the water at that point, can move into its neighbors according to rules which caricature the basic physics of the Navier-Stokes equations. In that regard, we consider two

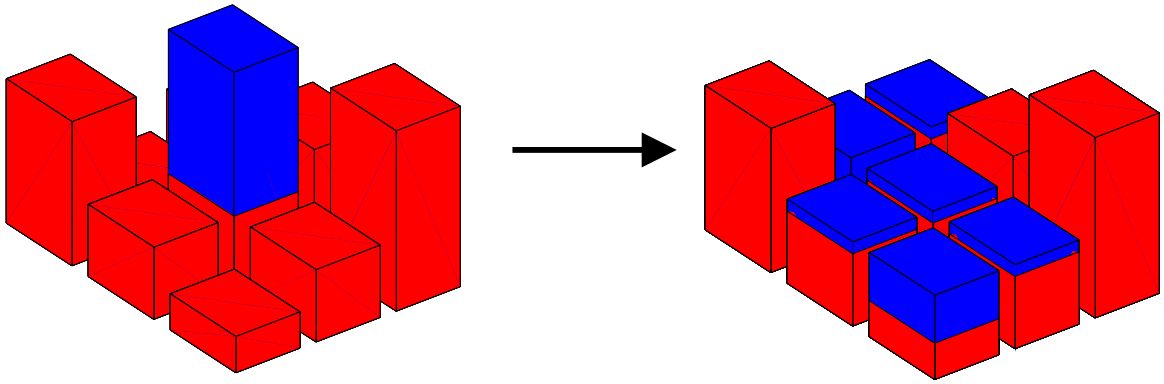

Figure 8.4: Illustration of 3D CDS water transport rule on a flat surface. The initial configuration on the left turns into the final configuration on the right, as the water (blue) initially in the center cell is moved to the lowest available configuration.

extremes. On a steep slope, water flows downhill subject to constraints from gravity and turbulence. On a flat underlying surface, water flattens out locally, as in Fig. 8.4. Which case controls each point is determined by this question: Can all of $W(\vec{r}_0)$ move into the neighbors, or will a lower energy configuration result if some water remains at the point $\vec{r}_0$?

To make this precise, consider a point on the lattice and its eight nearest neighbors. Label that point by $\vec{r}_0$, and its neighbors by $\vec{r}_j$ ($j \in [1, 8]$). Let $S_j = H(\vec{r}_0) - H_T(\vec{r}_j)$, where $H_T \equiv H_S + W$. Set $S_j = 0 \; \forall \; S_j < 0$, so that we can have an unrestricted sum. Define $N = \sum_j S_j$. If $N \leq W(\vec{r}_0)$, then the lattice site is flat relative to the the amount of fluid. In this case, we distribute $W(\vec{r}_0)$ among all *nine* cells (the site under consideration and its nearest neighbors) so as to minimize the total height of the resulting configuration (Fig. 8.4). This is analogous to the 2D transport rule.

If, on the other hand, $N > W(\vec{r}_0)$, then we are in a situation where all of the fluid will flow from the cell into its neighbors, as on a steep slope. In this case, the flow is taken to be turbulent and the flux is proportional to $\sqrt{S_j}$ in accord with Chézy's law. The flux from $\vec{r}_0$ to $\vec{r}_j$ is given by Eqn. 8.4[137]. In contrast to the model of Murray

et al., this equation does not allow any uphill flow of water (because S_j is constrained to be positive definite).

$$Q_{\vec{r}_0 \to \vec{r}_j} = \frac{W(\vec{r}_0)\sqrt{S_j}}{\sum_k \sqrt{S_k}} \tag{8.4}$$

Note that our bipartite fluid transport model extends the work of Murray et al.[137] even to flat surfaces, where $\sum_k \sqrt{S_k} \approx 0$.

In thin films near the rim of a pond[185] or on travertine domes, the water is subject to surface tension and contact line effects. We model these by requiring that water at a point ($\vec{r}_0$) is propagated only if $W(\vec{r}_0)$ exceeds a small threshold at that point.

Packets of water carry with them advected variables T, C and V, which are updated by the volumetrically-weighted average of all the neighborhood points from which the water packet originated. Finally, at each time step, we impose zero boundary conditions on our model: $W(x, y) = 0 \; \forall \; (x, y) \in$ perimeter.

Note that the order in which cells are visited does not matter in our transport rule. $W'(x, y)$ is computed for *all* cells (using the current W) before any cells are updated. This will be true for our chemistry and precipitation rules as well.

8.3.3 Step 3: Water chemistry

Our water chemistry model caricatures two physical processes: outgassing and precipitation.[4] V changes when there is outgassing and C does not. Both C and V change to reflect mass balance (Eqn. 5.1) as $CaCO_3$ is precipitated.

[4]The components of the chemistry model also change due to mixing water, as described above.

We model outgassing according to the following considerations:

$$\lim_{v(x,y)\to\infty} \frac{\delta V(x,y)}{V(x,y)} = 1 \tag{8.5a}$$

$$\lim_{v(x,y)\to 0} \frac{\delta V(x,y)}{V(x,y)} = f_1 \tag{8.5b}$$

$$\lim_{W(x,y)\to 0} \frac{\delta V(x,y)}{V(x,y)} = 1 \tag{8.5c}$$

$$\lim_{W(x,y)\to\infty} \frac{\delta V(x,y)}{V(x,y)} = 0 \tag{8.5d}$$

Here f_1 is a small constant, and v is the fluid velocity.

Eqns. 8.5b, 8.5c, and 8.5d are explicitly captured by the rule[5]:

$$V' = \left(f_1 V + f_2 \frac{\sqrt{S_j}}{1+\sqrt{S_j}} V \right) f_W \tag{8.6}$$

Here f_2 is also a small constant, and $f_W \equiv W/(W+\delta_w)$. f_W reflects increased outgassing in very thin films, but it is not essential for reproducing the terraced morphology. δ_w represents thickness of the boundary layer through which CO_2 diffuses at the air - water interface. Eqn. 8.6 allows a constant fraction of CO_2 to outgas at each time step, with an additional outgassing component proportional to $\sqrt{S_j}/(1+\sqrt{S_j})$ to reflect the influence of slope-initiated turbulent flow. Note that this latter term involves S_j from the previous rule, and is actually most conveniently implemented during fluid transport.

As $CaCO_3$ is deposited, the concentrations C and V change to reflect Eqn. 5.1. If, in step (4), a mass amount of $CaCO_3$ with volume δH precipitates ($\delta H = H' - H$), then the two chemistry fields change according to:

$$C' = (C - \delta H) f_W \tag{8.7}$$

$$V' = (V + \delta H) f_W \tag{8.8}$$

Temperature is updated along with water chemistry. Because it is not changed by either outgassing or precipitation, it is useful for tracking fluid flow. We update

[5] 8.5d is irrelevant in our system, as W is never large, except in ponds, which are treated separately.

temperature according to: $T' = f_3 T f_W$, where f_3 is a constant denoting the fractional change in temperature during each time step (e.g., $f_3 \sim 0.99$).

8.3.4 Step 4: Precipitation

The evolution of the height field H is the final step in the lattice update, with a change given by:

$$\delta H = \Omega \times f_P \left(R_1 + R_2 \vec{F} \cdot \hat{n} + R_3 \sqrt{S} \right) \frac{f_W}{\alpha} \tag{8.9}$$

As this expression is somewhat complicated, we explain each of the terms individually. $\Omega \equiv C - V$ mimics the real measure of saturation state (Eqn. 5.2), and describes the tendency of Eqn. 5.1 to run forward or backwards. Note that Ω can be negative, allowing for dissolution. Combined with the initial condition $V_0 > C_0$, this prevents deposition from immediately occurring at the source, and mimics the real system.

The second term, $f_P \equiv W_p/(\delta_p + W_p)$ describes the lack of precipitation in deep ponds due to slow mixing. δ_p is the characteristic pond depth at which this effect becomes important. Because $CaCO_3$ precipitation (Eqn. 5.5) occurs at the bottom of the pond, while CO_2 exsolution occurs at the air-water interface (Eqn. 5.4), the individual reactions are limited by diffusion or mixing within the water column. This effect has been documented by Lu et al.[134].

The third term caricatures the kinetics of crystallization which determine the *rate* of precipitation. Here R_1, R_2 and R_3 are constants, $\vec{F}$ is the flux between cells, with a direction given by the component of the gradient of the energy surface H_T between cells and magnitude given by the volume of water propagation per unit time step, $\hat{n}$ is the unit vector normal to the underlying surface H, and $\sqrt{S}$ represents the increased precipitation due to Bernoulli effects and local turbulent degassing. The overall term must be less than or equal to one, as it represents the rate at which the reaction tends toward equilibrium as determined by Ω.

The R_1 term represents the relatively small basal precipitation rate which occurs as the system tends toward equilibrium. The R_2 term characterizes precipitation due mass transport incident on the underlying surface, as in the 2D CDS model. The R_3 term reflects our experimental observations. The balance between R_2 and R_3 results in ponds forming where there is low flux, and flat regions forming where there is large flux.

The final term, f_W/α arises because of a new consideration. As before, f_W reflects the fact that very thin flows have less material to precipitate, and should result in small δH. However, α is a new constant which satisfies $\alpha \geq 2$. This constraint arises because of the requirement that Ω cannot change sign as a result of Eqn. 5.1. If $C > V$ before precipitation, then $C' > V'$ afterward, and conversely for dissolution. Using Eqns. 8.7, 8.8, and 8.9, it is readily shown that this requirement is always satisfied if $\alpha \geq 2$.

8.4 Conclusions

In this chapter, we have explained and motivated why we are building minimal models to describe macroscopic carbonate patterns at YNP. We have reviewed other examples of these kinds of models which have successfully solved problems in carbonate pattern formation and geomorphology. We have presented in detail 2D and 3D cellular models. These models use simple rules to capture the essential fluid dynamics, chemistry, and precipitation processes in our system.

In the next chapter, we will describe the results of our modelling approach, including the landscapes which result from the 3D CDS simulations described in section 8.3. Beginning with both flat and inclined surfaces, we iterate the governing rules until the flux from the source is "choked off" by the growing carbonate landscape, as in real carbonate hot springs. By varying flux, initial chemistry, as well as the roughness of the starting landscape, we reproduce qualitative and quantitative features of real

travertine hot springs, including both terraces and circularly symmetric domes.

8.5 Appendix 1: Pond filling algorithm

While the rules governing fluid transport are conceptually straightforward, there are some algorithmic technicalities, most significantly in calculating W_p, the amount of water it takes to fill the local minima termed *ponds*. The desired solution is easily described: Put a large amount of water in every cell on the lattice, and let it drain away through the edges. W_p is given by the water which cannot drain from the landscape, and the surface $H_S = H + W_p$ is flat wherever $W_p > 0$. Unfortunately this approach can be computationally inefficient. Unless care is taken, finding and filling ponds becomes a bottleneck in the CDS models. In this appendix, we present the algorithm we used to mitigate this problem.

First, identify the maximum possible pond depth (P_{max}), either by computing the maximum height difference in the landscape or by considering the maximum growth allowable given the specified chemistry (V_0, C_0). Algorithm speed does *not* depend on this value; one could just as easily pick an absurdly large number. Initialize $W_p = P_{max}$ at all lattice sites, and then set $W_p(\vec{r}) = 0 \; \forall \; \vec{r} \in$ edges.

Next, initialize a stack which consists of the coordinates of all lattice sites that are adjacent to sites with $W_p = 0$. By construction, this is the "ring" of cells one removed from the edge of the simulation.

We "pop"[6] a pair of coordinates from the stack ($\vec{r}$)[7], and determine whether the pond water at this site, $W_p(\vec{r})$, can drain through the empty cell(s) to which it is adjacent. We determine this by looping over all neighboring cells ($\vec{r}'$), and determining whether $W_p(\vec{r}') = 0$. If it is, we ask whether that neighbor is lower than our cells ($H(\vec{r}') < H(\vec{r})$). If so, $\vec{r}$ is not in a pond, and we accordingly set $W_p(\vec{r}) = 0$. If

[6] "Popping" removes the top element from the stack.

[7] This doesn't depend on whether the stack is last-in-first-out (LIFO) or first-in-first-out (FIFO).

$H(\vec{r}') > H(\vec{r})$, we then adjust $W_p(\vec{r})$ such that $H_S(\vec{r}) = H(\vec{r}')$.[8] If $W_p(\vec{r}) = 0$ due to either scenario, we then add the coordinates of *all* of the nearest neighbors which satisfy $W_p(\vec{r}') > 0$ to the original stack.[9] This process repeats until the stack is empty.

Next, we identify contiguous "ponds", groups of cells which all have $W_p(\vec{r}) > 0$, and are connected. There are several ways to accomplish this. For instance, loop over all cells, stopping if $W_p(\vec{r}) > 0$. Create a list, which will be filled with the coordinates of all cells within a contiguous "pond", add the coordinates $\vec{r}$ to the list, and set $W_p(\vec{r}) = 0$. Then visit each of the nearest neighbors, denoted by $\vec{r}'$. If $W_p(\vec{r}') > 0$, add $\vec{r}'$ to the list, and set $W_p(\vec{r}') = 0$. Repeating this process recursively leads to a list which contains the coordinates of all cells within that pond, and repeating the process until all $W_p(\vec{r}) = 0$ leads to a set of lists, one for each pond.

At this point, we have identified all points which lie in ponds, but we have not determined the appropriate values of W_p which result in a locally flat surface H_S. To accomplish this, we apply the following procedure to each pond. Loop over all cells within the pond. At each cell ($\vec{r}$), loop over its nearest neighbors ($\vec{r}'$). If $H_S(\vec{r}') < H_S(\vec{r})$, adjust $W_p(\vec{r})$ s.t. $H_S(\vec{r}') = H_S(\vec{r})$. Repeat the outer loop (over all ponds) until there are no changes in W_p.[10]

In passing, we note that the latter part of the algorithm may be inefficient, but assert that the first part (which identified locations which are in ponds) is nearly optimal.

[8] Note that $H(\vec{r}') = H_S(\vec{r}')$ by construction.

[9] We could check to avoid duplicate insertions, but this is not obviously more efficient.

[10] This is done to within some tolerance which determines how flat we need the pond surfaces to be.

Chapter 9

Simple models can explain complex phenomena

In this chapter, we present the results of our modelling program. Some of these results have already been submitted for publication[20], and others will be shortly. We begin by describing the success of both analytical and CDS models in explaining the formation of travertine domes. We unify these two approaches by exploring the role of surface tension in these patterns. We use the analytical theory due to Pak Yuen Chan to predict θ_c, the angle at the point on the dome where capillary effects become significant, and verify these predictions using the CDS model. Finally, we compare the predictions of both models to field observations.

We also explain both qualitative and quantitative aspects of terrace formation. By applying a linear stability analysis to our analytic theory, we show that precipitation is unstable toward perturbations of all length scales. This demonstrates that a scale free landscape could arise from a simple depositional instability. Moreover, we use the CDS model to explore the fully non-linear regime. Our numerical models produce landscapes which bear striking qualitative resemblance to real hot springs, including features that correspond to individual facies. Furthermore, the CDS model predicts dynamical aspects in pond growth and aggregation that are also seen in our time-lapse

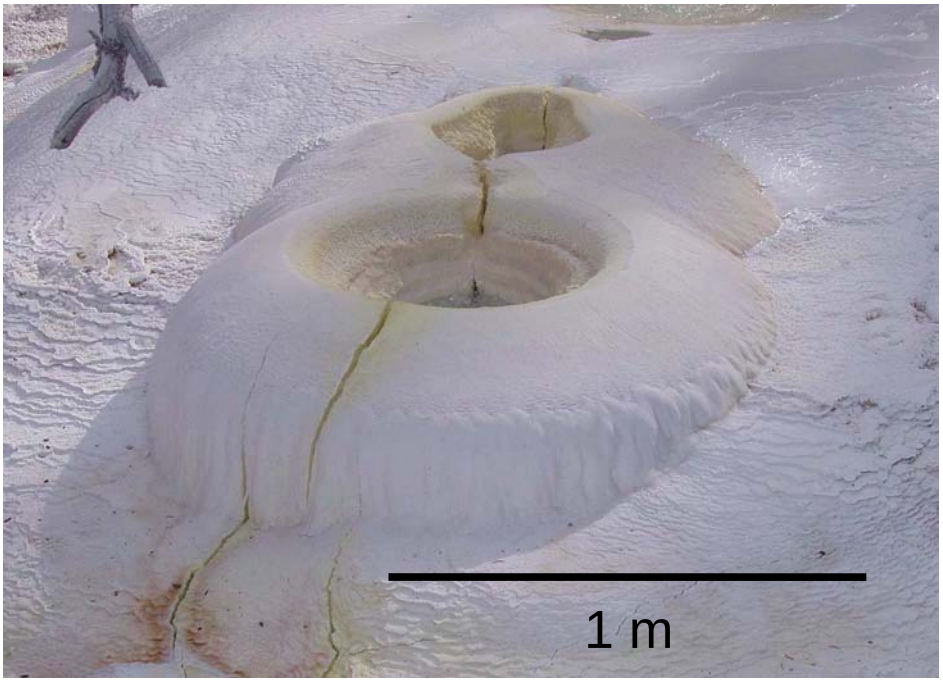

Figure 9.1: A circularly symmetric travertine *dome*. Photo by Nigel Goldenfeld.

movies. Finally, we conclude with quantitative measurements of the simulated landscapes. Both our results and limited experimental data indicate that some statistical measures of the landscapes, such as the pond size distribution function, are fractal. Future field work and extended simulation studies will further explore the statistical properties of terraced landscapes, but that is outside of the scope of this thesis.

9.1 Travertine Domes

We begin by analyzing the dynamics of pattern motifs, ignoring interactions between them, by analogy with earlier work on solidification patterns[186], or the monopole configurations that characterize the ordering defects during the isotropic-nematic transition in liquid crystals[6]. Figures 9.1 and 5.6 show pictures of a circularly symmetric carbonate feature which we have termed a *dome*. These domes arise from point sources, occur at low fluxes of spring water, are somewhat uncommon, grow to roughly 1-2 meters in diameter, and only reach about a half meter high. The short height follows

from the requirement of low flux; for the total flux to be small, the pressure head at the spring source must also be small.

The symmetry of this motif reduces it to a one dimensional problem, rendering it exactly solvable analytically. The next subsection summarizes Pak Yuen Chan's exact analytical solution for the dome[20].

9.1.1 Analytical Results

We formulate a boundary-layer model[186] of the growth of travertine domes, coupling the evolution of the travertine substrate to the fluid dynamics in a thin film around it. The kinematic equation governing time evolution of the curvature, κ, of a curve in two dimensions is given by[186, 187]:

$$\left.\frac{\partial \kappa}{\partial t}\right|_{\theta} = -\kappa^2 \left(1 + \frac{\partial}{\partial \theta^2}\right) v_n, \tag{9.1}$$

where θ is the angle between the surface and the vertical axis and v_n is the normal growth velocity of the surface. The time derivative in the equation is defined with respect to fixed θ. This equation is purely geometrical; for any given function v_n of water chemistry, surface kinetics, and fluid flow state, the evolution of κ is determined. Here we follow Wooding[188] and make the simple assumption that v_n is directly proportional to the depth-averaged tangential fluid velocity U: $v_n = GU$ where G is a mass transfer coefficient, depending on the water chemistry and the turbulent flow near the growing surface[189].

In general, we need to couple the Navier-Stokes equation to Eqn. 9.1 to obtain a complete description of the coupled fluid dynamics and surface kinematics, but there are a number of simplifications. First, because the growth rate is of order $1-5$mm/day and the fluid flow rate is of order 1mm/sec, there is a separation of time scales. So if we are interested in the morphological evolution, we can neglect the change in flow rate, represented by the time derivative in the Navier-Stokes equation. Second, our field

observations indicate that the thickness of the fluid film flowing over the domes is very small compared to the curvature of the surface; thus, we make the approximation that the fluid is flowing down a (locally) constant slope. In addition, the flow is apparently laminar, so that we can use the Poiseuille-Hagen profile for the velocity in thin film to give a depth-averaged mean flow velocity U:

$$U = \left(\frac{\alpha \sin\theta}{r^2} \right)^{1/3}, \tag{9.2}$$

where $\alpha \equiv gQ^2/12\pi\nu$, g is the gravitational acceleration, Q is the total mass flux coming out of the vent, ν is the viscosity of the fluid and r is the axial distance from the vent. Circular symmetry is imposed to arrive at Eqn. 9.2. We will later see that the assumption of laminar flow is self-consistently verified. For a dome, steadily translating upwards without change of shape with velocity v_t, Eqn. 9.1 gives

$$G \left(\frac{\alpha \sin\theta}{r^2} \right)^{1/3} = v_t \cos\theta, \tag{9.3}$$

Rearranging terms gives the shape of the dome as a one-parameter family of curves

$$r(\theta)/r_0 = \sqrt{\frac{\sin\theta}{\cos^3\theta}}. \tag{9.4}$$

where the scale factor $r_0 \equiv \sqrt{G^3\alpha/v_t^3}$. Eqn. 9.4 is plotted in Fig. 9.2. Good agreement is obtained between our theory and the observations below a critical angle θ_c. From the fit, and the typical parameter values $G \sim 10^{-8}$, $v_t \sim 1$mm/day and $Q \sim 1\text{cm}^3$/sec, we obtain $U \sim 25$mm/sec and $h \sim 1 - 10$mm, and a Reynold's number, $\text{Re} \equiv Uh/\nu \sim 10 - 100$. The assumption of laminar flow is self-consistently verified.

For angles $\theta > \theta_c$, the theoretical profile does not fit the field observations; the point of departure closely follows the point where we also observe a fluting pattern around the dome. We will show below that this is due to the effects of surface tension at the air-water-travertine interface.

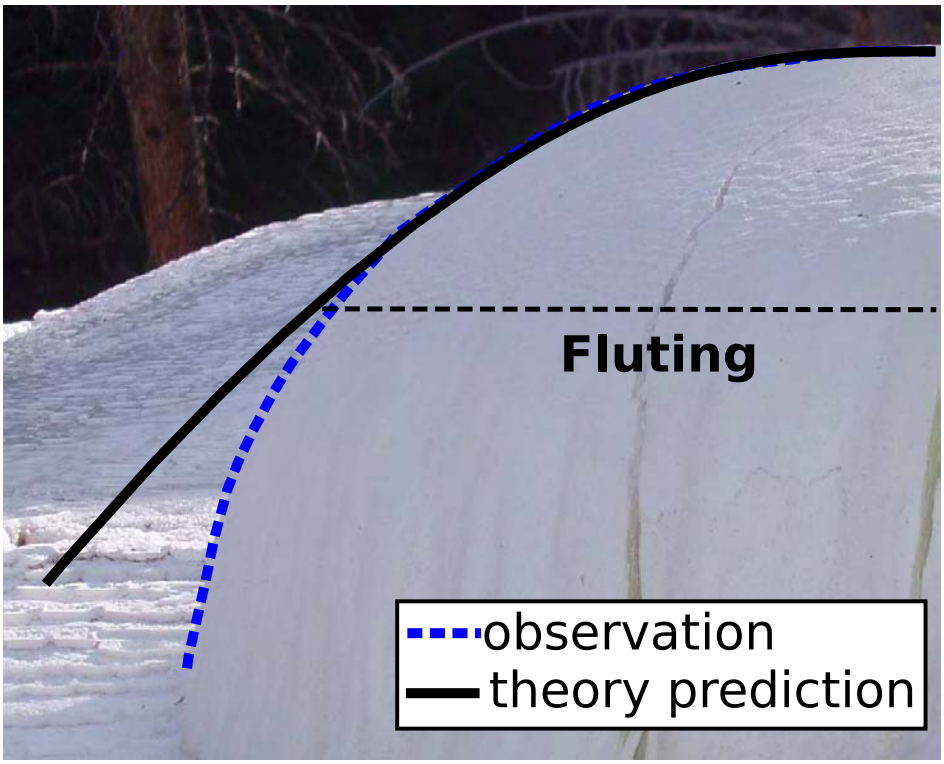

Figure 9.2: Dome profile compared with theory. The black curve is the analytical prediction from Eqn. 9.4. The blue dashed line traces the profile of the dome shown in the photograph. Photograph by Nigel Goldenfeld.

9.1.2 The Role of Surface Tension

As the water flows out of the dome, it is spread over an increasingly larger area and thus the fluid thickness decreases, ultimately reaching a point where contact lines form and surface tension cannot be ignored. The above theory neglects surface tension, and so is not valid beyond that point. This effect is clearly seen in all domes, and also near the lips of some ponds (Fig. 9.3). The signature fluting pattern begins exactly where Eqn. 9.4 breaks down in Fig. 9.2. We verify this explanation for the breakdown of the analytic theory by using the CDS models; these models show that a contact line develops exactly at $\theta = \theta_c$.

Figure 9.3: The lefthand photograph (by Nigel Goldenfeld) shows a closer view of the beginning of the "fluting" pattern seen on travertine domes. The righthand one (by John Veysey) shows a similar pattern sometimes seen on the lips and steep faces of ponds. Both phenomena arise due to capillary effects in very thin films.

9.1.3 Simulated Domes

When we begin the 3D CDS model from an initially horizontal surface, with or without small random noise, the result is a circularly symmetric dome, as shown in Fig. 9.4. The CDS model agrees quantitatively with the observed dome, and allows us to explicitly see the importance of surface tension in the overall dome shape. Figure 9.5 shows the results of the CDS model in comparison to the analytical prediction and the observed travertine dome. We see that the CDS model agrees with the analytical theory when the surface tension is switched off. Moreover, the same CDS model predicts the exact shape of the observed travertine domes when surface tension is switched on, *and* predicts the formation of a contact line at θ_c, where fluting emerges.

9.1.4 θ_c: Combining analytical and numerical results

Our analytic theory allows us to predict the scaling dependence of the critical angle on the model parameters. The inclusion of surface tension introduces an additional length scale, namely, the capillary length, d_c, into the problem. Now, the only other

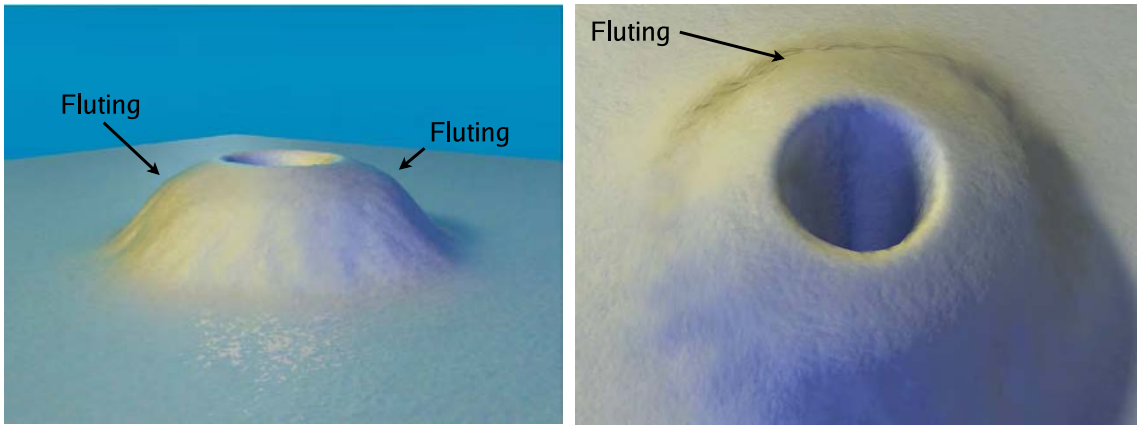

Figure 9.4: (left) Side profile of a simulated dome, showing the onset of "fluting" effects at $\theta = \theta_c$. (right) Overhead shot of the same dome. Rendering done with Blender[190] and the invaluable assistance of Nicholas Guttenberg.

length scale in the problem is $r_0 = \sqrt{gG^3Q^2/\nu v_t^3}$. Since θ_c is dimensionless, it can only depend on the ratio r_0/d_c and G. For a given chemical environment, G is fixed and we are left with the prediction, derived from our analytical solution, that

$$\theta_c = \hat{f}(\sqrt{(gQ^2/\nu v_t^3)}/d_c), \tag{9.5}$$

In this equation, there are two free parameters: d_c and Q. v_t is implicitly determined by the chemistry G and the details of the fluid flow. Within the CDS model, we can vary both d_c and Q. Since the radius at which the contact line forms is readily identified in the simulated water profile, we can measure θ_c, and thereby test the data collapse predicted by theory. Figure 9.6 shows that the scaling prediction Eq. (9.5) is verified over 5 decades of $Q^2v_t^3/d_c^2$. We ran the simulation for $0.1 < Q < 5000$. In these simulations, $v_t \sim 0.5$.

This fact that the simulated domes obey the scaling law (Eqn. 9.5) predicted by the analytic theory indicates that both of our models are capturing the same, essential physics needed to understand the formation of domes. This suggests that these patterns could be entirely understood through a simple analytic theory, if surface tension could be added to our description of fluid flow. While we would like to verify the predicted data collapse experimentally, domes are rare enough at hot springs (especially actively

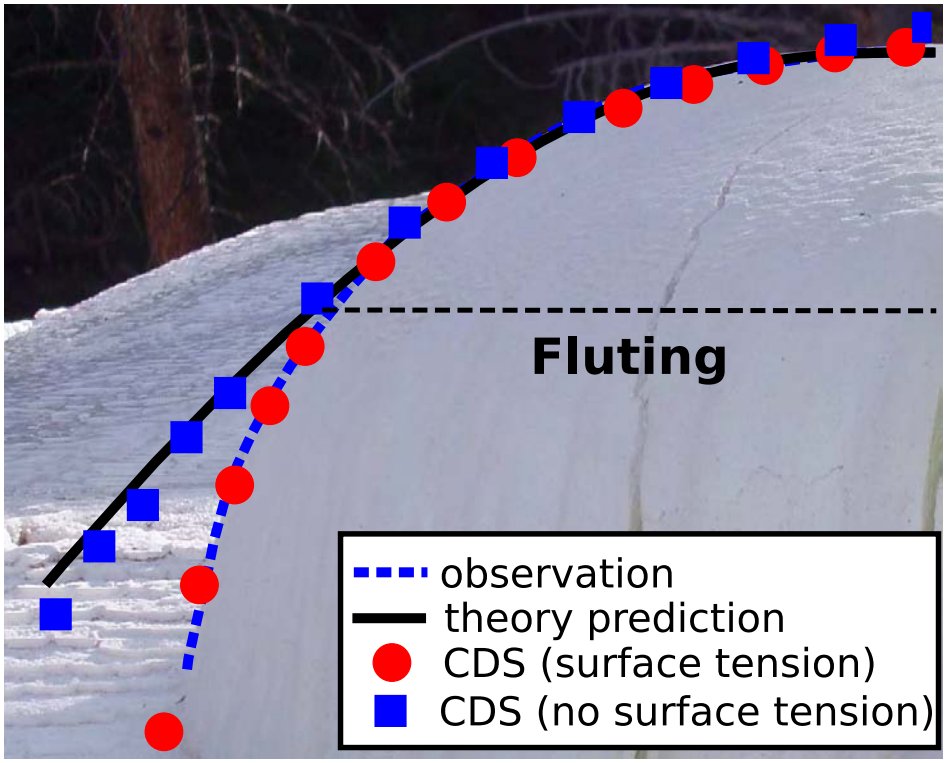

Figure 9.5: Dome profile compared with theory and simulation. The red filled circles show the profile of a simulated dome, including the effects of surface tension. The blue filled squares show the profile of a simulated dome without surface tension.

forming domes, where we could measure flux) that this has so far proved impossible. However, in our sampling of 3 domes, no significant departures from the profile shown in Fig. 9.5 have been observed.

9.2 Terraces

This section explains the ubiquitous growth of ponds and terraces at carbonate hot springs. We begin by using a linear stability analysis to prove that turbulent flow of saturated spring water results in depositional instabilities at all wavelengths. This work, which is presented in the next section, is due to Pak Yuen Chan. It is presented here because it complements my numerical studies, which extend it into the fully non-linear regime.

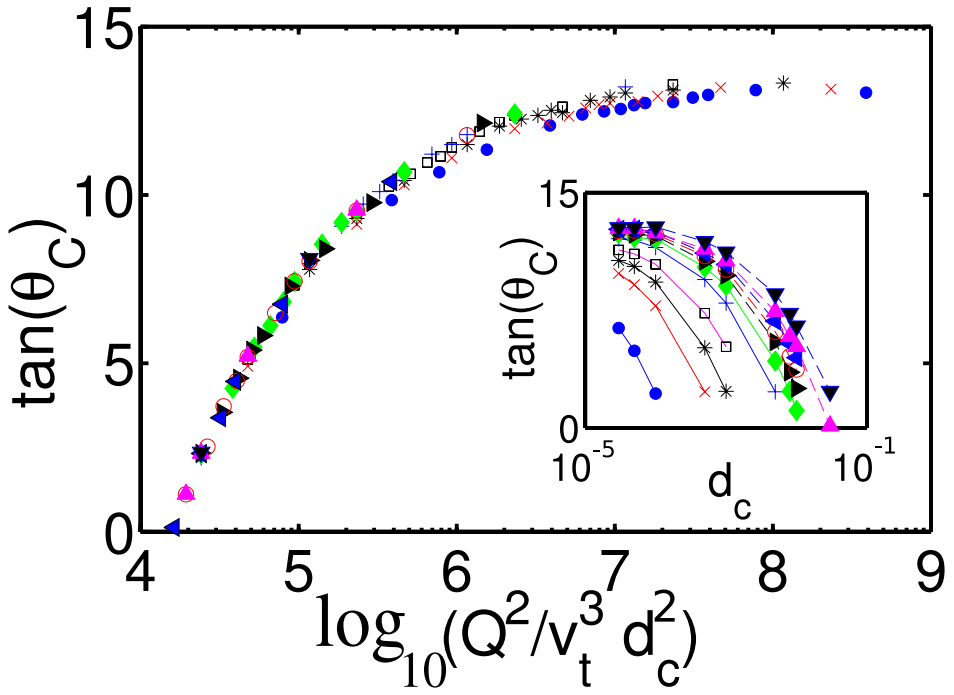

Figure 9.6: The critical angle at the contact line formation on a travertine dome, plotted according to Eqn. 9.5, showing data collapse as predicted by theory. The inset shows the raw data; each line (and color) shows a value of Q, and the plot shows θ_c as a function of d_c.

9.2.1 The damming instability

To understand the scale free structure of terraces, we consider the stability of the moving boundary problem for turbulent fluid flowing down a constant slope, on which deposition may occur. In order to capture the turbulent flow, we make two approximations. First, we use the thin film approximation attributed to de St. Venant, which is valid when the fluid film thickness is much less than the characteristic scale of variation of the flow in the streamwise direction, but include the lowest order corrections for the curvature of the underlying surface[191, 192].

$$\begin{aligned} \partial_t u_0 + \partial_s E &= -C_f g u^2/gh(1-\kappa h) \\ (1-\kappa h)\partial_t h - \partial_s q &= 0 \end{aligned} \tag{9.6}$$

where

$$\begin{aligned} u(s,n,t) &= u_0(s,t)/(1-\kappa h), \\ E(s,t) &= \zeta + h\cos\theta + u_0^2/(2g(1-\kappa h)^2), \\ q(s,t) &= -u_0/(\kappa \log(1-\kappa h)), \end{aligned} \tag{9.7}$$

where u is the fluid velocity, h is the fluid thickness, ζ is the height of the underlying surface measured from a fixed horizontal axis and s is the arc length measured from the top of the system. Secondly, we have modeled turbulent flow phenomenology by including the term proportional to C_f, the Chézy coefficient[193], which empirically describes the energy lost due to turbulence, in a manner consistent with Kolmogorov's 1941 scaling theory of turbulence (K41)[194, 195]. These two equations have to be solved together with the growth equation (9.1). The trivial solution to this set of equations can be easily found, and is simply uniform viscous flow down a slope.

To study the linear stability of this solution, we add a perturbation proportional to $\exp(ikx + \omega(k)t)$, and calculate the spectrum of the growth rate ω as a function of the wavenumber k for the linearized set of equations.

It is found that $\mathrm{Re}(\omega)$ is positive for all values of k, indicating that the system is unstable toward perturbations of all length scales. Although this calculation neglects the effects of non-linearities, we interpret this result as the origin of the fact that the terraced architecture is apparently scale-invariant.

9.2.2 Simulated terraces

When we run our CDS models, we see that the depositional instabilities identified above grow to form dams. Pond water pools behind the dams, and we see complicated interactions as terraces grow and interact. This section discusses the CDS results, beginning by presenting the striking resemblance between real hot springs and our simulated landscapes.

Figure 9.7 shows a snapshot from a typical time-dependent simulation, initiated on a

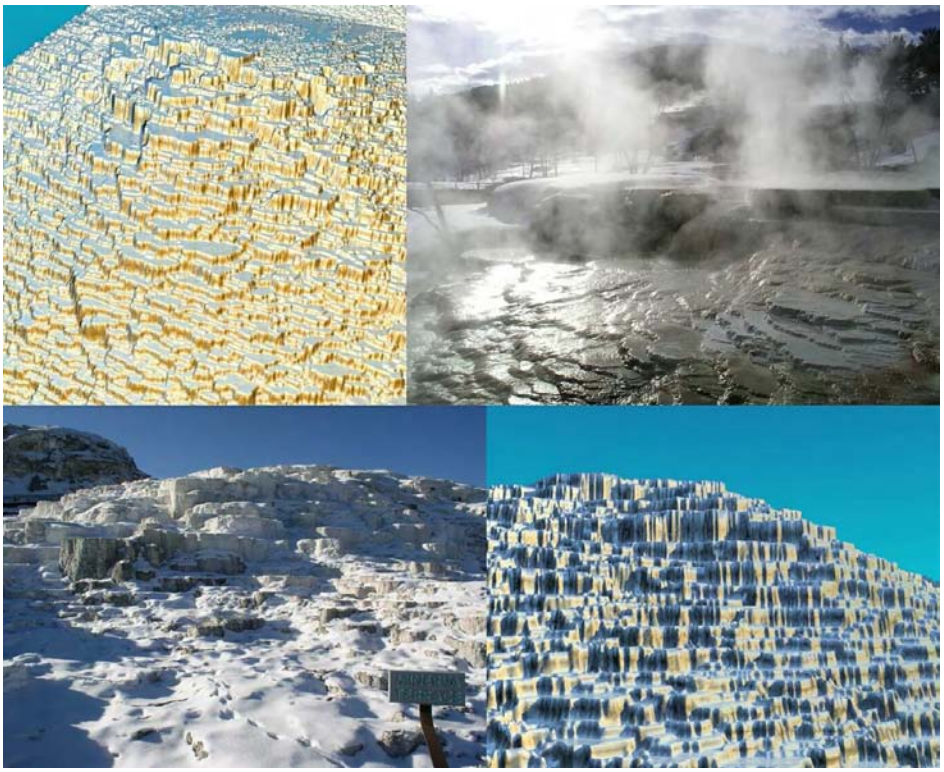

Figure 9.7: This photo compares two renderings of simulated terraces with two pictures from YNP. The upper right and lower left images are real photos, from AT-3 and Minerva Terrace, respectively. The simulated results were rendered by Nicholas Guttenberg from John Veysey's simulations. The upper right photo was taken by Roy Johnson, the lower left by the Goldenfeld group. Layout by Nigel Goldenfeld.

sloping plane with small initial roughness. For generic values of the model parameters, we observe the depositional instability predicted above, and the formation of ponds and terraces in broad qualitative agreement with field observations. We emphasize that this is produced by the *exact* same CA rules which led to the dome simulations discussed in the first part of this chapter. The differences lie in initial conditions, flux, and surface tension.[1]

The simulations allow us to examine more than just landscape morphology. We can compare water depths, flux, and our simple chemical model to the real system. Figure 9.8 shows an example of this sort of prediction, showing temperature and saturation

[1]The last two parameters actually only effect simulations according to the combined variable Q/d_c, as shown in Eqn. 9.5.

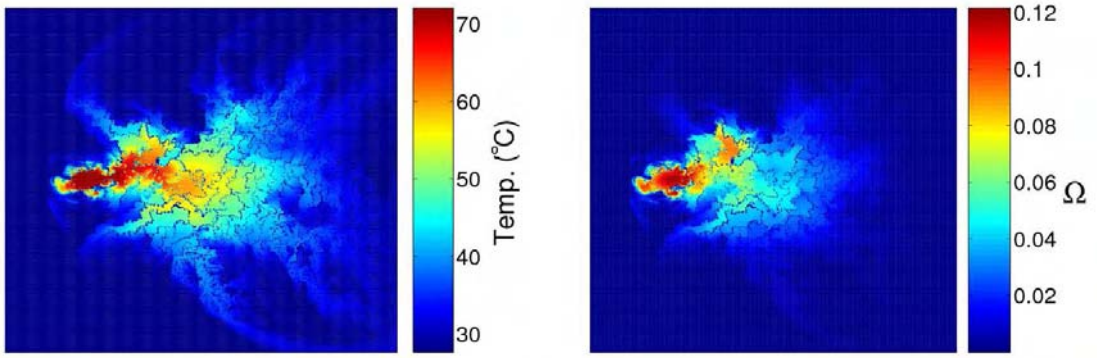

Figure 9.8: (left) "Temperature" profile of a simulated hot spring. (right) "Saturation state" (Ω, defined as in section 8.3.4) for the same simulation.

state for a simulated hot spring. This particular snapshot is taken from late in the simulation, near the end of spring water flow. As a result, $\Omega > 0$; precipitation is even occurring at the source! Because the flux is so small, and because of mixing, this can occur even though the initial conditions specify $\Omega_0 < 0$. The synthetic temperature profile can be compared to the real data in Figure 5.11. While there are similarities, this comparison does not result in new qualitative insights. In future work, we will subject our simulated hot springs to the same analysis detailed in chapter 6, identifying individual facies in the artificial spring and looking at the resulting distributions of temperature, C, and V. It is hoped that the CDS results exhibit the same sort of correlations seen between hot spring water and facies in the real system.

A qualitative analysis of the simulations allows us to identify both macroscopic patterns and dynamical processes which are also seen in the real system. First, as previously mentioned, we see that ponds form initially in areas with low flux. In addition to the pond morphology, which is easily mapped to the pond facies, we see features which correspond to the vent and apron channel facies. Figure 9.9 shows simulation results which closely resemble these two facies in real springs.

Near the spring water source, we see a pool of water (this can be more clearly seen in the overhead view shown in the upper left of Figure 9.7). In addition to harboring

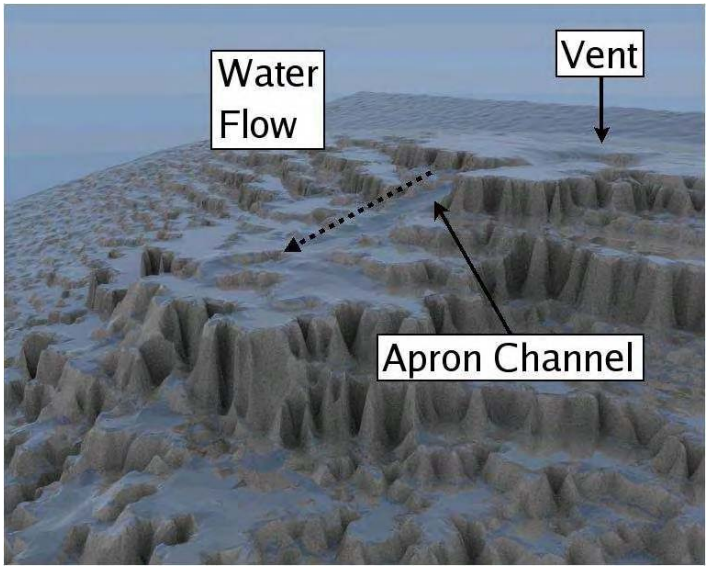

Figure 9.9: Rendering of a typical artificial hot spring, showing the formation of a "vent" pool near the spring water source, and a narrow high flux channel located immediately downstream from the pool.

the spring vent, this pool is characterized by high temperatures and $\Omega \lesssim 0$. In fact, simulations confirm that a pool is an inevitable consequence of source water which is highly charged with CO_2. As the water emerges, it can dissolve carbonate, and precipitation cannot begin until sufficient outgassing has occurred. Because the flow rate in the vent pool is largely determined by the flux Q, and CO_2 outgassing occurs at the same rate (for fixed geometry and Q), Ω will not be positive (and precipitation will not occur) until the water has travelled a fixed distance — the radius of the vent pool. If correct, this hypothesis implies that springs which have the same flux and emerge onto similar initial conditions will all have a the same vent pool radius, and that spring flux could be estimated from the size of the vent pool. Although field observations do not disprove this hypothesis, we do not have enough data (e.g. springs) to make quantitative comparisons.

Immediately downstream from the "vent pool," Figure 9.9 shows a narrow channel through which nearly all of the spring flux flows. As in the real spring, $\Omega > 0$ in this channel. Below the channel, water spreads out, draining into a network of ponds. Two

interesting dynamical features are associated with the artificial apron channels. First, in mature springs (e.g., after a sufficient number of time steps in the simulation), it often terminates in a large pond. This may be an inevitable consequence of pond inundation and aggregation, an observed process which we will explain in greater detail below. In spring AT-1, a mature hot spring which has dried up since we began our observations, the apron channel was terminated by a large pond, leading to speculation that this is a fundamental attribute of carbonate hot springs[103]. Secondly, simulations indicate that the apron channel is intrinsically transitory, and is associated only with large fluxes. At long times, as the spring water flux diminishes due to carbonate buildup, the apron channel is inevitably subsumed by ponds and buried by precipitation.

9.3 Time-Lapse Photography of Canary Springs

To help us understand time dependent processes at carbonate hot springs, we have undertaken a time-lapse photography project at Canary Springs (Fig. 5.4). For the past two years, YNP ranger Brian Suderman has been providing invaluable assistance by taking still camera photos at carefully chosen location, which we have subsequently processed to produce movies of travertine growth. These movies are the basis for the travertine growth "prediction" shown in Fig. 5.1. The details of these measurements and techniques are given in Appendix 9.6. By studying the resulting series of images, we have identified two dynamical processes which arise during pond formation and interaction. These processes were first seen in our artificial landscapes, prior to the correlation of the time-lapse movie.

Figure 9.10 shows a series of four stills from our time-lapse movie at Canary Springs. They show the process of pond *inundation*, or "drowning," in which faster growing downstream ponds merge with their upstream neighbors to form a single large pond. This mechanism may be the only one through which larger ponds are formed. It also

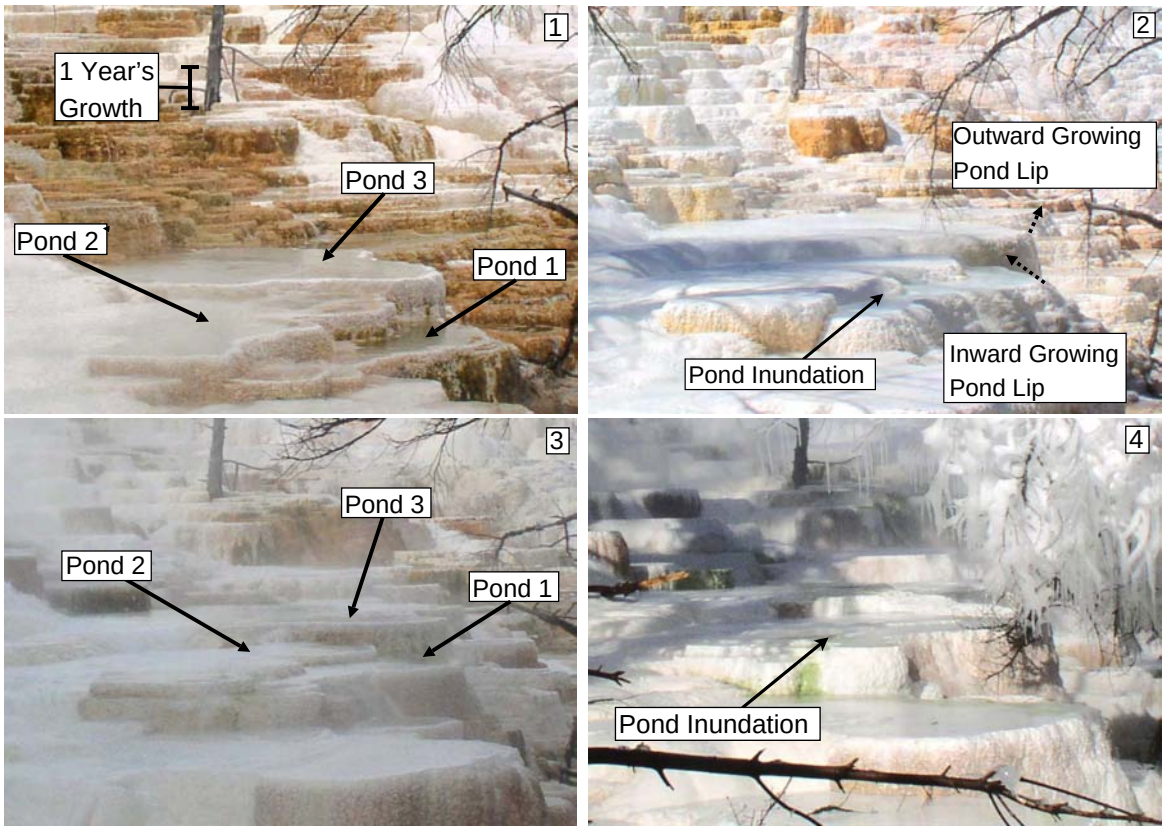

Figure 9.10: Four still frames from our time-lapse movie at Canary Springs. Beginning in the top left, we see the growth and interaction of two ponds. Three specific processes are illustrated. First, we see that faster growing downstream ponds "inundate" upstream ones to form a single large pond. Secondly, we see that pond lips grow both into and out of the flow, possibly depending on the flux in the pond. Spring water flows from left to right in the photos. Photos by Brian Suderman.

shows pond lips growing both into and along with the flow, an observation which constrains theories of lip growth. This dichotomy may be due to differing fluxes within the ponds. Simulations confirm that outward growth can arise in high flow areas, but we have not yet found inward pond lip growth, which may be a more subtle effect due to surface tension.

Figure 9.11 shows two of the same dynamical processes of pond inundation and outward lip growth, but as reproduced in our simulated hot springs. Examination of frame # 2 also shows how the apron channel seen in Fig. 9.9 disappears at longer times. Our results show both that we have captured the depositional processes important for

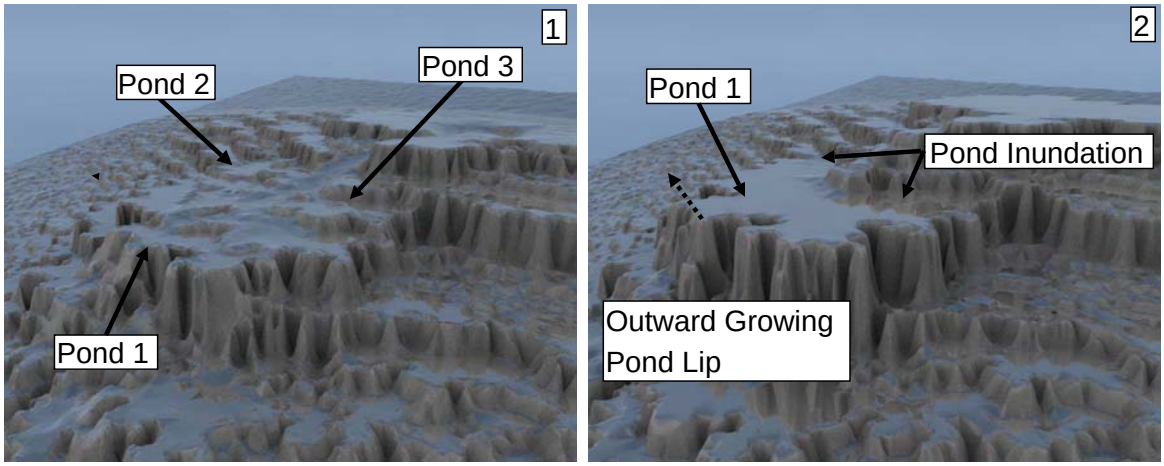

Figure 9.11: Still frames rendered from hot spring simulations. Both the inundation and outward pond lip growth seen in Figure 9.10 are also seen in the simulated landscapes.

terrace formation and that there are non-trivial dynamical process which arise from the interaction of simple motifs.

The distal slope and proximal slope facies have been absent from this qualitative discussion of simulation results. In the previous chapter, we demonstrated that the proximal slope facies is distinguished from the pond facies on the basis of flux and flow velocity. Simulations do produce landscapes which have features which correspond to the proximal slope. Macroscopically, these regions have higher flux, steeper slopes, and are usually adjacent to ponds. However, they necessarily disappear at long times. As the landscape grows, the flux at the vent drops, ponds form where there was once a "proximal slope". Like the apron channel, the proximal slope facies appears to be transitory, and is not present when the spring dries up. The distal slope is not clearly reproduced in simulations, although we have some evidence that it can be identified on the basis of elongated, step-like terraces. These have a different aspect ratio than pools in the pond facies. We would like to confirm this hypothesis numerically and experimentally, but do not yet have sufficient results for inclusion in this thesis.

9.4 Statistical Analysis of Simulated Landscapes

Ultimately, proving our model's success requires quantitative comparison with real systems. We have presented such evidence for travertine domes. In this section, we present some of the statistical properties of simulated terraced landscapes. Unfortunately, it has been hard to acquire sufficient data from real spring systems for comparisons. Our access to the springs is limited, and we have not been able to conduct aerial photography, lasergrammetry, or highly detailed surveys.[2] We have tried to make 3D measurements of hot spring topography using *photogrammetry*, which relies on multiple photographs taken from camera positions whose relative positions are known to a high degree of accuracy. Despite careful measurements and modelling of the camera's optics, we were unable to make measurements more accurate than $0.5m$, which is insufficient. Even if these efforts had succeeded, it is not clear that we could have surveyed enough hot springs to amass enough data to characterize hot spring landscapes in a statistically significant fashion.

The closest thing we have to aerial photography is a shot of small centimeter scale terraces from about two meters high. Figure 9.12 shows an example of such patterns, along with a careful tracing of all the lips in the picture due to Pak Yuen Chan. We can compute the statistical properties of the ponds in the traced image, and then compare them with simulation results. While this only captures the smallest ponds, and none of the actual large scale morphology which dominates the landscape, it nonetheless provides a basis for initial quantitative comparisons between simulation results and real landscapes. Although these comparisons are limited by the small number of samples (one photograph, which contains 1686 separate "ponds"), they agree surprisingly well with the simulation predictions.

[2]We have considered kite photography as an alternative to fly-overs. It is not obvious whether this would be acceptable to the park service!

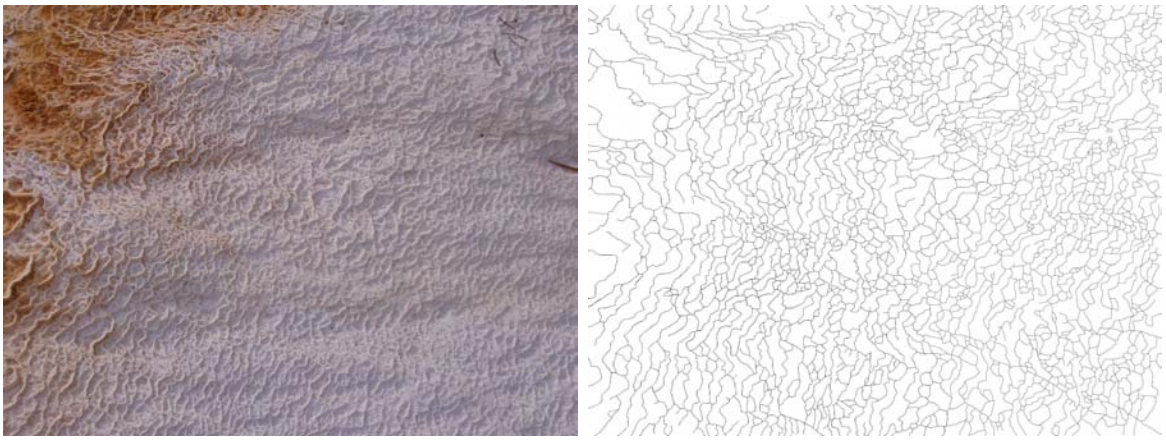

Figure 9.12: (left) Photograph of small terraces at Spring AT-3. (right) Pak Yuen Chan's painstaking tracing of the lips in the photo. We digitize this tracing and compare it with simulation results.

9.4.1 Distribution of Pond Areas

We begin by looking at the pond size distribution function. Figure 9.13 compares simulation results, based on 130 simulations of 600x500 cells and 28,000 time steps, with the photograph in Figure 9.12. To estimate the errorbars and to better calculate the distribution function, we used the *bootstrap* method[196]. In short, we resampled our original set of pond area measurments with replacement, to obtain a synthetic data set of with the same number of measurements. This new data set contains duplicate measurements (hence the requirement to sample with replacement). We then calculated the pond area distribution function based on the synthetic data set. We repeated this process 1,000 times for the simulation results and 10,000 times for the smaller data set which came from the digitized photograph. Looking at the distribution of each point on the resulting histograms, we then calculated the mean and standard deviation; these are our best estimate of the pond distribution function and the error in that estimate. All subsequent distributions were calculated in the same manner. We also note that any bins with less than three counts were considered to have insufficient data, and are not plotted.

The log-log plot in Figure 9.13 shows that the probability of finding a pond of a

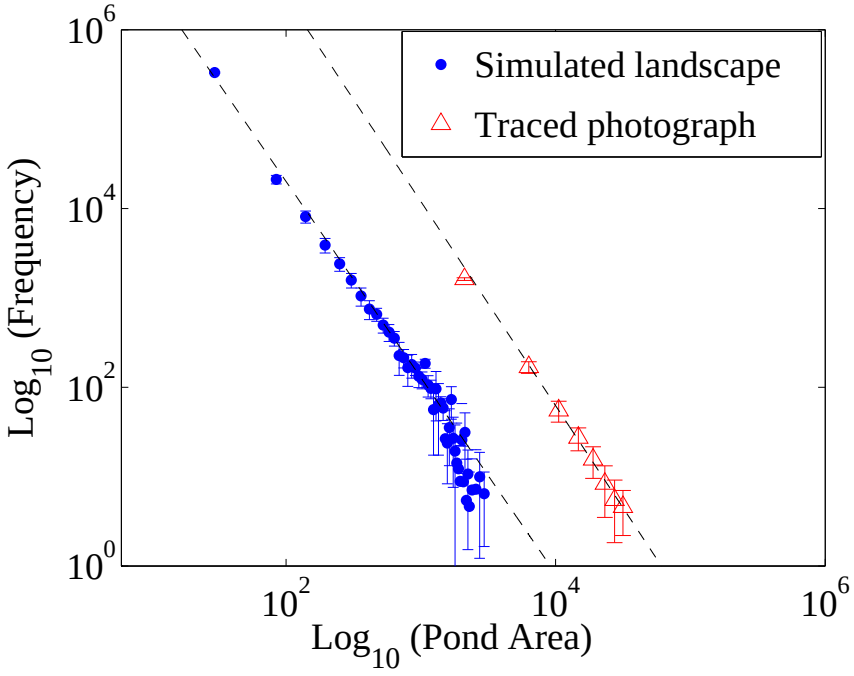

Figure 9.13: Log-Log plot showing the frequency of occurrence of a given pond area. Data comes from 130 simulated landscapes each of which began from a sloping plane with different random noise added. The simulations were 600x500 cells each, and ran approximately 30,000 time steps. Best fit lines are shown.

given area (measured in pixels) follows a power law, $P(A) \propto A^{-b}$, and suggests that ponds in simulated landscapes are fractal. The power law seen in simulation results ($b = -2.20 \pm 0.08$) is consistent with that seen in the limited data from the real system ($b = -2.29 \pm 0.54$). This agreement is satisfying, but requires further confirmation because the real data is limited in quantity, extent (only covers pond lips in a tiny area of a single spring), and because the observed power laws extend only for, at best, two decades. Nonetheless, this is a promising result which justifies future investigations, represents a preliminary validation of the CDS models, and suggests that the same processes control the formation of both small and large terraces. As it does not yield any insight into the shape of terraces, spatial variations in terrace shape and size, or temporal changes in macroscopic patterns, we turn now to more sophisticated methods of analyzing landscapes.

9.4.2 Calculating the Terrace Width Distribution

To examine terrace patterns in more detail, we draw on two previous ideas. First, height-height correlation functions have been used to describe the statistical properties of eroding landscapes[11, 197]. However, this approach relies on calculating two 1D correlation functions, along vectors parallel and perpendicular to the eroding slope. This cannot be easily applied to the terrace simulations shown in Fig. 9.7, because there is a radial symmetry imposed by our point source. Apart from the one line of symmetry through the source and parallel to the slope, we cannot compute sensible 1D height-height correlation functions along parallel and perpendicular directions. The approach of Pastor-Satorras and Rothman would be appropriate to characterize terraces which form on a slope with a *line* source. We hope to do this in future work.

Studies of the dynamics of step patterns on vicinal surfaces[198] have led to the idea of a *terrace-width distribution*[199, 200, 201]. Step like patterns arise in materials like Si(111), in which crystals are misoriented from a high-symmetry plane[201]. As they form, these terraces repel each other, giving rise to complicated interactions. Measuring the distribution of the step sizes allows determination of the dimensionless strength of the elastic repulsion between steps[199]. The physical resemblance between these terraces and those in our system, particularly those in the distal slope facies, inspired us to calculate terrace width distributions (TWDs) for our simulated hot springs, in the hopes that they would allow us to parametrize the governing dynamics.

Unfortunately, terraces in silicon have two fundamental differences from our system. First, the edges of their terraces do not cross or merge. Secondly, they do not have a preferred orientation; the width in silicon TWDs is always measured parallel to the slope. As with the height-height correlation functions, the radial symmetry imposed by the source renders this impossible. To resolve this problem, we make our measurements by defining the *maximum chord* in the pond. This is the longest chord which begins and ends on the edge a pond, as shown by C_1 in Fig. 9.14. This definition reduces

to the terrace width in vicinal surfaces in both an average sense and in the limit of infinitely long steps. All measurements of pond widths are then made perpendicular to C_1. In particular, we also define the largest width, C_2, which can be useful for some measurements of pond aspect ratio. Finally, we define the center of the pond as the intersection of C_1 and C_2. We later use this point to calculate the distance from the vent at which the pond is located.

We also find it useful to define and measure the aspect ratio of a terrace, roughly equal to C_2/C_1. Experimentally, it is difficult to measure aspect ratio more precisely, but in simulations, we can define it as $\alpha \equiv \int \mathrm{d}s\lambda(s)/C_1 = A/C_1$, where s is measured along C_1, and A is the area of the pond. In practice, we can compute α either by using the pond area or by performing the numerical integration. We choose the latter because we can perform subpixel interpolation along the edge of the pond, thereby obtaining a better estimate of both A and α.

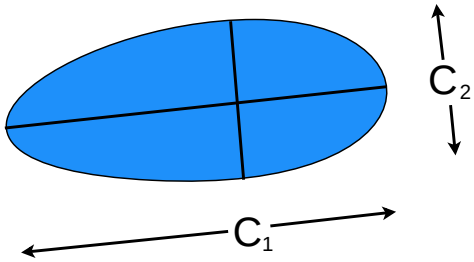

Figure 9.14: Schematic depicting the maximum chord in a pond (C_1), as well as the largest perpendicular chord, C_2.

TWDs are usually characterized by the probability of finding a particular terrace width, $P(s)$. Where $s = \lambda/\langle\lambda\rangle$, λ is the actual terrace width, and $\langle\lambda\rangle$ is the average of λ. For the purposes of our system, we calculate $\langle\lambda\rangle$ separately for each terrace or elongated pond. This differs from the definition in vicinal surface problems, where $\langle\lambda\rangle$ is averaged over all terraces. By construction, $P(s)$ will be normalized and have unit mean.

Figure 9.15 shows TWDs measured for both digitized photos and CDS simulations. The errors in the frequency distribution were again computed with the bootstrap method. This uncertainty was propagated in the usual manner as we calculated the normalized distribution $P(s)$, a step needed to compare the two distributions. Note

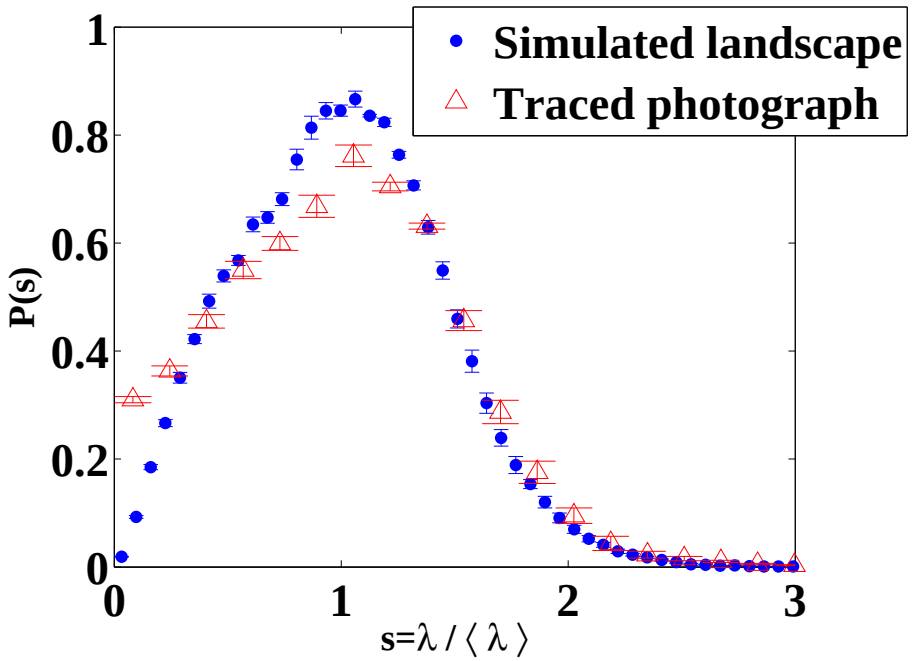

Figure 9.15: The bootstrapped terrace width distribution for simulations (8,000 resamplings) and experimental observations (10,000 resamplings). The vertical errorbars are shown; the width of the top and bottom of those errorbars represents horizontal uncertainty.

that there is also a horizontal uncertainty, due to the bin size, which is represented by the width of the top and bottom of the errorbars. We see a TWD which rises sharply and tails off more slowly. The form of this curve is roughly similar to distributions seen in vicinal surfaces, but differs at smaller s. The smallest data point in the experimental curve deviate from the simulation predictions, an effect likely due to our inability to resolve very small details in the traced terrace edges (Fig. 9.12).

In making this plot, we have amalgamated data from all parts of our simulated springs, and glossed over an important detail: in principle, $P(s) = P(s, r)$, and depends on the radius from the origin, r. As shown in the previous section, there are important qualitative differences between different parts of the spring, differences which can be associated with different pond sizes and shapes; λ and $\langle\lambda\rangle$ both depend on r. But Figure 9.15 shows that all of the data from our system falls onto the same curve when

plotted as a function of s. Because we see this data collapse for all of the simulated data, as well as for the photographic data (which was taken at one specific r), we conclude that s is independent of r[3], and that $P(s)$ represents a generic, facies independent property of terrace patterns at carbonate hot springs. This conclusion is supported by examining the bivariate distribution, $P(s, r)$. Moreover, the fact that all ponds have the same TWD is consistent with Figure 9.13: Scale invariant ponds must have the same TWD even if their absolute size, $\langle\lambda\rangle$, depends on r.

Workers studying vicinal surfaces have calculated $P(s)$ by mapping the problem onto a model of one-dimensional free fermions[202]. In this model, the edges of terraces only interact through steric repulsions. Joós et al. have numerically calculated $P(s)$, and derived analytic expressions for its asymptotic behavior. Figure 9.16 compares the TWD calculated from our simulation (Fig. 9.15) with their solutions to the free-fermion model. We see excellent agreement for $1.5 < s < 3.0$. We cannot truly evaluate our results at higher s because of a lack of data, and even if we had suitable statistics, we would eventually be limited by the finite system size in our models. At lower s the deviations imply that interactions in our system involve more than steric repulsions of terrace lips.

At $s > 1.5$, $P(s) = P_{\text{free fermion}}(s)$. This implies that any interactions between terrace lips are short range, and that at large s the lips do not interact. Consequently, in the large s limit, the TWD follows that of a random walking particle, and has a Gaussian tail (with logarithmic corrections)[202]. On a plot of $\log_{10} P(s)$ vs s^2, this asymptotic behavior appears as a straight line. Figure 9.16 shows explicitly that our TWD has this asymptotic behavior. Furthermore, the deviations from the free-fermion distribution at small s allow us to infer the nature of the effective interactions in our system. $P(s) > P_{\text{free fermion}}(s)$ for $s < 0.5$ (but greater for $0.5 < s < 1.0$) implies an effective attraction between the edges of terraces.

[3]This is possible because we defined $\langle\lambda\rangle$ locally, calculating it individually for each pond.

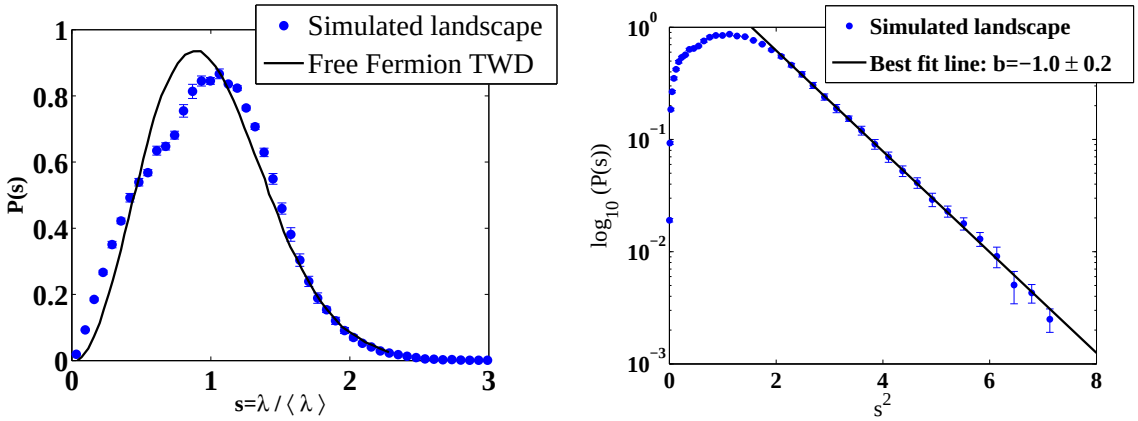

Figure 9.16: Comparison of our TWD with that predicted by a free fermion model. (left) Compares the TWD derived from our simulations with the form calculated by Joós et al. and digitized from their paper[202]. This plot shows agreement in the large s regime, implying that interactions in our system are short-range. (right) Shows that $P(s) \sim e^{-s^2}$ at large s. On a semi-log plot against s^2, this appears as a straight line.

9.4.3 Spatial Dependence of the Average Terrace Width

Figure 9.17 shows the spatial dependence $\langle\lambda\rangle$, and demonstrates that the data collapse seen in Figure 9.15 is non-trivial, rather than a consequence of λ and $\langle\lambda\rangle$ being independent of r. Figure 9.17 is complicated, and needs some interpretation. The most obvious feature is the number of small ponds in our simulations, a number which increases as r gets larger. This is not an artifact of small ponds which are formed due to initial randomness; we imposed a cutoff to avoid ponds with less than 10 pixels^2, as evidenced by the lack of any ponds on the far left of Fig. 9.17.

The spatial distribution of larger ponds is more interesting. Near $r = 0$ and $\langle\lambda\rangle = 20$, there is a cluster of large ponds which results from the pools which form at the vent in our simulation. As r increases from $r = 0$ to $r \sim 150$, the number of of large ponds also rises. Beyond $r \sim 150$, however, it begins to decrease, and essentially none are seen at $r \geq 400$, an area which corresponds to the distal slope facies. This is consistent with real hot springs, which do not have any large ponds in the distal slope. The trend in $\langle\lambda\rangle$ can be understood through changes in the saturation state. Ω initially rises due to exsolution, but reaches a peak as spring water gets depleted

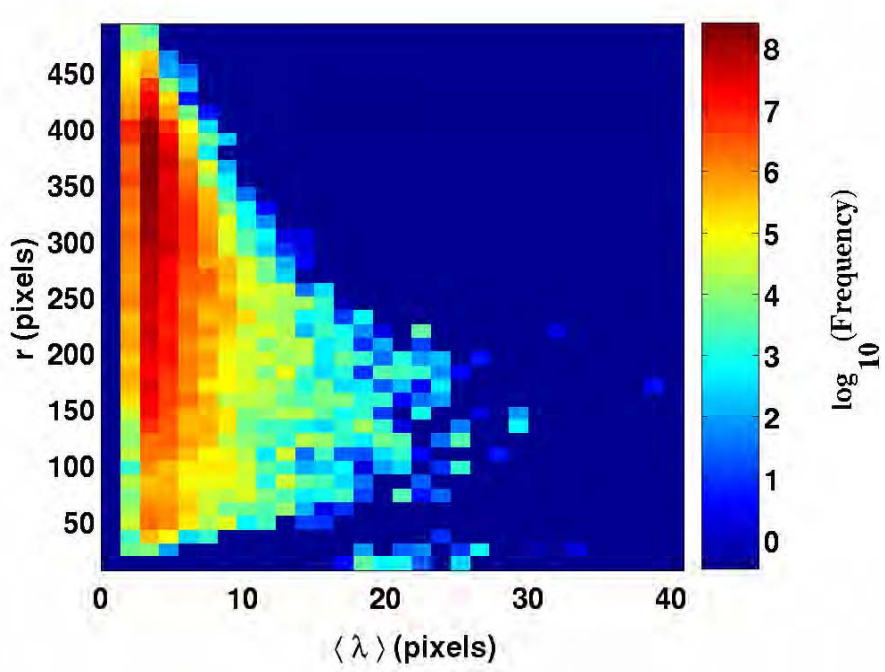

Figure 9.17: Dependence of $\langle\lambda\rangle$ on the distance from the source, r. Fewer large ponds occur as r increases. The contribution of vent pools can also be seen.

in both $[Ca^{2+}]$ and $[CO_2]$. The *average* precipitation rate follows this trend in Ω. With decreased growth rates beyond $r \sim 150$, the pond inundation and amalgamation described in the previous section occurs less frequently, resulting in fewer very large ponds. Unfortunately, our single photograph is too localized in r (e.g., it covers less than $1m^2$) to permit meaningful comparisons with these simulation predictions.

9.4.4 Pond Aspect Ratio

Examining the aspect ratio, α of ponds yields further insight into the origin of the data collapse in the TWD, the self-similar shape of ponds, and the differences between facies. Figure 9.18 shows the distribution of α for both the digitized photograph and simulated landscapes. We see that there is a characteristic pond aspect ratio ($\alpha \sim 0.35$), and that there are strong similarities between the experimental observations and the simulated landscapes. As with previous graphs, the observed distribution lacks sufficient data for

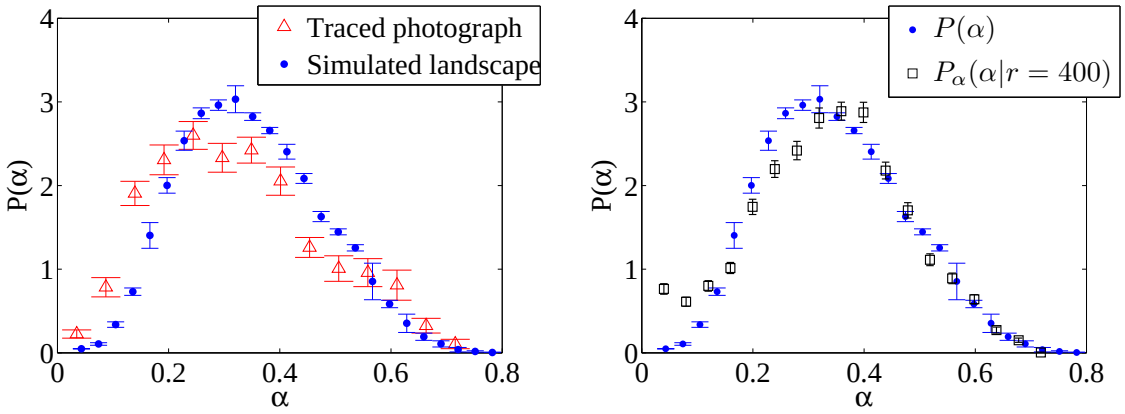

Figure 9.18: (left) The distribution of pond aspect ratios, comparing simulations with the digitized photograph. As before, horizontal uncertainty is shown by the width of the vertical errorbars. (right) α has a weak dependence on r. At large r, there is an increase in the number of long, narrow terraces, possibly associated with the beginning of the distal slope facies.

a compelling comparison.

The lefthand plot in Fig. 9.18 was made assuming that $P(\alpha)$ is independent of r. By examining the joint distribution $P(\alpha, r)$, we see that this is largely true, but that there are interesting exceptions. The righthand plot in Fig. 9.18 shows the conditional probability distribution for $P_\alpha(\alpha|r)$ at large r; it compares $P(\alpha)$ with $P_\alpha(\alpha|r = 400) \equiv \int_{395}^{405} P(\alpha, r)\mathrm{d}r$. Although the distributions are very similar, we see an increase in the number of wide, narrow terraces with small aspect ratio ($\alpha \lesssim 0.23$).

The fact that $P(\alpha)$ is independent of r for most r indicates that nearly terraces in our system have the same aspect ratio. This is consistent with a scale invariant pond shape and with a TWD which does not depend on r. The weak r dependence seen in simulation for $r \gtrsim 400$ may explain the differences between simulation and the traced photograph in Fig. 9.18; the photograph in Fig. 9.12 was taken in the distal slope facies, corresponding to large r. Moreover, the hypothesis that the distal slope facies is associated with small α terraces is supported qualitatively by field observations. Combined with the results in the previous section, this implies that terraces in the distal slope facies are characterized by small α and $\langle\lambda\rangle$.

9.5 Conclusions

In this chapter, we showed that travertine precipitation pattern formation results from an interplay between fluid flow, capillarity and chemistry, which can be captured quantitatively from both analytical and cell dynamical system approaches. We have demonstrated that the shape of circularly symmetric travertine domes can be largely explained by an analytical model which couples thin film fluid flow with precipitation, and that the shape can be completely explained through CDS models which include surface tension effects. We then used the analytical theory to predict a scaling law describing the dome shape where the contact line forms, and confirmed this theory using the CDS models; this demonstrated that both models capture the same essential physics.

We applied these same models to the problem of travertine dams, and explained the formation and statistical properties of travertine terraces. The analytic theory demonstrated that the carbonate depositional process was fundamentally unstable to linear perturbations of any wavelength. We used the CDS model to examine the nonlinear regime, and showed that the same simple rules which explained the formation of domes on flat surfaces also leads to the formation of terraces and pond lips. These simulations reproduce qualitative and quantitative features of real hot springs. In particular, we demonstrated that these ponds obey the same terrace width distribution regardless of their size or location within the spring system, a result we confirm using a photograph of small terraces. This distribution is consistent with our reported self-similar distribution of pond areas. Finally, on the basis of simulation results, we advanced the hypothesis that the distal slope facies can be identified by its narrow terrace aspect ratio and lack of wide terraces.

9.6 Appendix 1

To make the time-lapse movies, which we used to better understand the dynamical processes that control travertine pattern formation, we relied first and foremost on the help of Yellowstone Park Rangers. Without the assistance of Brian Suderman, none of this work would have been possible. As often as possible (about 40 times over 1.5 years), Brian took the supplied digital camera to Canary Springs, clipped it into the attached permenant mount which fixed its orientation, and took a picture.[4] He regularly emails these images, which we then carefully align to produce the movie.

Knowing that the rangers could take photographs no more frequently than once a week, we determined the camera location according to the following considerations. Table 9.1 relates the physical sensitivity of 1 camera pixel on our 4 megapixel camera at a given distance. We know that precipitation rates are $0.04 - 0.08 cm/day$. This corresponds to $0.28 - 0.56 cm/week$ or $0.15 - 0.3 m/year$. Therefore, if we are photographing once a week, we want to make our movies from 10 meters away. At this distance. the growth in one week, $\sim 0.4 cm$, is on the same order as our camera sensitivity ($0.3 cm$).

Once we received the photographs, we processed them using the program `autopanosift`[203] to automatically identify matching control points between the photographs. These control points required a great deal of manual fine tuning due to the evolving nature of the landscape. We edited the control points using a piece of software called `hugin`[204], which allows us to build a detailed model of camera lens abberations on the basis of EXIF focal lengths and the fixed control points. This software relies on the `panorama-tools` package to perform the necessary nonlinear image transformations which align the series of images according to the control points while accounting for lens distortion. Although we shot all pictures from a fixed mount, a substantial amount of fine tuning was needed for the images to align properly.

[4] Due to rapid $CaCO_3$ precipitation, we were forced to move our photography site once; the mounting point was overrun by the hot springs!

Target Distance (m)	1 pixel resolution
0.5	0.02 cm
1	0.03 cm
1.5	0.05 cm
2	0.07 cm
5	0.16 cm
10	0.3 cm
15	0.5 cm
20	0.7 cm
30	1.0 cm
60	2 cm

Table 9.1: Camera sensitivity to targets at a given distance.

To use `hugin`, we would first run the alignment optimizer (starting with a subset of parameters). Next we review the control points, adding or subtracting points to improve the worst alignments. We repeat this until mean error < 1 pixel, and the biggest errors are ~ 5 pixels. We then assemble the images, using `PTStitcher` with rectilinear output.

When this is complete, we can save the time-lapse movie as layers in a layered image format (e.g., a `Photoshop Document`, or `.psd` file). Cropping this file in an image editor allows us to select the desired scene, and remove any portions which do not appear in all frames. The cropped layers are exported as separate files and then assembled into a movie (e.g. using `Adobe Premiere`).

Chapter 10

Conclusions

In this dissertation, I have presented analytical, experimental, and numerical work relating to two problems in fluid flow.

In order to develop my modelling skills, I first addressed the oldest, hardest, and arguably one of the most important problems in *boundary layer* theory: determining the drag on a sphere and an infinite cylinder moving at a fixed speed in a highly viscous fluid. After a much-needed extensive review of previous experimental and theoretical work, I used techniques based on the perturbative renormalization group to produce optimal "coarse-grained" approximations, resulting in a new prediction for the drag coefficient. I demonstrated that RG could reproduce and improve upon the results of matched asymptotics, without the need for multiple intricate expansions. The more economical RG expansions result in a superior approximate drag coefficient, one which does not behave pathologically at non-infinitesimal Reynolds number. Given the historical importance of this problem, it is hoped that this successful demonstration of the RG approach will result in further interest in, and more widespread application of the technique.

Following this success, I presented my investigations into the biocomplex system at Yellowstone National Park. This research, which includes field work, experimental measurements, and numerical modelling, has shown that microbes do not play an im-

portant role in the formation of large scale carbonate terraces. Although we found that microbial communities are highly correlated with depositional facies, we demonstrated that this is a consequence of tight correlations between aqueous chemistry and the underlying travertine. We showed how these correlations can be used to reconstruct ancient depositional environments. We also demonstrated explicitly that any impact of microbial metabolic activities on precipitation is negligible in comparison to physical controls (such as carbon dioxide degassing), although microbes may be important as passive nucleation sites.

This conclusion is further supported by our highly successful minimal models. We used both a cellular model and analytical calculations to explain the formation of large scale travertine motifs, including the characteristic terrace architecture. Neither of these models requires any microorganisms. The role of microbes in the vast majority of the system is limited to contributing to the kinetics of growth, something we cannot address at our level of description. The cellular model predicts dynamical phenomena which have been seen subsequently at real hot springs. It also produces patterns which have the same static statistical properties as those we have been able to quantify at real hot springs, and indicates that hot spring terraces are indeed scale invariant; we showed that the distribution of pond areas in both simulated landscapes and real hot springs obeys the same power law. By analogy with studies of vicinal surfaces, we identified another universal statistical characterization, the *terrace width distribution*. We presented evidence from both simulated landscapes and experimental data that this distribution is a universal property of carbonate terraces, applying regardless of the details of local chemistry and biology. Finally, we showed that the geomorphology of the geologists' facies model is a generic outcome of the overall physics.

10.1 Reflections on Interdisciplinary Research

The work presented in this dissertation included diverse experimental, analytical, and numerical techniques. While I am proud of this accomplishment, I view the fundamentally interdisciplinary nature of this work as a harder and more important achievement. Although my doctoral work began with a calculation lengthy and intricate enough to make an applied mathematician weep, it developed into something richer when I began work on the Yellowstone project. I have listened to a microbiologist speak interminably on the relative virtues of lysing cells with freeze thawing or bead beating — and he never spared the details. I have lost arguments with geologists when the rock under the microscope proved that I was a physicist; I have taught a geochemist the virtues of Monte Carlo simulations; and I have sat in the sulfur-laden steam of a hot spring at 3am, trying to figure out why the pH probe would not calibrate in the freezing air.

These experiences have made me a richer person and a better scientist. Beyond the intellectual excitement that comes from being exposed to such a wide range of disciplines, I have been forced to consider other ways of thinking about problems, forced to learn how to speak the languages of other fields, and forced to temper my own ignorance so as not to alienate my coworkers. Through weekly meetings, field work, and evenings at the Two Bit Saloon (Gardiner, Montana), I have forged relationships with members of the Fouke group and developed a taste for Moose Drool. These relationships have led to sharing of ideas, sometimes through loud arguments. Without these disagreements, however, I could not have built the computer models which are my primary contribution to the Yellowstone project. I needed to have taken hundreds of pH measurements, to have stared at the flowing spring water and growing terraces, to have painstakingly assembled frame after frame in our time-lapse movie, and to have tried and failed to measure DIC. Without these experiences, and numerous conversations with geologists who have spent vastly more time studying the springs, I could not have known what to put into my models; sitting and reading papers would have led to paper

models and I would have had no idea if they resembled real hot springs.

Interdisciplinary research always looks good on paper. Everybody does their part, and we can understand a system which nobody could have explained individually. But that's not the way it works. We physicists didn't like the way geologists calculated their error bars, and we didn't know a rock hammer from a pH probe when we first went into the field. So we branch into each other's work, teach each other new skills, and we advance our understanding through small steps, confusion, and lots of meetings in which we have no idea what the other side finds important. But by repeating this process over and over again, we have surprisingly learned each other's language, and even advanced our understanding of hot springs.

The process of writing a paper embodies this experience. As the lead author of our recent *Geology* paper, I described the results of our analysis of extensive aqueous chemistry data. I produced a decent first draft and gave it to my co-authors for review. A month and twenty drafts later, after countless hours in meetings, we were finally done. The geologists, of course, had to completely rework the thematic structure, couching it in the language and context needed to explain its relevance to that field. What we physicists called "thin film flow" became "thin sheet flow. The chemist forced me to be precise about the difference between DIC and CO_2, and the microbiologists drove home the fact that not all "microbes" are "bacteria". This process (or ordeal) played itself out over hours of rewriting sentences word by word, with each person having their say. But it resulted in a paper — an excellent paper — which none of us could have produced independently. It forced us all to combined our separate results and viewpoints in a very tangible way.

Interdisciplinary research is done in the trenches. Dividing a complex problem into little pieces, each of which is given to the appropriate expert simply does not work. The very difficulty inherent in true interdisciplinary science is what makes it so rewarding. My collaborations on this project have turned intellectual relationships into

personal ones, and I used those to build the understanding of other disciplines needed to collectively solve an immensely complicated problem. Using this understanding to forge explanations is a deeply rewarding process, and the resulting papers embody research which could not have been done any other way.

References

[1] N. D. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group*. Reading, MA: Addison-Wesley, 1992.

[2] B. P. Athreya, N. Goldenfeld, and J. A. Dantzig, "Renormalization group theory for the phase field crystal equation," 2006.

[3] K. R. Elder and M. Grant, "Modeling elastic and plastic deformations in nonequilibrium processing using phase field crystals," *Physical Review E*, vol. 70, 2004.

[4] N. Goldenfeld, B. P. Athreya, and J. A. Dantzig, "Renormalization group approach to multiscale simulation of polycrystalline materials using the phase field crystal model," *Physical Review E*, vol. 72, 2005.

[5] M. Mondello and N. Goldenfeld, "Scaling and vortex dynamics after the quench of a system with a continuous symmetry," *Phys. Rev. A*, vol. 42, pp. 5865–5872, 1990.

[6] T. Nagaya, H. Orihara, and Y. Ishibashi, "Coarsening dynamics of +1 and -1 disclinations in two-dimensionally aligned nematics: spatial distribution of disclinations," *J. Phys. Soc. Jpn.*, vol. 64, pp. 78–85, 1995.

[7] T. S. Kuhn, *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press, 3 ed., 1996.

[8] K. R. Popper, *The logic of scientific discovery*. New York: Scientific Editions, 1961.

[9] P. K. Feyerabend, *Against method*. New York: Verso, 3 ed., 1993.

[10] R. W. Batterman, *The devil in the details : asymptotic reasoning in explanation, reduction, and emergence*. Oxford studies in philosophy of science, New York: Oxford University Press, 2002.

[11] R. Pastor-Satorras and D. Rothman, "Scaling of a slope: The erosion of tilted landscapes," *J. Stat. Phys.*, vol. 70, pp. 477–500, 1998.

[12] G. G. Stokes, "On the effect of the internal friction of fluids on the motion of pendulums," *Trans. Camb. Phil. Soc.*, vol. 9, pp. 8–106, 1851. Part II.

[13] W. Chester and D. R. Breach, "On the flow past a sphere at low Reynolds number," *Journal of Fluid Mechanics*, vol. 37, pp. 751–760, 1969.

[14] E. R. Lindgren, "The Motion of a sphere in an incompressible viscous fluid at Reynolds Numbers considerably less than one," *Physica Scripta*, vol. 60, pp. 97–110, 1999.

[15] C. M. Bender and S. A. Orzag, *Advanced Mathematical Methods for Scientists and Engineers, I.* New York: Springer-Verlag New York, Inc., 1999.

[16] L. Chen, N. Goldenfeld, and Y. Oono, "Renormalization group and singular perturbations: Multiple scales, boundary layers, and reductive perturbation theory," *Physical Review E*, vol. 54, pp. 376 – 394, Jul 1996.

[17] L. Chen, N. Goldenfeld, and Y. Oono, "Renormalization-Group Theory For Global Asymptotic Analysis," *Physical Review Letters*, vol. 73, pp. 1311 – 1315, Sep 5 1994.

[18] B. W. Fouke, G. T. Bonheyo, B. Sanzenbacher, J. Frias-Lopez, and J. Veysey, "Partitioning of bacterial communities between travertine depositional facies at Mammoth Hot Springs, Yellowstone National Park, USA," *Canadian Journal of Earth Science*, vol. 40, pp. 1531–1548, 2003.

[19] J. Veysey, T. Schickel, B. Fouke, M. Kandianis, R. Johnson, and N. Goldenfeld, "Toward a Predictive Model of Hot Spring Water from Modern and Ancient Travertine Depositional Facies," *Geology*, 2006. Submitted for publication.

[20] N. Goldenfeld, P. Chan, and J. Veysey, "Dynamics of precipitation pattern formation at geothermal hot springs," *Physical Review Letters*, vol. 96, 27 June 2006. http://arxiv.org/abs/nlin.PS/0601028.

[21] K. O. L. F. Jayaweera and B. J. Mason, "The behaviour of freely falling cylinders and cones in a viscous fluid," *Journal of Fluid Mechanics*, vol. 22, pp. 709–720, 1965.

[22] D. J. Tritton, "Experiments on the flow past a circular cylinder at low Reynolds numbers," *Journal of Fluid Mechanics*, vol. 37, pp. 547–567, 1959.

[23] T. Maxworthy, "Accurate Measurements of Sphere Drag at Low Reynolds Numbers," *Journal of Fluid Mechanics*, vol. 23, pp. 369– 372, 1965.

[24] B. P. Le Clair and A. E. Hamielec, "A Numerical Study of the Drag on a Sphere at Low and Intermediate Reynolds Numbers," *Journal of the Atmospheric Sciences*, vol. 27, pp. 308–315, 1970.

[25] S. C. R. Dennis and J. D. A. Walker, "Calculation of the steady flow past a sphere at low and moderate Reynolds numbers," *Journal of Fluid Mechanics*, vol. 48, pp. 771–789, 1971.

[26] I. Proudman and J. R. A. Pearson, "Expansions at small Reynolds numbers for the flow past a sphere and a circular cylinder," *Journal of Fluid Mechanics*, vol. 2, pp. 237–262, 1957.

[27] S. Goldstein, "The Steady Flow of Viscous Fluid past a Fixed Spherical Obstacle at Small Reynolds Numbers," *Proc. Roy. Soc. A*, vol. 123, pp. 225–235, 1929.

[28] E. J. Hinch, *Perturbation Methods*. Cambridge: Cambridge University Press, 1991.

[29] M. Van Dyke, *Perturbation Methods in Fluid Mechanics*. Stanford, CA: The Parabolic Press, 1975.

[30] S. Kaplun, "Low Reynolds Number Flow past a circular cylinder," *J. Math. Mech.*, vol. 6, pp. 595–603, 1957.

[31] S. Tomotika and T. Aoi, "The steady flow of viscous fluid past a sphere and circular cylinder at small Reynolds number," *Quart. J. Mech. Appl. Math*, vol. 3, pp. 140–161, 1950.

[32] C. W. Oseen, "Ueber die Stokes'sche Formel, und über eine verwandte Aufgabe in der Hydrodynamik," *Ark. f. Mat. Astr. og Fys*, vol. 6, no. 3,4,29, 1910.

[33] L. Landau and E. Lifschitz, *Fluid Mechanics*. Butterworth-Heinemann, 2 ed., 1999.

[34] P. Lagerstrom, L. Howard, and C. Liu, eds., *Fluid Mechanics and Singular Perturbations, A Collection of Papers by Saul Kaplun*. Academic Press, Inc., 1967. Includes reprints of other Kaplun references.

[35] S. Goldstein, ed., *Modern Develoments in Fluid Mechanics: An Account of Theory and Experiment Relating to Boundary Layers, Turbulent Motion, and Wakes*, vol. I, II. 180 Varick St. New York, NY 10014: Dover Publications, Inc., 1 ed., 1965.

[36] H. Liebster, "Über den Wilderstand von Kugeln," *Annalen der Physik*, vol. 82, pp. 541–562, 1927.

[37] H. Liebster and L. Schiller, "Vorträge Und Diskussionen Von Der 88. Versammlung Deutscher Naturforscher und Ärzte in Innsbruck," *Physikalische Zeitschrift*, vol. 25, pp. 670–672, 1924.

[38] L. A. Skinner, "Generalized Expansions for Slow Flow Past A Cylinder," *Q. Jl. Mech. Appl. Math*, vol. 28, pp. 333–340, 1975.

[39] H. S. Allen, "The Motion of A Sphere in a Viscous Fluid," *Philosophical Magazine*, vol. 5, no. 50, pp. 323–338,519–534, 1900.

[40] Arnold H. D., "Limitations imposed by slip and inertia terms upon Stokes's law for the motion of spheres through liquids," *Philosophical Magazine*, vol. 6, no. 22, pp. 755–775, 1911.

[41] W. E. Williams, "On the Motion of a Sphere in a Viscous Fluid," *Philosophical Magazine, Series 6*, no. 29, pp. 526–552, 1915.

[42] C. Wieselsberger, "Weitere Festellungen über die Gesetze des Flüssigkeits- und Luftwiderstandes," *Physikalische Zeitschrift*, vol. 23, pp. 219–224, 1922.

[43] J. Schmiedel, "Experimentelle Unterscuhungen über die Fallhewegun von Kugeln und Scheiben in reibenden Flüssigkeiten," *Physikalische Zeitschrift*, vol. XXIX, 1 Sept. 1928.

[44] R. A. Castleman, "N.A.C.A. Technical Note," Tech. Rep. 231, Bureau of Standards, U.S.A., 1925.

[45] J. Perry, ed., *Chemical Engineering Handbook*. 3 ed., 1950.

[46] D. Tritton, *Physical Fluid Dynamics*. Clarendon Press, 3 ed., 1988.

[47] V. Fidleris and R. L. Whitmore, "Experimental determination of the wall effect for spheres falling axially in cylindrical vessels," *British Journal of Applied Physics*, vol. 12, pp. 490–494, 1961.

[48] H. Pruppacher and B. P. Le Clair, "Some relations between drag and flow pattern of viscous flow past a sphere and a cylinder at low and intermediate Reynolds numbers," *Journal of Fluid Mechanics*, vol. 44, part 4, pp. 781–790, 1970.

[49] H. R. Pruppacher and E. H. Steinberger, "An Experimental Determination of the Drag on a Sphere at Low Reynolds Numbers," *Journal of Applied Physics*, vol. 39, pp. 4129–4132, Aug. 1968.

[50] K. V. Beard and H. R. Pruppacher, "A Determination of the Terminal Velocity and Drag of Small Water Drops by Means of a Wind Tunnel," *Journal of the Atmospheric Sciences*, vol. 26, pp. 1066–1072, Sept. 1969.

[51] A. E. Hamielec, T. W. Hoffman, and L. Ross, "Numerical solution of the Navier-Stokes Equation for Flow Past Spheres: Part I. Viscous flow around spheres with and without radial mass efflux," *A. I. Ch. E. Journal*, vol. 13, pp. 212–219, Mar. 1967.

[52] Y. Rimon and S. I. Cheng, "Numerical solutions of a uniform flow over a sphere at intermediate Reynolds numbers," *Physics of Fluids*, vol. 12, pp. 949–959, 1969.

[53] V. G. Jenson, "Viscous flow around a sphere at low Reynolds numbers," *Proceedings of the Royal Society of London, A*, vol. 249, pp. 346–366, 1959.

[54] M. Kawaguti, "Numerical Solution for the Viscous Flow past a Sphere," *Tokyo University Institute of Science and Technology Reports*, vol. 4, pp. 154–158, 1950.

[55] P. Brown and D. Lawler, "Sphere Drag and Settling Velocity Revisited," *Journal of Environmental Engineering*, pp. 222–231, Mar. 2003.

[56] Lamb H., "On the Uniform Motion of a Sphere Through a Viscous Fluid," *Philosophical Magazine*, no. 21, pp. 112–121, 1911.

[57] E. Relf, "Discussion of the results of measurements on the resistance of wires, with some additional tests on the resistance of wires of small diameter," *British A.R.C. Reports and Memoranda*, no. 102, 1914.

[58] C. Wieselsberger, "Neuere Feststellungen über die Gesetze des Flüssigkeits- und Luftwiderstandes," *Physikalische Zeitschrift*, vol. 22, no. 13, pp. 321–328, 1921.

[59] W. Linke, "Neue Messungen zur Aerodynamik des Zylinders, insbsondere seines reinen Reuibungswiderstandes," *Physickalische Zeitschrift*, vol. 32, pp. 900–914, 1931. Originalmitteilungen (Leipziger Dissertation).

[60] H. Lamb, *Hydrodynamics*. Cambridge: Cambridge University Press, 6 ed., 1932.

[61] A. Thom, "The flow past circular cylinders at low speeds," *Proceedings of the Royal Society A*, vol. 141, pp. 651–669, 1933.

[62] C. M. White, "The drag of cylinders in fluids at slow speeds," *Proceedings of The Royal Society A*, vol. 186, pp. 472–479, 1946.

[63] R. K. Finn, "Determination of the Drag on a Cylinder at Low Reynolds Numbers," *Journal of Applied Physics*, vol. 24, pp. 771–773, June 1953.

[64] R. L. Underwood, "Calculation of incompressible flow past a circular cylinder at moderate Reynolds number," *Journal of Fluid Mechanics*, vol. 37, pp. 95–114, 1969.

[65] J. S. Son and T. J. Hanratty, "Numerical solution for the flow around a cylinder at Reynolds numbers of 40, 200, and 500," *Journal of Fluid Mechanics*, vol. 35, pp. 369–386, 1969.

[66] M. Kawaguti and P. Jain, "Numerical study of a viscous fluid past a circular cylinder," MRC Summary Report 590, University of Wisconsin, 1966.

[67] S. C. R. Dennis and M. Shmshoni, "The steady flow of a viscous fluid past a circular cylinder," *Aero. Res. Counc. Lond. Current Paper*, no. 797, 1965.

[68] C. J. Apelt, "The steady flow of a viscous fluid past a circular cylinder at Reynolds numbers 40 and 44," *Aero. Res. Counc. Lond. R & M*, no. 3175, 1961.

[69] D. N. Allen and R. V. Southwell, "Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder," *Quarterly Journal of Mechanics and Applied Mathematics*, vol. 8, pp. 129–145, 1955.

[70] J. Happel and H. Brenner, *Low Reynolds number hydrodynamics, with special applications to particulate media*. Mechanics of fluids and transport processes Monographs and textbooks on mechanics of solids and fluids, Leyden: Noordhoff International Publishing, 2 ed., 1973.

[71] E. Buckingham, "On Physically Similar Systems; Illustrations of the use of Dimensional Equations," *Phys. Rev.*, vol. 4, p. 345, 1914.

[72] G. I. Barenblatt, *Scaling, Self-Similarity, and Intermediate Asymptotics.* New York: Consultant Bureau, 1 ed., 1979.

[73] M. Krakowski and A. Charnes Tech. Rep. 37, Carnegie Institute of Technolody Dept. of Mathematics, 1953.

[74] A. N. Whitehead, "Second approximations to viscous fluid motion," *Quart. J. Math.*, vol. 23, pp. 143–152, 1888.

[75] Lord Rayleigh, "On the Flow of Viscous Liquids, especially in Two Dimensions," *Philosophical Magazine, Series 5*, no. 36, pp. 354–372, 1893.

[76] H. Faxén, "Exakte Losung der Oseenschen Differentialgleichungen einer Zahen Flussigkeit fur den Fall der Translationsbewegung eines Zylinders," *Nova Acta Societatis Scientiarum Upsaliensis*, vol. Volumen extra ordine, pp. 1–55, 1927.

[77] F. Noether, "Über den Gültigkeitsbereich der Stokesschen Widerstandsformel," *Zeitschrift Für Mathematik Und Physik*, vol. 62, no. 1, pp. 1–39, 1913.

[78] C. W. Oseen, "Über den Gülitigkeitsbereich der Stokesschen Widestrandsformel," *Arkiv För Matematik Astronomi och Fysik*, vol. 9, no. 16, pp. 1–15, 1913.

[79] Burgess R.W., "The Uniform Motion of a Sphere Through A Viscous Liquid," *American Journal of Mathematics*, vol. 38, pp. 81–96, Jan. 1916.

[80] S. Kaplun and P. Lagerstrom, "Asymptotic Expansions of Navier-Stokes solutions for small Reynolds numbers," *J. Math. Mech.*, vol. 6, pp. 585–593, 1957.

[81] L. Bairstow and Cave B.M., "The Resistance of a Circular Cylinder Moving in a Viscous Fluid," *Phil. Trans. Roy. Soc. A*, vol. 223, p. 383, 1923.

[82] H. Faxén, "Die Bewegung einer starren Kugel längs der Achse eines mit zäher Flüssigkeit gefüllten Rohres," *Arkiv För Matematik, Astronomi Och Fysik*, vol. 17, no. 27, pp. 1–28, 1923.

[83] H. Faxén, *Einwirkung der Gefässwände auf den Widerstand gegen die Bewegung einer kleinen Kugel in einer zähen Flüssigikeit.* PhD thesis, Upsalla University, 1921.

[84] T. Bohlin Master's thesis, Royal Institute of Technology, 1960.

[85] W. Haberman and R. M. Saure, "Motion of Rigid and Fluid Spheres in Stationary and Moving Liquids Iinside Cylindrical Tubes," *David Taylor Model Basin Report*, vol. 1143, Oct. 1958.

[86] D. Shanks, "Non-linear transformations of divergent and slowly convergent sequences," *J. Mathematics and Physics*, vol. 34, pp. 1–42, 1955.

[87] M. Van Dyke, "Extension of Goldstein's series for the Oseen drag of a sphere," *Journal of Fluid Mechanics*, vol. 44, pp. 365–372, 1970.

[88] Imai Isao, "A new method of solving Oseen's equations and its application to the flow past an inclinded elliptic cylinder," *Proceedings of The Royal Society*, vol. 224, pp. 141–160, 22 June 1954.

[89] W. Chester, "On Oseen's approximation," *Journal of Fluid Mechanics*, vol. 37, pp. 751–760, 1962.

[90] P. A. Lagerstrom and J. D. Cole, "Examples illustrating expansion procedures for the Navier-Stokes equations," *Journal of Rational Mechanics and Analysis*, vol. 4, pp. 817–882, 1955.

[91] S. Kaplun, "The role of coordinate systems in boundary-layer theory," *Z. Angew. Math. Phys.*, vol. 5, pp. 115–135, 1954.

[92] J. Keller and M. Ward, "Asymptotics beyond all orders for a low Reynolds number flow," *Journal of Engineering Mathematics*, vol. 30, pp. 253–265, 1996.

[93] G. Carrier, "On slow viscous flow, Final Rept.," tech. rep., Office of Naval Research Contract Nonr-653(00), Brown University, 1953.

[94] P. A. Lagerstrom and R. G. Casten, "Basic concepts underlying singular perturbation techniques," *SIAM Review*, vol. 14, pp. 63–120, 1972.

[95] Y. Oono, "Renormalization and asymptotics," *International Journal Of Modern Physics B*, vol. 14, pp. 1327 – 1361, May 30 2000.

[96] T. Kunihiro, "A Geometrical Formulation Of The Renormalization-Group Method For Global Analysis," *Progress Of Theoretical Physics*, vol. 94, pp. 503 – 514, Oct 1995.

[97] S. L. Woodruff, "A uniformly valid asymptotic solution to a matrix system of ordinary differential equations and a proof of its validity," *Stud. Appl. Math.*, vol. 94, no. 4, pp. 393–413, 1995.

[98] S. L. Woodruff, "The use of an invariance condition in the solution of multiple-scale singular perturbation problems: ordinary differential equations," *Stud. Appl. Math.*, vol. 90, no. 3, pp. 225–248, 1993.

[99] G. T. Bonheyo, H. Garçía Martín, J. Veysey II, J. Frias-Lopez, N. Goldenfeld, and B. W. Fouke, *Statistical Evaluation of Bacterial 16S rRNA Gene Sequences in Relation to Travertine Mineral Precipitation and Water Chemistry at Mammoth Hot Springs, Yellowstone National Park, USA*. 2006.

[100] T. D. Ford and H. M. Pedley, *A review of tufa and travertine deposits of the world*, vol. 41, pp. 117–175. 1996.

[101] D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics*. Cambridge University Press, 1992.

[102] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group*. Reading, MA: Addison-Wesley, 1992.

[103] B. W. Fouke, J. D. Farmer, D. J. Des Marais, L. Pratt, N. C. Sturchio, P. C. Burns, and M. K. Discipulo, *Depositional facies and aqueous-solid geochemistry of travertine-depositing hot springs (Angel Terrace, Mammoth Hot Springs, Yellowstone National Park, USA)*, vol. 70, pp. 265–285. 2000.

[104] I. Friedman, *Some investigations of the deposition of travertine from hot springs: I. The isotope chemistry of a travertine-depositing spring*, vol. 34, pp. 1303–1315. 1970.

[105] A. Pentecost, *The formation of travertine shrubs: Mammoth Hot Springs, Wyoming*, vol. 127, pp. 159–168. 1990.

[106] H. S. Chafetz and R. L. Folk, *Travertines: Depositional morphology and the bacterially constructed constituents*, vol. 54, pp. 289–386. 1984.

[107] Chafetz, H.S., Wikinson, B.H., and Love, K.M., *Morphology and Composition of Non-Marine Carbonate Cements in Near-Surface Settings*, vol. 36, pp. 337–347. 1985.

[108] NASA, "Life on Mars?." http://www.jsc.nasa.gov/er/seh/marslife.html.

[109] D. Mittlefehldt, "ALH84001, A cumulate orthopyroxenite member of the martian meteorite clan," *Meteoritics*, vol. 29, no. 6, p. 900, 1994.

[110] C. P. McKay and C. R. Stoker, *The early environment and its evolution on Mars: implications for life*, vol. 27, pp. 189–214. 1989.

[111] D. S. McKay, E. K. Gibson, K. L. Thomas-Keptra, H. Vali, C. S. Romanek, S. J. Clemett, X. D. F. Chillier, C. R. Maechling, and R. N. Zare, *Search for past life on Mars: possible relic biogenic activity in Martian meteorite ALH84001*, vol. 273, pp. 924–930. 1996.

[112] D. S. McKay, E. K. Gibson Jr, K. L. Thomas-Keprta, H. Vali, C. S. Romanek, S. J. Clemett, X. D. F. Chillier, C. R. Maechling, and R. N. Zare, "Search for past life on mars: possible relic biogenic activity in martian meteorite ALH84001," *Science*, vol. 273, no. 5277, pp. 924–930, 1996.

[113] R. A. Kerr, "Requiem for life on Mars? Support for microbes fades," *Science*, vol. 282, no. 5393, p. 1398, 1998.

[114] H. S. Chafetz and S. A. Guidry, "Bacterial shrubs, crystal shrubs, and ray-crystal shrubs: bacterial vs. abiotic precipitation," *Sediment. Geol.*, vol. 126, no. 1-4, pp. 57–74, 1999.

[115] H. S. Chafetz, P. F. Rush, and N. M. Utech, *Microenvironmental controls on mineralogy and habit of $CaCO_3$ precipitates: an example from an active travertine system*, vol. 38, pp. 107–126. 1991.

[116] W. Dreybrodt, L. Eisenlohr, B. Madry, and S. Ringer, "Precipitation kinetics of calcite in the system $CaCO_3$-H_2O-CO_2: The conversion to CO_2 by the slow process $H^+ + HCO_3 \rightarrow CO_2 + H_2O$ as a rate limiting step," *Geochim. Cosmochim. Acta*, vol. 61, no. 18, pp. 3897–3904, 1997.

[117] E. Allen and A. Day, "Hot Springs of the Yellowstone National Park," vol. Publication Number 466, p. 525, Carnegie Institution of Washington, 1935.

[118] K. E. Bargar, *Geology and thermal history of Mammoth Hot Springs, Yellowstone National Park, Wyoming*, vol. 1444, pp. 1–54. 1978.

[119] N. C. Sturchio, M. T. Murrell, K. L. Pierce, and M. L. Sorey, *Yellowstone Travertines: U-series Ages and Isotope Ratios (C, O, Sr, U)*, pp. 1427–1430. Rotterdam: Balkema, 1992. ISBN 9054100753.

[120] N. C. Sturchio, K. L. Pierce, M. T. Murrell, and M. L. Sorey, *Uranium-series ages of travertines and timing of the last glaciation in the northern Yellowstone area, Wyoming-Montana*, vol. 41, pp. 265–277. 1994.

[121] D. E. White, R. O. Fournier, L. P. J. Muffler, and A. H. Truesdell, *Physical results of research drilling in thermal areas of Yellowstone National Park, Wyoming.* Menlo Park, California: United States Geological Survey, 1975.

[122] Y. K. Kharaka, M. L. Sorey, and J. J. Thordsen, *Large-scale hydrothermal fluid discharges in the Norris-Mammoth corridor, Yellowstone National Park, USA*, vol. 69, pp. 201–205. 2000.

[123] M. L. Sorey, *Effects of potential geothermal development in the Corwin Springs known geothermal resources area, Montana, on the thermal features of Yellowstone National Park.* Menlo Park, CA: United States Geological Survey, 1991.

[124] N. C. Sturchio, *Radium isotopes, alkaline earth diagenesis, and age determination of travertine from Mammoth Hot Springs, Wyoming, USA*, vol. 5, pp. 631–640. 1990.

[125] K. Pierce, K. Adams, and N. Sturchio, *Geologic setting of the Corwin Springs Known Geothermal Area-Mammoth Hot Springs area in and adjacent to Yellowstone*, vol. Water-resources Investigations Report 91-4052, pp. C1–C37. U.S. Geological Survey, 1991.

[126] Y. Kharaka, R. Mariner, T. Bullen, B. Kennedy, and N. Sturchio, *Geochemical investigations of hydraulic connections between Corwin Springs Known Geothermal Area and adjacent parts of Yellowstone National Park*, vol. Water-Resources Investigations Report 91-4052, pp. F1–F38. U.S. Geological Survey, 1991.

[127] E. Busenberg and L. Plummer, *A comparative study of the dissolution and crystal growth kinetics of calcite and aragonite*, pp. 139–168. 1986.

[128] W. Dreybrodt, D. Buhman, J. Michaelis, and E. Usdowski, *Geochemically controlled calcite precipitation by CO_2 outgassing: Field measurements of precipitation rates in comparison to theoretical predictions*, vol. 97, pp. 285–294. 1992.

[129] W. Dreybrodt, "Precipitation kinetics of calcite in the system $CaCO_3$-H_2O-CO_2: The conversion to CO_2 by the slow process $H^+ + HCO_3 \rightarrow CO_2 + H_2O$ as a rate limiting step - Reply," *Geochim. Cosmochim. Acta*, vol. 62, no. 23-24, pp. 3791–3792, 1998.

[130] W. Dreybrodt, J. Lauckner, Z. H. Liu, U. Svensson, and D. Buhmann, "The kinetics of the reaction $CO_2 + H_2O \rightarrow H^+ + HCO_3^-$ as one of the rate limiting steps for the dissolution of calcite in the system $H_2O - CO_2 - CaCO_3$," *Geochim. Cosmochim. Acta*, vol. 60, no. 18, pp. 3375–3381, 1996.

[131] T. M. L. Wigley, L. Plummer, and J. Pearson, F. J., *Mass transfer and carbon isotope evolution in natural water systems*, vol. 42, pp. 1117–1139. 1978.

[132] Z. H. Liu and W. Dreybrodt, "Dissolution kinetics of calcium carbonate minerals in $H_2O - CO_2$ solutions in turbulent flow: The role of the diffusion boundary layer and the slow reaction $H_2O + CO_2 \leftrightharpoons H^+ + HCO_3$," *Geochim. Cosmochim. Acta*, vol. 61, no. 14, pp. 2879–2889, 1997.

[133] D. D. Zhang, Y. J. Zhang, A. Zhu, and X. Cheng, "Physical mechanisms of river waterfall tufa (travertine) formation," *J. Sediment. Res.*, vol. 71, no. 1, pp. 205–216, 2001.

[134] G. Lu, C. Zheng, R. J. Donahoe, and W. B. Lyons, *Controlling processes in a $CaCO_3$ precipitating stream in Huanlong Natural Science District, Sichuan, China*, vol. 230, pp. 34–54. 2000.

[135] M. M. Lorah and J. S. Herman, "Calcite precipitation rates in the field: Measurement and prediction for a travertine-depositing spring," *Geochimica et Cosmochimica Acta*, vol. 52, pp. 2347–2355, 1988.

[136] G. M. Friedman, *Depositional Environments in Carbonate Rocks*, vol. 14, p. 209. 1969.

[137] A. B. Murray and C. Paola, "A cellular model of braided rivers," *Nature*, vol. 371, pp. 54–57, 1994.

[138] E. Flügel, *Microfacies of Carbonate Rocks: Analysis, Interpretation, and Application*. Berlin, New York: Springer, 2004.

[139] A. Gressly, *Observations geologiques sur le Jura Solerois*, vol. 2, pp. 1–349. 1838.

[140] H. Reading, *Sedimentary Environments: processes, Facies and Stratigraphy*. London, England: Blackwell Science Ltd., 1996.

[141] J. L. Wilson, *Carbonate Facies in Geologic History*. New York: Springer-Verlag, 1975. 551.81 W69c.

[142] R. G. Walker, *Facies Models*. St. Johns, Newfoundland: Geological Association of Canada, 1984.

[143] R. Amundson and E. Kelly, *The chemistry and mineralogy of a CO_2-rich travertine depositing spring in the California Coast Range*, vol. 51, pp. 2883–2890. 1987.

[144] D. L. Suarez, *Determination of calcite precipitation rates in irrigated arid land soils*, vol. 15, p. 701. 1983.

[145] J. S. Herman and M. M. Lorah, *Calcite precipitation rates in the field: Measurement and prediction for a travertine-depositing spring*, vol. 52, pp. 2347–2355. 1988.

[146] J. Kanakis and E. Dalas, *The crystallization of vaterite on Fibrin*, vol. 219, pp. 277–282. 2000.

[147] R. C. Kerr and J. S. Turner, "Crystallization and gravitationally controlled ponding during the formation of mound springs, terraces, and "black smoker" flanges," *J. Geophys. Res.-Solid Earth*, vol. 101, no. B11, pp. 25125–25137, 1996.

[148] N. Spanos and P. Koutsoukos, *Kinetics of Precipitation of Calcium Carbonate in Alkaline pH at Constant Supersaturarion. Spontaneous and Seeded Growth*, vol. 102, pp. 6679–6684. 1998.

[149] E. Usdowski, J. Hoefs, and G. Menschel, *Relationship between d13C and d18O fractionation and changes in major element composition in a recent calcite-depositing spring - a model of chemical variations with inorganic $CaCO_3$ precipitation*, vol. 42, pp. 267–276. 1979.

[150] P. Van Cappellen, L. Charlet, W. Stumm, and P. Wersin, "A Surface Complexation Model of the Carbonate Mineral-Aqueous Solution Interface," *Geochim. Cosmochim. Acta*, vol. 57, no. 15, pp. 3505–3518, 1993.

[151] C. Zhang, B. Fouke, G. Bonheyo, D. White, Y. Huang, and C. S. Romanek, *Lipid biomarkers and carbon isotopes of modern travertine deposits (Yellowstone National Park, USA): Implications for biogeochemical dynamics in hot-spring systems*, vol. 68, pp. 3157–3169. 2004.

[152] B. W. Fouke, J. D. Farmer, D. J. Des Marais, L. Pratt, N. C. Sturchio, P. C. Burns, and M. K. Discipulo, *REPLY-Depositional facies and aqueous-solid geochemistry of travertine-depositing hot springs (Angel Terrace, Mammoth Hot Springs, Yellowstone National Park, USA)*, vol. 71, pp. 497–500. 2001.

[153] T. Brock, M. Madigan, J. Martinko, and J. Parker, *Biology of Microorganisms*. Prentice Hall, 1999.

[154] C. R. Woese, *Bacterial evolution*, vol. 51, pp. 221–271. 1987.

[155] C. Woese, O. Kandler, and M. Wheelis, "Towards a natural system of organisms - proposal for the domains archaea, bacteria, and eucarya," *PNAS*, vol. 87, pp. 4576–4579, 1990.

[156] H. G. Martín and N. Goldenfeld, "Estinmation of microbial cover distributions at Mammoth Hot Springs using a multiple clone library resampling method," *Enivronmental Microbiology*, 8 Mar. 2006. In press; online early edition available.

[157] R. Rossello-Mora and R. Amann, "The species concept for prokaryotes," *FEMS Microbiology Reviews*, vol. 25, pp. 39–67, 2001.

[158] J. Frias-Lopez, A. L. Zerkle, G. T. Bonheyo, and B. W. Fouke, *Partitioning of bacterial communities between seawater and healty, black band diseased, and dead coral surfaces*, vol. 68, pp. 2214–2228. 2002.

[159] W. Barnes, *The fidelity of Taq polymerase catalyzing PCR is imporved by an N-terminal deletion*, vol. 112. 1992.

[160] K. Tindall and T. Kunkel, *Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase*, vol. 27, pp. 6008–6013. 1988.

[161] E. Stackebrandt and B. M. Goebel, *Taxonomic Note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology*, vol. 44, pp. 846–849. 1994.

[162] R. K. Colwell and J. Coddington, "Estimating terrestrial biodiversity through extrapolation," *Philosophical Transactions of the Royal Society (Series B)*, vol. 345, pp. 101–118, 1994.

[163] C. Krebs, *Ecological methodology*. New York, N.Y.: Harper and Row, 1989.

[164] S.-M. Lee and A. Chao, "Estimating population size via sample coverage for closed capture-recapture models," *Biometrics*, vol. 50, pp. 88–97, 1994.

[165] J. Hughes, J. Hellman, T. Ricketts, and B. Bohannon, *MINIREVIEW- Counting the Uncountable: Statistical approaches to estimating microbial diversity*, vol. 67, pp. 4399–4406. 2001.

[166] S. Miller and R. Castenholtz, *Evolution of thermotolerance in hot spring cyanobacteria of the genus Synechococcus*, vol. 66, pp. 4222–4229. 2000.

[167] H. S. Chafetz and J. R. Lawerance, *Stable isotope variability within modern travertines*, vol. 48, pp. 257–273. 1994.

[168] J. Butler, *Carbon Dioxide Equilibria and Practice*. Mass: Addison-Wesley, 1982.

[169] W. Varnedoe, "A hypothesis for the formation of rimestone dams and gours," *National Speleogical Society Bulletin*, vol. 27, pp. 39–60, 1965.

[170] W. Williams, H. L. Gorton, and T. C. Vogelmann, "Surface gas-exchange processes of snow algae," *PNAS*, vol. 100, pp. 562–566, 21 Jan. 2003.

[171] T. Sobczuk, F. Camacho, F. Rubio, F. Acien Fernandez, and M. Grima, "Carbon Dioxide Uptake Efficiency by Outdoor Microalgal Cultures in Tubular Airlift Photobioreactors," *Biotechnology and Bioengineering*, vol. 67, 20 Feb. 2003.

[172] J. P. Grotzinger and D. H. Rothman, "An abiotic model for stromatolite morphogenesis," *Nature*, vol. 383, pp. 423–425, 1996.

[173] J. P. Grotzinger and A. H. Knoll, "Stromatolites in Precambrian Carbonates: Evolutionary Mileposts or Environmental Dipsticks?," *Annu. Rev. Earth Planet. Sci.*, vol. 27, pp. 313–358, 1999.

[174] M. Kardar, G. Parisi, and Y.-C. Zhang, "Dynamic Scaling of Growing Interfaces," *Physical Review Letters*, vol. 56, pp. 889–892, 3 Mar. 1986.

[175] A. B. Murray, "Contrasting the goal, strategies, and predictions associated with simplified numerical models and detailed simulations," in *Prediction in Geomorphology, Geophysical Monograph 135 American Geophysical Union* (D. Iverson and P. Wilcock, eds.), pp. 151–165, 2003.

[176] A. Doeschl, M. Davison, H. Rasmussen, and G. Reid, "Assessing Cellular Automata Based Models using Partial Differential Equations," *Conditionally accepted by Mathematical and Computer Modelling*, 2004.

[177] M. B. Short, J. C. Baygents, J. W. Beck, D. A. Stone, R. S. T. III, and R. E. Goldstein, "Stalactite growth as a free-boundary problem: A geometric law and its platonic ideal," *Phys. Rev. Lett.*, vol. 94, pp. 018501–018504, 2005.

[178] M. B. Short, J. C. Baygents, and R. E. Goldstein, "Stalactite growth as a free-boundary problem," *Phys. Fluids*, vol. 17, pp. 083101–083112, 2005.

[179] S. Wolfram, "Statistical mechanics of cellular automata," *Rev. Mod. Phys.*, vol. 55, pp. 601–644, 1983.

[180] D. H. Rothman and S. Zaleski, "Lattice-gas models of phase separation: interfaces, phase transitions, and multiphase flow," *Rev. Mod. Physics*, vol. 66, p. 1417, 1994.

[181] R. Pandey, "Cellular automata approach to interacting cellular network models for the dynamics of cell population in an early HIV infection," *Physica A*, vol. 179, pp. 442–470, 1991.

[182] G. B. Ermentrout and L. Edelstein-Keshet, "Cellular automata approaches to biological modeling," *Journal of Theoretical Biology*, vol. 160, pp. 97–133, Jan 1993.

[183] S. Wolfram, *A New Kind of Science*. Champaign, IL: Wolfram Media, 2002.

[184] Y. Oono and S. Puri, "Computationally efficient modeling of ordering of quenched phases," *Phys. Rev. Lett.*, vol. 58, pp. 836–839, 1987.

[185] R. C. Kerr and J. S. Turner, "Crystallization and graviationally controlled ponding during the formation of mound springs, terraces and black smoker flanges," *J. Geophysical Res.*, vol. 101, pp. 25125–25137, 1996.

[186] E. Ben-Jacob, N. Goldenfeld, J. S. Langer, and G. Schon, "Dynamics of interfacial pattern formation," *Phys. Rev. Lett.*, pp. 1930–1932, 1983.

[187] R. C. Brower, D. A. Kessler, J. Koplik, and H. Levine, "Geometrical models of interface evolution," *Phys. Rev. A*, vol. 29, pp. 1335–1342, 1984.

[188] R. A. Wooding, "Growth of natural dams by deposition from steady supersaturated shallow flow," *J. Geophysical Res.*, vol. 96, pp. 667–682, 1991.

[189] J. A. Campbell and T. J. Hanratty, "Mass transfer between a turbulent fluid and a solid boundary: linear theory," *AlChE Journal*, vol. 28, p. 988, 1982.

[190] http://www.blender3d.org/.

[191] R. F. Dressler, "New nonlinear shallow-flow equations with curvature," *J. Hydraul. Res.*, vol. 16, p. 205, 1978.

[192] N. S. Sivakumaran, R. Hosking, and T. Tingsanchali, "Steady shallow flow over a spillway," *J. Fluid Mech.*, vol. 111, p. 411, 1981.

[193] A. Chézy, "Formula to find the uniform velocity that the water will have in a ditch or in a canal of which the slope is known," 1776. File No. 847, Ms. 1915 in the library of Ecole des Ponts et Chaussées. English translation in H. Clemens, On the origin of the Chézy formula, Journal Association of Engineering Societies, v. 18, pp. 363-369, (1897).

[194] A. N. Kolmogorov, "Local structure of turbulence in incompressible fluid at a very high reynolds number," *Dokl. Akad. Nauk. SSSR*, vol. 30, pp. 299–302, 1941. [English translation in Proc. R. Soc. London Ser. A 434 (1991)].

[195] K. R. Sreenivasan, "Fluid turbulence," *Rev. Mod. Phys.*, vol. 71, pp. S383–S395, 1999.

[196] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. 40 West 20th St., New York, NY 10011-4211, USA: Cambridge University Press, 2 ed., 1992.

[197] R. Pastor-Satorras and D. H. Rothman, "Stochastic Equation for the Erosion of Inclined Topography," *Physical Review Letters*, vol. 80, pp. 4349–4353, 11 May 1998.

[198] H.-C. Jeong and E. D. Williams, "Steps on Surfaces: Experiment and Theory," *Surface Science Reports*, vol. 34, pp. 171–294, 1999.

[199] T. L. Einstein, H. L. Richards, S. D. Cohena, and O. Pierre-Louisc, "Terrace-width distributions and step-step repulsions on vicinal surfaces: symmetries, scaling, simplifications, subtleties, and Schrödinger," *Surface Science*, vol. 493, pp. 460–474, 1 Nov. 2001.

[200] H. L. Richards and T. L. Einstein, "Beyond the Wigner Distribution: Schrödinger Equations and Terrace Width Distributions," *Physical Review E*, vol. 72, 2005.

[201] A. Pimpinelli, H. Gebremariam, and T. L. Einstein, "Evolution of Terrace-width Distributions on Vicinal Surfaces: Fokker-Planck Derivation of the Generalized Wigner Surmise," *Physical Review Letters*, vol. 95, 2005.

[202] B. Joós, T. L. Einstein, and N. C. Bartelt, "Distribution of terrace widths on a vicinal surface within the one-dimensional free-fermion model," *Physical Review B*, vol. 43, no. 10, pp. 8153–8162, 1991.

[203] http://user.cs.tu-berlin.de/ nowozin/autopano-sift/.

[204] http://hugin.sourceforge.net/.

Vita

John Veysey II was born on August 30, 1975 in Malden, Massachusetts. He graduated first in his class at North Middlesex Regional High School, where he was a National Merit Scholar semi-finalist. He enrolled at Yale University in 1993, graduated four years later with a bachelors of science degree in physics, and was awarded distinction in the major by a faculty vote. After working for two years at MIT / Lincoln Laboratories, he entered the graduate program in the Department of Physics at the University of Illinois at Urbana-Champaign, received a GAANN (Graduate Assistance in Areas of National Need) fellowship, and joined Nigel Goldenfeld's research group in 2000. He took a brief break from his graduate studies in 2004, when he received an American Academy for the Advancement of Science Mass Media Fellowship. In this program, he interned as a science writer at the Milwaukee Journal Sentinel, writing articles on subjects ranging from geochemistry to ecology. He returned to UIUC as a teaching assistant in 2005, receiving the "Excellence in Teaching" award and being included on the "Incomplete List of Teachers Ranked as Excellent". In 2006, he was awarded the "John Bardeen Award" in recognition of his doctoral work in condensed matter physics. He has been a member of the American Physical Society since 2003.