UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Star and Planet Formation Through Cosmic Time

Permalink
https://escholarship.org/uc/item/46952888

Author
Lee, Aaron Thomas

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Star and Planet Formation Through Cosmic Time

By

Aaron Thomas Lee

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Astrophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Christopher F. McKee, Chair
Adjunct Professor Richard I. Klein
Professor Eliot Quataert
Researcher Steven W. Stahler
Professor Leo Blitz
Professor James Demmel

Summer 2017

Star and Planet Formation Through Cosmic Time

Copyright 2017
by
Aaron Thomas Lee

Abstract

Star and Planet Formation Through Cosmic Time

by

Aaron Thomas Lee

Doctor of Philosophy in Astrophysics

University of California, Berkeley

Professor Christopher F. McKee, Chair

The computational advances of the past several decades have allowed theoretical astrophysics to proceed at a dramatic pace. Numerical simulations can now simulate the formation of individual molecules all the way up to the evolution of the entire universe. Observational astrophysics is producing data at a prodigious rate, and sophisticated analysis techniques of large data sets continue to be developed. It is now possible for terabytes of data to be effectively turned into stunning astrophysical results. This is especially true for the field of star and planet formation. Theorists are now simulating the formation of individual planets and stars, and observing facilities are finally capturing snapshots of these processes within the Milky Way galaxy and other galaxies. While a coherent theory remains incomplete, great strides have been made toward this goal.

This dissertation discusses several projects that develop models of star and planet formation. This work spans large spatial and temporal scales: from the AU-scale of protoplanetary disks all the way up to the parsec-scale of star-forming clouds, and taking place in both contemporary environments like the Milky Way galaxy and primordial environments at redshifts of $z \sim 20$.

Particularly, I show that planet formation need not proceed in incremental stages, where planets grow from millimeter-sized dust grains all the way up to planets, but instead can proceed directly from small dust grains to large kilometer-sized boulders. The requirements for this model to operate effectively are supported by observations. Additionally, I draw suspicion toward one model for how you form high mass stars (stars with masses exceeding $\sim 8\,M_\odot$), which postulates that high-mass stars are built up from the gradual accretion of mass from the cloud onto low-mass stars. I show that magnetic fields in star forming clouds thwart this transfer of mass, and instead it is likely that high mass stars are created from the gravitational collapse of large clouds. This work also provides a sub-grid model for computational codes that employ sink particles accreting from magnetized gas. Finally, I analyze the role that radiation plays in determining the final masses of the first stars to ever form in the universe. These stars formed in starkly different environments than stars

form in today, and the role of the direct radiation from these stars turns out to be a crucial component of primordial star formation theory.

These projects use a variety of computational tools, including the use of spectral hydrodynamics codes, magneto-hydrodynamics grid codes that employ adaptive mesh refinement techniques, and long characteristic ray tracing methods. I develop and describe a long characteristic ray tracing method for modeling hydrogen-ionizing radiation from stars. Additionally, I have developed Monte Carlo routines that convert hydrodynamic data used in smoothed particle hydrodynamics codes for use in grid-based codes. Both of these advances will find use beyond simulations of star and planet formation and benefit the astronomical community at large.

This thesis is dedicated to those whose love and never-faltering support made its completion possible: my sister, my parents, and Marta Belcher.

Contents

Acknowledgments

Writing the acknowledgments section has proven to be one of the most challenging components of this dissertation. This thesis would never have even started without the support and love of countless individuals, and it is an incredible challenge to resist listing the names of every one of them here.

First off, I acknowledge the support and facilities of the NSF, NASA, UCB, and any other source that assisted in the completion of the projects detailed in this thesis. Furthermore, I thank all of the employees at these institutions for the work they do.

I thank my adviser, Christopher McKee, as well as other members of the UCB astronomy department, Richard Klein, Steven Stahler, and Eugene Chiang, for taking me on as a student and for teaching me how high-quality research in astrophysics is done. Your persistent demand for excellence has molded my own scientific thinking. Also, working with each of you has showed me what is necessary for a strong mentor/mentee relationship, and our times together will serve as my guide in my life's future chapters. I also thank Stella Offner, for taking me on as a postdoctoral scholar, and for being both patient and generous during the final stages of completing this thesis.

The current and past members of the McKee/Klein star formation group have contributed to my intellectual growth during my time at Berkeley. Thank you to Andrew Cunningham, Robert Fisher, William Gray, Charles Hansen, Mark Krumholz, Pak Shing Li, Shule Li, Andrew Myers, Jeff Oishi, Nathan Roth, Aaron Skinner, Chalence Safranek-Shrader, Athena Stacy, and any other members I may have crossed paths with during my tenure.

As a graduate student, I became committed to a career of teaching in higher education. With regards to teaching, I personally thank Jeffrey Silverman and Peter Williams for running an excellent Ay300/Ay375 course on "How to be a GSI," which struck the initial spark that grew to my current passion for higher education teaching. Thank you to Linda von Hoene and the GSI Teaching and Resource Center for all the opportunities you gave me to lead discussions and workshops about teaching, and for all the work you do to improve the quality of instruction at UCB. Furthermore, I thank Alex Filippenko and Geoffrey Marcy for demonstrating what top-quality higher education teaching looks like. Your enthusiasm and passion for teaching undergraduates was infectious, and I have gained so much from our numerous conversations on pedagogy. Being your head GSI, Alex, has been one of my most memorable experiences as a graduate student.

Furthermore, I have had the good fortune to learn from some fantastic teachers and mentors, including Donna Dennis and Mr. Wiz from high school, Fred Rasio and Giles Novak from Northwestern, and Eliot Quataert and Steven Stahler at Berkeley. Thank you for the dedication you show toward teaching.

One could not ask for a better group of undergraduate students, graduate students, and postdoctoral scholars to work among at UCB, and I thank each of you. Particularly, I wish to thank Josh Shiode, Theresa Jones, Francesca Fornasini, and Carina Cheng for agreeing to teach Ay300/Ay375 with me. That class was amazing. I also thank Doris Lee for spoiling me as my undergraduate mentee. You, with very little help from me, encapsulate the type of student that every mentor hopes to have. You have set a high bar for all my future students (and myself). Also, thank you Christopher Klein for being my old-Campbell office mate and HFA cubical mate for most of graduate school. Live long and prosper.

A special acknowledgment for Athena Stacy is absolutely necessary. Our rants and tangents on any topic whatsoever were often the highlights of the week. And I still firmly believe that our "bread method" for the projection code is genius. Thank you for your guidance and friendship. Now it is finally time for us to go and read *Watership Down*.

A similar special acknowledgement is needed for Steven Stahler, even though he has already been mentioned several times already. Even as a graduate student, working with you made me feel more like a colleague rather than like a student, and this went a long way toward developing my own self-confidence in science. I value and cherish our professional and personal relationships, both which I hope will continue long after graduate school.

I cannot end this section without thanking my family and friends for their love and support. It has been a long road, and you have all played a significant role in helping me reach the end. And, lastly, I thank Geo the cat. After finding you as a stray, I gave you a home, and you gave me companionship during a time when I needed it most. It is easy for me to say that we saved each other. Thank you.

This dissertation was typeset using the ucastrothesis LaTeX template. Thank you, Peter Williams, for creating this template. It simplified the writing process tremendously.

Chapter 1

Introduction

> *Somewhere, something incredible is waiting to be known.*
> – Carl Sagan

> *Now, my suspicion is that the universe is not only queerer than we suppose, but queerer than we can suppose.*
> – J. B. S. Haldane, in *Possible Worlds*

We live in a miraculous universe. Expanding from a Big Bang state nearly 14 billion years ago, a seemingly homogenous and isotropic universe has developed incredible complexity on smaller scales. Matter has collected into galaxies clusters and individual galaxies. Stars have formed in those galaxies, producing heavy elements that were recycled into later generations of stars. One of these stars formed about 4.5 billion years ago in a galaxy called the Milky Way. This star was born with a host of planets that formed in a flattened disk surrounding it. One of these planets soon became covered with life that gradually evolved into ever more complex forms. And today, the most advanced life-form on this planet, human beings, can look back on this series of events and understand how the universe created conditions suitable for their own existence.

Since pre-historic times, humans have gazed at the night sky and wondered about the nature and origin of the light we see. The night sky is full of stars, and while we have made considerable progress toward understanding the nature of these stars, many aspects about their origin remain the subject of intense study today. Indeed, star formation is perhaps one of the least understood phenomena in cosmic evolution, due to the complex physics involved, as well as the wide range of spatial scales and temporal scales. It is also one of the most important branches of astrophysics, because it is central to making progress on many fundamental problems, including understanding how individual stars evolve, the structure and evolution of galaxies, the energy balance and dynamics of the interstellar medium, and the formation of planetary systems. Today astronomy can utilize the entire electromagnetic spectrum to study the stars, and the last two decades of observational advances, particularly at radio and infrared wavelengths, have painted an increasingly detailed picture of the star formation process (McKee & Ostriker 2007; Dunham et al. 2014). Our work is far from complete, however, and further observations from current and future facilities like the Atacama Large Millimeter-Submillimeter Array (Brown et al. 2004) and the James Webb Space Telescope (Gardner et al. 2006) will continue to delve into the process by which light-year-sized clouds of gas transform into compact shining stars.

As our observational methods develop in sophistication, it becomes more apparent that star formation is a challenging problem. Unlocking the story of star formation has thus required the use of new techniques and ever-more sophisticated theoretical models. That said, the previous simple models still yielded incredible information and have played an important role in the development of more complicated theories. The theory of star formation has proceeded in three stages. First, order of magnitude estimates based on fundamental physics yielded information regarding, for example, timescales and the necessity of certain physical processes. These have led to the development of simple analytical models, which have, along with linear perturbation analysis, offered a simplified look at the dynamics of star forming regions (e.g., Jeans 1902; Mouschovias 1976; Shu 1977). Observations showed that sometimes the approximations of these models were reasonable, but sometimes they were not. Eventually there came a point where many of these analytical models could no longer grow in sophistication. From here, star formation theorists have turned to numerical simulations. The computational capabilities and techniques of numerical simulations themselves have also developed over the past few decades (compare, e.g., Foster & Chevalier 1993; Myers et al. 2014), and there is little sign that this rate of development is slowing down (Klein et al. 2007; Dale 2015). Today, numerical simulations continue to unveil to nature of stellar nurseries across all of cosmic time.

This dissertation explores models of the star and planet formation process through the use of numerical simulations. The next section briefly outlines the star and planet formation process and identifies several unanswered questions of the field. The final section outlines the remaining chapters of this thesis and how it will address these questions. Section 1.1 is not meant as an exhaustive or pedagogical review; for that, the reader should seek out the many excellent reviews and texts on the subject (e.g., Krumholz 2014; McKee & Ostriker 2007; Klessen et al. 2011; Kennicutt & Evans 2012; Stahler & Palla 2005; Bodenheimer 2011).

1.1 Star and Planet Formation

1.1.1 Contemporary Environments

One could joke that the answer to almost every question in astrophysics is ‘gravity,’ but this is indeed the case for star formation. In the Milky Way and local galaxies, stars form via gravitational collapse in $\sim 1-100$ pc-sized molecular clouds. These clouds are composed predominately of molecular hydrogen and range in mass from a paltry $10^2\,M_\odot$ (e.g., Magnani et al. 1985) all the way up to $10^7\,M_\odot$ (e.g., Oka et al. 2001). Molecular clouds are located primarily in the spiral arms of disk-like galaxies (e.g., the Milky Way; Blitz & Rosolowsky 2005), in starburst galaxies (e.g., M82), irregular galaxies (e.g., the Magellanic Clouds), and merging or interacting galaxies (e.g., Antennae).

The nature of these clouds has been revealed through decades of observational work. Local clouds exhibit broad linewidths, which is interpreted as evidence for supersonic turbulence (Zuckerman & Evans 1974). Turbulence is a result of kinetic energy cascading from large scales down to small scales, and its origin likely ties to the clouds' formation process (Dobbs et al. 2014). While the bulk of the energy remains at cloud-size scales, turbulence appears to be dynamically important all the way down to the sub-parsec scales where individual star formation is occurring (Larson 1981; Mac Low & Klessen 2004).

Observations using Zeeman splitting and the Chandrasekhar-Fermi effect also show that these clouds are threaded by magnetic fields (Crutcher 1999; Troland & Crutcher 2008; Lai et al. 2002) with strengths on order of a few tens of μG. These strengths make magnetic fields potentially an important player in the star formation process, as measured by the ratio of the thermal pressure to the magnetic pressure (β): clouds typically have $\beta < 1$ (Crutcher 1999). While magnetic pressures exceed thermal pressures, a virial analysis of local clouds finds that clouds are typically magnetically supercritical, meaning that the mass of the cloud exceeds the minimum mass necessary for gravitational collapse to occur in the presence of opposing magnetic forces (Crutcher et al. 2003). Even so, as collapse proceeds and magnetic flux is conserved, magnetic fields likely play an important role in how mass is transferred from the cloud to the star.

Molecular cloud temperatures are determined by the balance between heating and cooling mechanisms, both which depend on the chemical and dust content of clouds. Clouds are observed to be ~ 10 Kelvin at densities below $\sim 10^{11}$ cm^{-3}, a result of heating from either cosmic rays or photoelectric heating and cooling from line transitions (e.g., from CO, H_2O, CII, O) or thermal dust emission (e.g., Jijina et al. 1999; Evans 1999; Larson 2005). Above this number density–reached near localized sites of star formation–the dust opacity increases and the temperature rises rapidly.

The interplay of these turbulent, magnetic, and thermal components generate the observed clumpy and often filamentary structure of molecular clouds (e.g., Ward-Thompson et al. 2010). Inside these filaments live dense cores of gas with distribution of masses (e.g., André et al. 2010, for Aquila). Cores that eventually become gravitationally unstable collapse into stars.

In particular, the larger cores are likely the formation sites for the massive OB stars ($\gtrsim 8\,M_\odot$) seen in star clusters. Massive stars play a dominant role in terms of their incredible mechanical and radiative feedback on their local environments. They also synthesize and eventually disperse the heavier elements of the periodic table. Two theories of massive star formation are under active study. The core accretion model extends the model of low-mass stars, where individual or few-N multiples of stars form from the collapse of the molecular core (Shu et al. 1987). The competitive accretion model (Bonnell et al. 2001) posits that low-mass stars form in these cores instead, but then accrete as they roam through the clumps and the wider, non self-gravitating region. Competitive accretion models require that stars can efficiently accrete gas from their surroundings. Further investigation is necessary to assess whether or not this process is thwarted by the presence of magnetic fields.

During the collapse of molecular cores, conservation of angular momentum prevents all the material from falling directly onto the star. Instead, it falls onto a surrounding flattened disk (e.g., Terebey et al. 1984; Watson et al. 2007). The central star continues to accrete material from this disk, but sufficiently heavy disks can further fragment into smaller protostars or brown dwarfs (Toomre 1964; Dullemond et al. 2007). The remaining disk material can eventually form a planetary system.

Early observations of stars emerging from their dense core cocoons revealed an excess of near- and mid-infrared radiation, which was interpreted as the presence of a circumstellar disk. Many circumstellar disks have also been spatially resolved (e.g., White et al. 2016), seen in the infrared and radio from micron-sized dust emission. The process by which dust becomes a planet occurs in three stages. First, collisions via Brownian motion and chemical bonding gently builds centimeter- to meter-sized objects. These objects eventually become planetesimals, self-gravitating objects that are typically kilometer in size. Planetesimals then coalesce via gravitational focusing to form the rocky planets and gas gaint cores (Goldreich et al. 2004a; Kokubo & Ida 1996). The most challenging step is the second one, when the solids are $\sim$meters in size. Their surfaces are insufficiently sticky and their individual gravity is too feeble to grow in size through collisions, and strong aerodynamic drag from the gas causes radial drifts that are too quick to form planets (Chiang & Youdin 2010). The presence of planets indicates that this challenge can be overcome, but the exact mechanism to do so remains disputed.

1.1.2 Primordial Environments

The fact that star formation occurs suggests there was a point in time when the very first stars formed. These first stars (so called Population III stars, or Pop III stars) marked a fundamental transition in the history of the universe. Ending the so-called 'cosmic dark ages,' when the universe contained no visible light, they lit up the universe at redshifts $\gtrsim 20$ (Bromm & Larson 2004; Glover 2005; Bromm 2013). They initiated the transformation of the homogeneous intergalactic medium (IGM) to one filled with the rich structure we observe today (Barkana & Loeb 2001). Sufficiently massive Pop III stars carved out the first HII regions (e.g., Kitayama et al. 2004; Alvarez et al. 2006; Johnson et al. 2007a), left behind the first metal-enriched SN remnants (e.g., Mori et al. 2002; Bromm et al. 2002; Greif et al. 2010; Wise et al. 2012; Karlsson et al. 2008), or collapsed to some of the first gamma-ray bursts (Woosley 1993; Stacy et al. 2011, 2013).

These stars are believed to have formed in the first gravitationally-bound dark matter 'minihalos' that developed from the cosmological large-scale structure. With virial masses of $M_{\rm vir} \sim 10^6 M_\odot$, these halos pulled nearby baryonic gas into their gravitational potential well. A gravitationally-bound core of pristine $\sim 10^3\, M_\odot$ gas settled to the center of each halo, and this core provided the reservoir from which primordial protostars formed and grew. Though initially too hot to collapse, and devoid of coolants made of elements heavier than helium, these halos formed a sufficient abundance of H_2 such that the energy release due to rovibrational transitions cooled the halos down to ~ 200 Kelvin, allowing for eventual collapse (e.g., Haiman et al. 1997; Tegmark et al. 1997; Yoshida et al. 2003). The higher temperatures result in a higher Jeans mass, and the higher Jeans mass implies that these stars experienced high accretion rates ($\sim 10^{-3} M_\odot/\mathrm{yr}$). These stars are believed to be much more massive than typical stars today, perhaps hundreds of solar masses in size (Abel et al. 1997, 2002; Bromm et al. 2002).

Can the first stars be this large? Perhaps, but gravitational collapse does not proceed unimpeded in these halos. While cosmological magnetic fields are too weak to be dynamically relevant at these early stages (though dynamo fields may eventually develop, Turk et al. 2012), the forming protostars radiate considerable energy back into their environments. Comparing the accretion timescale to the Kelvin-Helmholtz timescale for Pop III stars, they are likely to undergo Kelvin-Helmholtz contraction above $\sim 10\,M_\odot$ (Hosokawa et al. 2010a). As these stars contract, their surface temperatures increase so to produce considerable hydrogen-ionizing radiation (energies exceeding 13.6 eV). Ionization raises the local temperature of the gas to $\sim 10^4$ Kelvin, and this highly pressured region will begin to drive gas from the halo once it can escape from the nearby star(s). Tan & McKee (2004) considers a 1D model of one of these stars forming in isolation. Once the star's mass exceeds $\sim 40\,M_\odot$ it will begin to develop an HII region that reduces and eventually shuts off accretion above and below its accretion disk. Within a few $\times 10^4$ years, this star will grow beyond $100\,M_\odot$, at which point the radiative feedback will begin photoevaporating the accretion disk (Hollenbach et al. 1994), further reducing the accretion rate. McKee & Tan (2008) extend the 1D model of Tan & McKee (2004) and show that accretion is effectively shut off once the star exceeds $\sim 150\,M_\odot$, though this value depends on several assumptions about the host cloud. The 2D simulations of Hosokawa et al. (2011) and the 3D simulations of Stacy et al. (2012) have confirmed this simple picture, at least for the early times of the star's growth. However, their stars reached a final mass that was tens of solar masses compared to hundreds, which may have been due to poor resolution near the protostars. The work of Hosokawa et al. (2011) was extended by Hirano et al. (2014), which simulated a large variety of primordial environments. They found that Pop III stars can range in mass from $10\,M_\odot$ to $10^3\,M_\odot$, and the final stellar masses and multiplicity not only depended on the radiative feedback, but also depended sensitively on the thermal evolution of the cloud during the initial collapse phase. Additional simulations are needed to assess the role that ionizing radiative feedback plays in the Pop III star formation process.

1.2 Outline

I explore the formation of stars and planets in several chapters. Chapters 2 and 3 explores how micron-sized dust accumulates into km-sized planetesimals. These chapters show that where the individual fails the collective succeeds: finer mm-sized particles can settle to the midplane of the protoplanetary disk so that their collective gravity can draw them together, forming planetesimals. This mechanism skips the centimeter- to meter-sized regime entirely, avoiding the concern of rapid radial drift due to aerodynamic drag. However, this mechanism requires that the thin dust sub-layer remains stable as the dust settles, as velocity gradients off the disk midplane can trigger Kelvin-Helmholtz instabilities. This was studied through shearing box simulations employing spectral methods (Barranco 2009). In early stages of settling, vertical perturbations are transformed into radial perturbations through the Coriolis force, which are then sheared away by disk's radial velocity gradient. As the dust settles and the vertical shear increases in strength, the dust withstands if either (1) there is locally more dust compared to gas, which can occur globally, through radial pileups, or gas photoevaporation; or (2) planet formation is not 100% efficient and the total disk mass is at least a few times the minimum mass required to form the Solar System. Support for our model comes from observations: low-mass stars with exoplanets tend to be metal (dust)-rich, and high-mass stars likely host more massive disks (Johnson et al. 2010).

Chapter 4 investigates how massive stars may get their mass. If stars grow through the competitive accretion scenario, stars should be able to effectively accrete gas from their (magnetized) surroundings. Additionally, low mass stars may also accrete from the molecular cloud after they have formed, albeit at a much lower rate. How stars accrete from a non self-gravitating magnetized gas was studied using the adaptive mesh refinement (AMR) code RAMSES (Teyssier 2002). Placing a stationary sink particle at the center of the box, initially uniform magnetized gas was allowed to flow past the particle at some sonic Mach number $\mathcal{M}$. A range of magnetic field strengths are considered, as well as flow speeds ranging from stationary to highly supersonic. For $\mathcal{M} \sim 1$ and $\beta \sim 0.04$, the magnetic accretion rate to the hydrodynamic rate is $\sim 1/10$. This ratio decreases for lower $\mathcal{M}$ and β. Competitive accretion models therefore underestimate the timescale required for making massive stars.

Finally, Chapter 5 investigates how the star formation process occurred for the first generation of stars. This chapter describes the preliminary work done to improve our knowledge on the Pop III mass function. Using the adaptive mesh refinement hydrodynamic code ORION2 (Li et al. 2012), we simulate the formation of Pop III stars in a dark matter minihalo. This simulation is the first ORION2 simulation to incorporate the hydrogen-ionizing radiative feedback from the forming stars (Appendix C). Initial conditions are drawn from the spherical cloud models of Tan & McKee (2004) and McKee & Tan (2008) and monitor the growth of the HII region once the star exceeds $\sim 200 M_\odot$. Future work will draw from the cosmological smoothed particle hydrodynamics (SPH) simulations of Stacy & Bromm (2013). As of this writing, we have already developed a Monte Carlo routine that transforms SPH data into AMR data; we describe this routine in Appendix D. This routine conserves all fluid quantities to better than 0.01%, whereas previous routines written to convert SPH to AMR artificially conserved mass by applying a uniform correction factor to each particle (inconsistent with its non-uniform smoothing kernel).

Chapter 2

Forming Planetesimals by Gravitational Instability: I. The Role of the Richardson Number in Triggering the Kelvin-Helmholtz Instability

[1] Gravitational instability (GI) of a dust-rich layer at the midplane of a gaseous circumstellar disk is one proposed mechanism to form planetesimals, the building blocks of rocky planets and gas giant cores. Self-gravity competes against the Kelvin-Helmholtz instability (KHI): gradients in dust content drive a vertical shear which risks overturning the dusty subdisk and forestalling GI. To understand the conditions under which the disk can resist the KHI, we perform three-dimensional simulations of stratified subdisks in the limit that dust particles are small and aerodynamically well coupled to gas, thereby screening out the streaming instability and isolating the KHI. Each subdisk is assumed to have a vertical density profile given by a spatially constant Richardson number Ri. We vary Ri and the midplane dust-to-gas ratio μ_0 and find that the critical Richardson number dividing KH-unstable from KH-stable flows is not unique; rather $Ri_{\rm crit}$ grows nearly linearly with μ_0 for $\mu_0 = 0.3$–10. Plausibly a linear dependence arises for $\mu_0 \ll 1$ because in this regime the radial Kepler shear replaces vertical buoyancy as the dominant stabilizing influence. Why this dependence should persist at $\mu_0 > 1$ is a new puzzle. The bulk (height-integrated) metallicity is uniquely determined by Ri and μ_0. Only for disks of bulk solar metallicity is $Ri_{\rm crit} \approx 0.2$, close to the classical value. Our empirical stability boundary is such that a dusty sublayer can gravitationally fragment and presumably spawn planetesimals if embedded within a solar metallicity gas disk ~4× more massive than the minimum-mass solar nebula; or a minimum-mass disk having ~3× solar metallicity; or some intermediate combination of these two possibilities. Gravitational instability seems possible without resorting

---

[1] Large portions of this chapter have been previously published as Lee, A. T., Chiang, E., Asay-Davis, X., Barranco, J. 2010. Forming Planetesimals by Gravitational Instability. I. The Role of the Richardson Number in Triggering the Kelvin-Helmholtz Instability. ApJ, 718, 1367.

to the streaming instability or to turbulent concentration of particles.

2.1 Introduction

In the most venerable scenario for forming planetesimals, dust particles in circumstellar gas disks are imagined to settle vertically into thin sublayers ("subdisks") sufficiently dense to undergo gravitational instability (Safronov 1969; Goldreich & Ward 1973; for a review of this and other ways in which planetesimals may form, see Chiang & Youdin 2010, hereafter CY10). Along with this longstanding hope comes a longstanding fear that dust remains lofted up by turbulence. Even if we suppose that certain regions of the disk are devoid of magnetized turbulence because they are too poorly ionized to sustain magnetic activity (Gammie 1996; Bai & Goodman 2009), the dusty sublayer is susceptible to a Kelvin-Helmholtz shearing instability (KHI; Weidenschilling 1980).[2]

2.1.1 Basic Estimates

The KHI arises because dust-rich gas at the midplane rotates at a different speed from dust-poor gas at altitude. The background radial pressure gradient $\partial P/\partial r$ causes dust-free gas at disk radius r to rotate at the slightly non-Keplerian rate

$$\Omega_{\rm F} = \Omega_{\rm K}(1-\eta) \tag{2.1}$$

where $\Omega_{\rm K}$ is the Kepler angular frequency,

$$\eta = \frac{-(1/\rho_{\rm g})\partial P/\partial r}{2\Omega_{\rm K}^2 r} \approx 8\times 10^{-4}\left(\frac{r}{\rm AU}\right)^{4/7} \tag{2.2}$$

is a dimensionless measure of centrifugal support by pressure, and $\rho_{\rm g}$ is the density of gas (e.g., Nakagawa et al. 1986; Cuzzi et al. 1993). The numerical evaluation is based on the minimum-mass solar nebula derived by CY10. Unlike dust-free gas, dust-rich gas is loaded by the extra inertia of solids and must rotate at more nearly the full Keplerian rate to remain in centrifugal balance. Variations in the dust-to-gas ratio $\rho_{\rm d}/\rho_{\rm g}$ with height z result in a vertical shear $\partial v_\phi/\partial z$ from which free energy is available to overturn the dust layer.

The shearing rate across a layer of thickness Δz is given to order of magnitude by

$$\begin{aligned}\left|\frac{\partial v_\phi}{\partial z}\right| \sim \frac{\Delta v_\phi}{\Delta z} &= \frac{1}{\Delta z}\frac{\mu_0}{1+\mu_0}\eta\Omega_{\rm K} r \\ &\approx \frac{25}{\Delta z}\frac{\mu_0}{1+\mu_0}\left(\frac{r}{\rm AU}\right)^{1/14}\ {\rm m\,s^{-1}} \qquad (2.3)\end{aligned}$$

where $\rho_{\rm d}/\rho_{\rm g} = \mu_0$ at the midplane and $\rho_{\rm d}/\rho_{\rm g} \ll 1$ at altitude (for more details see CY10, or §§2.2.1–2.2.2 of this chapter). For $\mu_0 \gg 1$ the velocity difference Δv_ϕ saturates at a speed

---

[2] Goldreich & Ward (1973) also recognized that the sublayer would be shear-unstable, but unlike Weidenschilling (1980), overlooked the possibility that the KHI may forestall gravitational instability.

$\eta\Omega_{\rm K} r \sim 25(r/{\rm AU})^{1/14}\,{\rm m\,s^{-1}}$, well below the gas sound speed $c_{\rm s} \sim 1\,{\rm km\,s^{-1}}$. That the flow is highly subsonic motivates what simulation methods we employ in our study.

We might expect the flow to be stabilized if the Brunt-Väisälä frequency

$$\begin{aligned} \omega_{\rm Brunt} &= \left(\frac{-g}{\rho}\frac{\partial\rho}{\partial z}\right)^{1/2} \\ &\sim \left(\frac{\mu_0}{1+\mu_0}\right)^{1/2}\Omega_{\rm K} \end{aligned} \tag{2.4}$$

of buoyant vertical oscillations is much larger than the vertical shearing rate. For the order-of-magnitude evaluation in (2.4) we approximate the vertical gravitational acceleration g as the vertical component of stellar gravity $-\Omega_{\rm K}^2\Delta z$ (no self-gravity), and the density gradient $\rho^{-1}\partial\rho/\partial z \sim (\rho_{\rm d}+\rho_{\rm g})^{-1}\Delta(\rho_{\rm d}+\rho_{\rm g})/\Delta z \sim (\rho_{\rm d}+\rho_{\rm g})^{-1}\Delta\rho_{\rm d}/\Delta z$. The last approximation relies in part on the dust density $\rho_{\rm d}$ changing over a lengthscale Δz much shorter than the gas scale height. Both $|\partial v_\phi/\partial z|$ and $\omega_{\rm Brunt}$ shrink as μ_0 decreases.[3]

For two-dimensional, heterogeneous, unmagnetized flow, a necessary but not sufficient condition for instability is given by the Richardson number:

$$Ri \equiv \frac{-(g/\rho)(d\rho/dz)}{(dv_\phi/dz)^2} < 1/4 \text{ is necessary for instability} \tag{2.5}$$

(Miles 1961; see the textbook by Drazin & Reid 2004). The Richardson number is simply the square of the ratio of the stabilizing Brunt frequency (2.4) to the destabilizing vertical shearing frequency (2.3). The critical value of 1/4 arises formally but can also be derived heuristically by energy arguments (e.g., Chandrasekhar 1981). The Richardson criterion does not formally apply to our dusty subdisk, which represents a three-dimensional flow: the KHI couples vertical motions to azimuthal motions, while the Coriolis force couples azimuthal motions to radial motions. (For how the Richardson criterion may not apply to magnetized flows, see Lecoanet et al. 2010.) Nevertheless we may hope the Richardson number is useful as a guide, as previous works have assumed (Sekiya 1998; Youdin & Shu 2002; Youdin & Chiang 2004).

In this spirit let us use the Richardson criterion to estimate the thickness of a marginally KH-unstable dust layer. Substitution of (2.3) and (2.4) into (2.5) reveals that[4]

$$\Delta z \approx \left(\frac{\mu_0}{1+\mu_0}\right)^{1/2} Ri^{1/2}\eta r\,. \tag{2.6}$$

---

[3] But not indefinitely. In the limit $\mu_0 \to 0$, the vertical shearing and Brunt frequencies reach minima set by pressure and temperature gradients in gas (see, e.g., Knobloch & Spruit 1985). The limit $\mu_0 \to 0$ is not relevant for our study and not captured by either (2.3) or (2.4).

[4] This order-of-magnitude expression for the dust layer thickness, and the related equation (2.3) which approximates the vertical shear, are each smaller than their counterparts given by Youdin & Shu (2002, page 499, first full paragraph) by a factor of $(1+\mu)$. This is because Youdin & Shu (2002) evaluate quantities deep inside the layer, within a density cusp at the midplane, whereas we are interested in quantities averaged across the entire layer. The difference does not change either our conclusions or theirs.

Since the gas scale height $H_{\rm g} = c_{\rm s}/\Omega_{\rm K}$ and $\eta \sim (H_{\rm g}/r)^2$, equation (2.6) indicates that for $\mu_0 > 1$ the marginally unstable dust sublayer is $\sim Ri^{1/2} H_{\rm g}/r \sim 0.02 Ri^{1/2}$ times as thin as the gas disk in which it is immersed. Those KH-unstable modes that disrupt the layer should have azimuthal wavelengths—and by extension radial wavelengths, because the Kepler shear turns azimuthal modes into radial ones—that are comparable to Δz. Shorter wavelength modes cannot overturn the layer, while longer wavelength modes grow too slowly (Gómez & Ostriker 2005).

How does disk rotation affect the development of the KHI? In a linear analysis, Ishitsu & Sekiya (2003) highlight the role played by the Keplerian shear, characterized by the strain rate

$$\left|\frac{\partial \Omega_{\rm K}}{\partial \ln r}\right| = \frac{3\Omega_{\rm K}}{2}\,, \tag{2.7}$$

in limiting the growth of KH-unstable modes. The radial shear is implicated because azimuthal motions excited by the KHI are converted to radial motions by the Coriolis force; moreover, the non-axisymmetric pattern excited by the KHI is wound up, i.e., stretched azimuthally by the radial shear. The Kepler rate $|\partial\Omega_{\rm K}/\partial \ln r|$ is at least as large as $\omega_{\rm Brunt}$, and can dominate the latter when μ_0 is small. This suggests that Ri does not capture all the relevant dynamics—a concern already clear on formal grounds. In this chapter we address this concern head-on, using fully 3D numerical simulations to assess the role of the Richardson number in governing the stability of the dust layer.

2.1.2 Our Study in Relation to Previous Numerical Simulations

Three-dimensional shearing box simulations of the KHI in dusty subdisks, performed in the limits that dust is perfectly coupled to gas and disk self-gravity is negligible, demonstrate the importance of the Kepler shear. Compared to rigidly rotating disks (Gómez & Ostriker 2005; Johansen et al. 2006), radially shearing disks are far more stable (Chiang 2008; Barranco 2009). The relevance of Ri, or lack thereof, may be assessed by simulating flows with initially spatially constant Ri (Sekiya 1998; Youdin & Shu 2002), and varying Ri from run to run to see whether dust layers turn over. Chiang (2008, hereafter C08) found that when $\mu_0 > 1$, dust layers for which $Ri < 0.1$ overturn, while those for which $Ri > 0.1$ do not. In retrospect, we might have anticipated this result, that the critical value $Ri_{\rm crit}$ dividing stable from unstable runs lies near the canonical value of 1/4, at least for $\mu_0 > 1$, because in this regime of parameter space all the frequencies of the problem are comparable to each other: $|\partial v_\phi/\partial z| \sim \omega_{\rm Brunt} \sim |\partial\Omega_{\rm K}/\partial \ln r| \sim \Omega_{\rm K}$ when $\mu_0 > 1$ and $Ri \approx 0.1$–1. But other simulations of C08 also make clear that Ri does not alone determine stability under all circumstances. For $\mu_0 \approx 0.2$–0.4, $Ri_{\rm crit}$ was discovered to drop substantially to $\sim$0.02 (see his runs S9–S12). Chiang (2008) speculated that the baroclinic nature of the flow may be responsible (Knobloch & Spruit 1985, 1986), but no details were given.

In addition to being left unexplained, the findings of C08 require verification. Parameter space was too sparsely sampled to discern trends with confidence. Concerns about numerics—e.g., biases introduced by box sizes that were too small, resolutions too coarse, and runs terminated too early—also linger. At least one numerical artifact marred the simulations of C08: the KHI manifested first at the "co-rotation" radius where the mean azimuthal flow speed was zero (see his figure 8). But in a shearing box, by Galilean invariance, there should be no special radius. It was suspected, but not confirmed, that errors of interpolation associated with the grid-based advection scheme used by C08 artificially suppressed the KHI away from co-rotation.

For the problem at hand, the spectral code developed by Barranco & Marcus (2006) and modified by Barranco (2009, hereafter B09) to treat mixtures of dust and gas is a superior tool to the grid-based ZEUS code utilized by C08. Working in Fourier space rather than configuration space, the simulations of the KHI by B09 did not betray the co-rotation artifact mentioned above. Spectral methods, often used to model local (WKB) dynamics, are appropriate here because the structures of interest in the subdisk have dimensions tiny compared to the disk radius (by at least a factor $Ri^{1/2}\eta$ according to equation 2.6) and even the gas scale height. At the same computational expense, spectral algorithms typically achieve greater effective spatial resolution than their grid-based counterparts (Barranco & Marcus 2006). Another advantage enjoyed by the B09 code is that it employs the anelastic approximation, which is designed to treat subsonic flows such as ours. Having filtered away sound waves, anelastic codes are free to take timesteps set by how long it takes fluid to advect across a grid cell (which themselves move at the local orbital velocity in a shearing coordinate system). By contrast, codes such as ZEUS take mincing steps limited by the time for sound waves to cross a grid cell. The latter constraint is the usual Courant condition for numerically solving problems in compressible fluid dynamics. It was unnecessarily applied by C08 to a practically incompressible flow.

In this chapter we bring all the advantages of the spectral, anelastic, shearing box code of B09 to bear on the problems originally addressed by C08. We assess numerically the stability of flows characterized by constant Richardson number Ri, systematically mapping out the stability boundary in the parameter space of Ri, midplane dust-to-gas ratio μ_0, and bulk metallicity $\Sigma_\mathrm{d}/\Sigma_\mathrm{g}$ (the height-integrated surface density ratio of dust to gas). Though our simulations may still be underresolved, we rule out box size as a major influence on our results. We offer some new insight into why Ri is not a sufficient predictor of stability. And in the restricted context of our constant Ri flows, we assess the conditions necessary for the midplane to become dense enough to trigger gravitational instability on a dynamical time.

2.1.3 The Perfect Coupling Approximation vs. The Streaming Instability vs. Turbulent Concentration Between Eddies

Following C08 and B09, we continue to work in the limit that dust is perfectly coupled to gas, i.e., in the limit that particles are small enough that their frictional stopping times t_stop in gas can be neglected in comparison to the dynamical time Ω_K^{-1}. The perfect coupling approximation allows us to screen out the streaming instability which relies on a finite stopping time and which is most powerful when particles are marginally coupled, i.e., when $\tau_\mathrm{s} \equiv \Omega_\mathrm{K} t_\mathrm{stop} \sim 0.1$–$1$ (Youdin & Goodman 2005). Numerical simulations have shown that when an order-unity fraction of the disk's solids is in particles having $\tau_{\rm s} = 0.1$–1, the streaming instability clumps them strongly and paves the way for gravitational instability (e.g., Johansen et al. 2007, 2009). The particle sizes corresponding to $\tau_{\rm s} = 1$ depend on the properties of the background gas disk, as well as on the particle's shape and internal density; under typical assumptions, marginally coupled particles are decimeter to meter-sized.

It remains debatable whether a substantial fraction of a disk's solid mass is in marginally coupled particles at the time of planetesimal formation, as current proposals relying on the streaming instability assume. Particle size and shape distributions are not well constrained by observations (though see, e.g., Wilner et al. 2005, who showed that centimeter-wavelength fluxes from a few T Tauri stars are consistent with having been emitted by predominantly centimeter-sized particles). Measuring $\tau_{\rm s}$ in disks also requires knowing the gas density, but direct measurements of the gas density at disk midplanes do not exist. Marginally coupled particles—sometimes referred to as "meter-sized boulders"—also face the longstanding problem that they drift onto the central star too quickly, within hundreds of years from distances of a few AU in a minimum-mass disk. Johansen et al. (2007) claimed to solve this problem by agglomerating all the boulders into Ceres-mass planetesimals via the streaming instability before they drifted inward. Their simulation presumed, however, that all of the disk's solids began boulder-sized. The concern we have is that even if particle-particle sticking could grow boulders (and sticking is expected to stall at centimeter sizes; Blum & Wurm 2008; CY10), the disk's solids may not be transformed into boulders all at once. Rather, marginally coupled bodies may initially comprise a minority population on the extreme tail of the particle size distribution. Unless they can transform themselves from a minority to a majority within the radial drift timescale, they would be lost from the nebula by aerodynamic drag.

By focussing on the dynamics of the smallest, most well entrained particles having $\tau_{\rm s} \ll 1$, our work complements that which relies on the streaming instability. We would argue further that the well coupled limit is potentially more relevant for planet formation. If even the smallest particles having sizes $\ll$ cm can undergo gravitational collapse to form kilometer-sized or larger planetesimals, nature will have leapfrogged over the marginally coupled regime, bypassing the complications and uncertainties described above.

Particle clumping is not restricted to marginally coupled particles via the streaming instability. Small $\tau_{\rm s}$ particles also clump within the interstices of turbulent, high vorticity eddies (Maxey 1987; Eaton & Fessler 1994; Cuzzi et al. 2008, and references therein; for a review, see CY10). This particle concentration mechanism presumes some gas turbulence, which may be present in the marginally KH-unstable state to which dust settles. Our simulations cannot capture this phenomenon. However, on the scales of interest to us, turbulent clumping might only be of minor significance. Particles of given $t_{\rm stop}$ are concentrated preferentially by eddies that turn over on the same timescale. Thus the degree of concentration depends sensitively on particle size and the turbulent spectrum. At least in Kolmogorov turbulence, the smallest eddies concentrate particles most strongly because they have the greatest vorticity. The smallest eddies at the inner scale of Kolmogorov turbulence have sizes $\ell_{\rm i} \sim \nu^{3/4} t_{\rm o}^{1/4} / \delta v_{\rm o}^{1/2}$, where ν is the molecular kinematic viscosity, and $t_{\rm o}$ and $\delta v_{\rm o}$ are the turnover time and velocity of the largest, outer scale eddy. Given $\delta v_{\rm o} \sim \eta \Omega_{\rm K} r \sim 25 (r/\mathrm{AU})^{1/14}$ m/s, $t_{\rm o} \sim \Omega_{\rm K}^{-1}$, and values of ν based on the nebular model of CY10, we estimate that $\ell_{\rm i} \sim 10^3 (r/{\rm AU})^{127/56}$ cm. This is far smaller than the sublayer thicknesses $\Delta z \sim 0.02 Ri^{1/2} H_{\rm g} \sim 2 \times 10^9 (Ri/0.1)^{1/2} (r/{\rm AU})^{9/7}$ cm considered in this chapter. Moreover, the lifetimes of the particle clumps on a given eddy length scale should roughly equal the eddy turnover times, which for the smallest eddies are of order $t_{\rm i} \sim \sqrt{\nu t_{\rm o}}/\delta v_{\rm o} \sim 10^2 (r/{\rm AU})^{55/28}$ s. We do not expect such rapid fluctuations in particle density, occurring on such small length scales, to affect significantly the evolution of the slower, larger scale KHI. Turbulent clumping may only serve as a source of noise on tiny scales. The possibility that turbulent clumping could still be significant on larger scales is still being investigated (Hogan & Cuzzi 2007; Cuzzi et al. 2008).

The perfect coupling approximation prevents us from studying how particles sediment out of gas into dusty sublayers, but it does not stop us from identifying what kinds of sublayers are dynamically stable to the KHI. A subdisk with a given density profile is either dynamically stable or it is not, and we can run the B09 code for many dynamical times (typically 60 or more) to decide the answer. In the next chapter we will combine the B09 code with a settling algorithm that will permit us to study how dust settles from arbitrary initial conditions, freeing us from the assumption that the density profile derives from a constant Richardson number.

2.1.4 Organization of this Chapter

Our numerical methods, including our rationale for choosing box sizes and resolutions, are described in §2.2. Results are presented in §2.3 and discussed in §2.4.

2.2 Methods

The equations solved by the B09 code are rederived in §2.2.1. Initial conditions for our simulations are given in §2.2.2. The code itself is briefly described in §2.2.3. Our choices for box size and resolution are explained in §2.2.4.

2.2.1 Equations

The equations we solve are identical to equations (12a–e) of B09. We outline their derivation here, filling in steps skipped by B09, adjusting the notation, and providing some clarifications. This section may be skimmed on a first reading.

We begin with the equations for an ideal gas perfectly coupled to pressureless dust in an inertial frame:

$$\begin{aligned}\frac{d\mathbf{v}}{dt} &= -\nabla\Phi - \frac{\nabla P}{\rho_d+\rho_g}, && (2.8)\\ \frac{d\rho_{\rm g}}{dt} &= -\rho_g\nabla\cdot\mathbf{v}, && (2.9)\\ \frac{d(\rho_{\rm d}/\rho_{\rm g})}{dt} &= 0, && (2.10)\\ \rho_g C_{\rm V}\frac{dT}{dt} &= -P\nabla\cdot\mathbf{v}, && (2.11)\\ P &= \Re\rho_{\rm g}T, && (2.12)\end{aligned}$$

where d/dt is the convective derivative, $\rho_{\rm g(d)}$ is the density of gas (dust), P is the gas pressure, and T is the gas temperature. Under the assumption that they are perfectly coupled, gas and dust share the same velocity $\mathbf{v}$, and the dust-to-gas ratio is conserved in a Lagrangian sense. The background potential is provided by the central star of mass M: $\Phi = -GM/\sqrt{r^2+z^2}$, where r is the cylindrical radius and z is the vertical distance above the disk midplane. There are five equations for the five flow variables $\mathbf{v}$, $\rho_{\rm g}$, $\rho_{\rm d}$, P, and T. The thermodynamic constants include the specific heat $C_{\rm V} = \Re/(\gamma-1)$ at constant volume, the ideal gas constant $\Re = C_{\rm P} - C_{\rm V}$, the specific heat $C_{\rm P}$ at constant pressure, and $\gamma = C_{\rm P}/C_{\rm V}$. Equation (2.11) is equivalent to the condition that the flow be isentropic $[d(P\rho_{\rm g}^{-\gamma})/dt = 0]$. The code which solves the fluid equations actually employs an artificial hyperviscosity to damp away the smallest scale perturbations (§2.2.3); in writing down equations (2.8)–(2.12), we have omitted the hyperviscosity terms for simplicity.

We move to a frame co-rotating with dust-free gas at some fiducial radius $r = R$. This frame has angular frequency $\Omega_{\rm F}$ given by (2.1) with $\Omega_{\rm K} = (GM/R^3)^{1/2}$. We define a velocity $v_{\rm max}$ using the pressure support parameter η, as given by (2.2):

$$v_{\rm max} \equiv \eta|_{r=R}\,\Omega_{\rm K}R. \tag{2.13}$$

The velocity $v_{\rm max}$ is the difference in azimuthal velocity between a strictly Keplerian flow and dust-free gas; it is the maximum possible difference in velocity, attained at large μ_0, between gas at the midplane and gas at altitude. The quantities $v_{\rm max}$, η, and the background radial pressure gradient are equivalent; specifying one specifies the other two. Our numerical models are labeled by $v_{\rm max}$.

In addition to moving into a rotating frame, we also replace the usual cylindrical coordinates (r,ϕ,z) with local Cartesian coordinates $x = r - R$, $y = (\phi - \Omega_{\rm F}t)R$, and z.[5] Keeping terms to first order in $|x| \sim |z| \sim \eta R$ (see the discussion surrounding equation 2.6) and dropping curvature terms, the momentum equation (2.8) reads

$$\frac{d\mathbf{v}}{dt} = -2\Omega_{\rm K}\hat{\mathbf{z}}\times\mathbf{v} + 3\Omega_{\rm K}^2 x\hat{\mathbf{x}} - \Omega_{\rm K}^2 z\hat{\mathbf{z}} - \frac{1}{\rho_{\rm d}+\rho_{\rm g}}\nabla P - 2\Omega_{\rm K}^2\eta R\hat{\mathbf{x}} \tag{2.14}$$

---

[5] Throughout this chapter we alternate freely between subscripts (x,y,z) and (r,ϕ,z).

where $d/dt = \partial/\partial t + v_i \partial/\partial x_i$ ($i = x, y, z$). On the right-hand side, the first term is the Coriolis acceleration, the second combines centrifugal and radial gravitational accelerations, the third represents the vertical gravitational acceleration from the star, and the last term arises from the centrifugal acceleration in a frame rotating at $\Omega_{\mathrm{F}} \neq \Omega_{\mathrm{K}}$. The remaining fluid equations appear the same as (2.9)–(2.12), except that $\mathbf{v}$ is now measured in a (rigidly) rotating frame.

We measure all flow variables relative to a time-independent reference state (subscripted "ref"):

$$\begin{aligned}
\mathbf{v} &= \mathbf{v}_{\mathrm{ref}} + \widetilde{\mathbf{v}} = \widetilde{\mathbf{v}} \\
P &= P_{\mathrm{ref}} + \widetilde{P} \\
\rho_{\mathrm{g}} &= \rho_{\mathrm{g,ref}} + \widetilde{\rho}_{\mathrm{g}} \\
T &= T_{\mathrm{ref}} + \widetilde{T} \\
\rho_{\mathrm{d}} &= \rho_{\mathrm{d,ref}} + \widetilde{\rho}_{\mathrm{d}} = \widetilde{\rho}_{\mathrm{d}}\,.
\end{aligned}$$

The reference state is defined as follows. It is dust-free ($\rho_{\mathrm{d,ref}} = 0$) and has constant gas temperature T_{ref}. The gas in the reference state does not shear, either in the radial or vertical directions, but rotates with a fixed angular frequency Ω_{F} in the inertial frame (hence $\mathbf{v}_{\mathrm{ref}} = 0$ in the rotating frame). In the reference state there exists a radial pressure gradient directed outward

$$-\frac{1}{\rho_{\mathrm{g,ref}}}\frac{\partial P_{\mathrm{ref}}}{\partial r} = 2\Omega_{\mathrm{K}}^2 \eta R = 2\Omega_{\mathrm{K}} v_{\max} \tag{2.15}$$

and a vertical pressure gradient balanced by vertical tidal gravity

$$-\frac{1}{\rho_{\mathrm{g,ref}}}\frac{\partial P_{\mathrm{ref}}}{\partial z} = \Omega_{\mathrm{K}}^2 z\,. \tag{2.16}$$

Equation (2.16) together with equation (2.12) and the assumption of constant T_{ref} implies that the reference gas density $\rho_{\mathrm{g,ref}}$ and pressure P_{ref} have Gaussian vertical distributions in z with scale height $H_{\mathrm{g}} = \sqrt{\Re T_{\mathrm{ref}}}/\Omega_{\mathrm{K}}$. For simplicity we neglect the radial density gradient ($\partial \rho_{\mathrm{g,ref}}/\partial r = 0$), as did B09. This reference state should not be confused with our equilibrium states of interest (§2.2.2), which do shear and which do contain dust. The reference state merely serves as a fiducial.

The flows of interest are subsonic. Mach numbers $\epsilon \equiv \widetilde{v}/c_{\mathrm{s}}$ peak at $v_{\max}/c_{\mathrm{s}} \sim c_{\mathrm{s}}/(\Omega_{\mathrm{K}} R) \sim 0.02$ for gas sound speeds $c_{\mathrm{s}} \sim 1$ km/s at $R \sim 1$ AU. Such flow is nearly incompressible: $|\widetilde{\rho}_{\mathrm{g}}|/\rho_{\mathrm{g,ref}} \sim |\widetilde{P}|/P_{\mathrm{ref}} \sim |\widetilde{T}|/T_{\mathrm{ref}} \sim \epsilon^2$. Invoking the anelastic approximation, we keep only terms leading in ϵ in any given equation. Equations (2.9), (2.10), and (2.12) reduce to:

$$\frac{d\rho_{\mathrm{g}}}{dt} + \rho_{\mathrm{g}} \nabla \cdot \mathbf{v} = \frac{\partial \rho_{\mathrm{g}}}{\partial t} + \nabla \cdot (\rho_{\mathrm{g}} \mathbf{v}) \approx \nabla \cdot (\rho_{\mathrm{g,ref}} \mathbf{v}) = 0 \tag{2.17}$$

$$\frac{d(\rho_{\mathrm{d}}/\rho_{\mathrm{g}})}{dt} \approx \frac{d(\widetilde{\rho}_{\mathrm{d}}/\rho_{\mathrm{g,ref}})}{dt} \equiv \frac{d\mu}{dt} = 0 \tag{2.18}$$

$$\frac{\widetilde{P}}{\rho_{\mathrm{g,ref}}} \equiv \widetilde{h} = \frac{\widetilde{\rho}_{\mathrm{g}}}{\rho_{\mathrm{g,ref}}} \Re T_{\mathrm{ref}} + \Re \widetilde{T} \tag{2.19}$$

where we define $\mu \equiv \widetilde{\rho}_{\rm d}/\rho_{\rm g,ref} = \rho_{\rm d}/\rho_{\rm g,ref}$ and the pressure-like enthalpy $\widetilde{h} \equiv \widetilde{P}/\rho_{\rm g,ref}$, and henceforth for convenience drop all tildes on $\rho_{\rm d}$, μ, and $\mathbf{v}$ (but not the other variables related to gas). The rightmost equalities of (2.17), (2.18), and (2.19) match equations (12b), (12c), and (12e) of B09. The anelastic approximation has been employed in the study of atmospheric convection (Ogura & Phillips 1962; Gough 1969), stars (Gilman & Glatzmaier 1981), and vortices in protoplanetary disks (Barranco & Marcus 2000, 2005, 2006). By eliminating the time derivative in the continuity equation (2.17), we effectively "sound-proof" the fluid. The simulation timestep is not limited by the sound-crossing time but rather by the longer advection time.

We rewrite our energy equation (2.11) as follows: replace $-\nabla \cdot \mathbf{v}$ with $d\ln\rho_{\rm g}/dt = -d\ln T/dt + d\ln P/dt$ to find that

$$\begin{aligned} C_{\rm P}\frac{d\widetilde{T}}{dt} &= \frac{1}{\rho_{\rm g}}\frac{dP}{dt} \\ &\approx \frac{1}{\rho_{\rm g,ref}}\mathbf{v}\cdot\nabla P_{\rm ref} \\ &\approx -\mathbf{v}\cdot\left(2\Omega_{\rm K}^2\eta R\hat{\mathbf{x}} + \Omega_{\rm K}^2 z\hat{\mathbf{z}}\right) \end{aligned} \tag{2.20}$$

where for the second line we dropped $d\widetilde{P}/dt$ in comparison to $\mathbf{v}\cdot\nabla P_{\rm ref}$, and for the third line we replaced $\rho_{\rm g,ref}^{-1}\nabla P_{\rm ref}$ using (2.15) and (2.16). Equation (2.20) matches (12d) of B09 except that for the right-hand side he has a coefficient equal to $1 + \widetilde{T}/T_{\rm ref}$, which we have set to unity.

Finally, to recover the form of the momentum equation (12a) of B09, first consider the pressure acceleration and isolate the contribution from dust-free gas ($-\rho_{\rm g}^{-1}\nabla P$):

$$\begin{aligned} -\frac{1}{\rho_{\rm d}+\rho_{\rm g}}\nabla P &= -\left(\frac{1}{\rho_{\rm d}+\rho_{\rm g}} - \frac{1}{\rho_{\rm g}}\right)\nabla P - \frac{1}{\rho_{\rm g}}\nabla P \\ &\approx \frac{\mu}{\mu+1}\left(\frac{1}{\rho_{\rm g}}\nabla P\right) - \frac{1}{\rho_{\rm g}}\nabla P\,. \end{aligned} \tag{2.21}$$

Now expand

$$\begin{aligned} \frac{1}{\rho_g}\nabla P &\approx \frac{1}{\rho_{\rm g,ref}}\nabla P_{\rm ref} + \frac{1}{\rho_{\rm g,ref}}\nabla\widetilde{P} - \frac{\widetilde{\rho}_{\rm g}}{\rho_{\rm g,ref}^2}\nabla P_{\rm ref} \\ &\approx \frac{1}{\rho_{\rm g,ref}}\nabla P_{\rm ref} + \nabla\widetilde{h} + \frac{\widetilde{T}}{T_{\rm ref}}\left(\frac{\nabla P_{\rm ref}}{\rho_{\rm g,ref}}\right) \\ &\approx -\left(1+\frac{\widetilde{T}}{T_{\rm ref}}\right)(2\Omega_{\rm K}^2\eta R\hat{\mathbf{x}} + \Omega_{\rm K}^2 z\hat{\mathbf{z}}) + \nabla\widetilde{h} \end{aligned} \tag{2.22}$$

where for the last line we used (2.15) and (2.16). Insertion of (2.21) and (2.22) into (2.14) yields the anelastic momentum equation (12a) of B09:

$$\begin{aligned}\frac{d\mathbf{v}}{dt} &= -2\Omega_{\mathrm{K}}\hat{\mathbf{z}}\times\mathbf{v} + 3\Omega_{\mathrm{K}}^2 x\hat{\mathbf{x}} + \frac{\widetilde{T}}{T_{\mathrm{ref}}}(2\Omega_{\mathrm{K}}^2\eta R\hat{\mathbf{x}} + \Omega_{\mathrm{K}}^2 z\hat{\mathbf{z}}) - \nabla\widetilde{h} \\ &\quad -\frac{\mu}{\mu+1}\left[\left(1+\frac{\widetilde{T}}{T_{\mathrm{ref}}}\right)(2\Omega_{\mathrm{K}}^2\eta R\hat{\mathbf{x}} + \Omega_{\mathrm{K}}^2 z\hat{\mathbf{z}}) - \nabla\widetilde{h}\right]. \end{aligned} \tag{2.23}$$

which isolates the driving term due to dust.

2.2.2 Initial Conditions

Equilibrium initial conditions (superscripted "†") are specified by five functions: $\mu = \mu^\dagger$, $\widetilde{T} = \widetilde{T}^\dagger$, $\widetilde{h} = \widetilde{h}^\dagger$, $\widetilde{\rho_{\mathrm{g}}} = \widetilde{\rho_{\mathrm{g}}}^\dagger$, and $\mathbf{v} = \mathbf{v}^\dagger$. For $\mu^\dagger$, we use flows characterized by a globally constant Richardson number (Sekiya 1998; Youdin & Shu 2002; Chiang 2008). The conditions $Ri =$ constant, $\partial\rho_{\mathrm{g}}/\partial z \ll \partial\rho_{\mathrm{d}}/\partial z$, and $g = -\Omega_{\mathrm{K}}^2 z$ (no self-gravity) yield

$$\mu^\dagger(z) = \left[\frac{1}{1/(1+\mu_0)^2 + (z/z_{\mathrm{d}})^2}\right]^{1/2} - 1, \tag{2.24}$$

where μ_0 is the initial midplane dust-to-gas ratio and

$$z_{\mathrm{d}} \equiv \frac{Ri^{1/2}\, v_{\mathrm{max}}}{\Omega_{\mathrm{K}}} \tag{2.25}$$

is a characteristic dust height. The dust density peaks at the midplane and decreases to zero at

$$z = \pm z_{\mathrm{max}} = \pm\frac{\sqrt{\mu_0(2+\mu_0)}}{1+\mu_0} z_{\mathrm{d}} \tag{2.26}$$

which is consistent with our order-of-magnitude expression (2.6). Neither equation (2.24) nor the code accounts for self-gravity and therefore we are restricted to modeling flows whose densities are less than that required for the Toomre parameter of the subdisk to equal unity (CY10; see also §2.4). For the minimum-mass disk of CY10, this restriction is equivalent to $\mu \lesssim 30$. Input model parameters include μ_0, Ri, and v_{max}.

For the gas, we assume

$$\widetilde{T}^\dagger = 0 \tag{2.27}$$

(initially isothermal) and solve vertical hydrostatic equilibrium for $\widetilde{h}^\dagger$ (the z-component of equation 2.23):

$$\frac{\partial\widetilde{h}^\dagger}{\partial z} = -\mu^\dagger\Omega_{\mathrm{K}}^2 z\,. \tag{2.28}$$

The functional form for $\widetilde{h}^\dagger(z)$ is not especially revealing and so we do not write it out here. For simplicity we assume that $\widetilde{h}^\dagger$ does not depend on x. From $\widetilde{h}^\dagger$ and $\widetilde{T}^\dagger = 0$ it follows from

(2.19) that

$$\widetilde{\rho_{\rm g}}^{\dagger} = \frac{\rho_{\rm g,ref}\widetilde{h}^{\dagger}}{\Re T_{\rm ref}} \,. \tag{2.29}$$

The fractional deviations $\widetilde{\rho_{\rm g}}^{\dagger}/\rho_{\rm g,ref}$ and $\widetilde{P}^{\dagger}/P_{\rm ref}$ from the reference state are very small, of order $\mu^{\dagger}(v_{\rm max}/c_{\rm s})^2 Ri$.

It remains to specify $\mathbf{v}^{\dagger}$. Using the conditions on $\widetilde{h}^{\dagger}$ stated above, we solve for the equilibrium (steady-state) solution to equation (2.23):

$$\begin{aligned} v_x^{\dagger} &= v_z^{\dagger} = 0 \\ v_y^{\dagger} &= -\frac{3}{2}\Omega_{\rm K} x + \left[\frac{\mu^{\dagger}(z)}{\mu^{\dagger}(z)+1}\right] v_{\rm max} \,. \end{aligned} \tag{2.30}$$

In our reference frame rotating with the velocity of dust-free gas at R, the first term on the right side of (2.30) accounts for the standard Kepler shear, while the second term describes how dust, which adds to inertia but not pressure, speeds up the gas.

To $\mu^{\dagger}$ we add random perturbations

$$\Delta\mu(x,y,z) = A(x,y)\mu^{\dagger}(z)[\cos(\pi z/2z_d) + \sin(\pi z/2z_d)] \,. \tag{2.31}$$

The amplitude $A(x,y)$ is constructed in Fourier space so that each Fourier mode has a random phase and an amplitude inversely proportional to the horizontal wavenumber: $\hat{A} \propto k_{\perp}^{-1} = (k_x^2 + k_y^2)^{-1/2}$. Because our box sizes are scaled to $z_{\rm max}$, our Fourier noise amplitudes are largest on scales comparable to the dust layer thickness. Thus those modes which are most likely to overturn the layer are given the greatest initial power. The perturbations are also chosen to be antisymmetric about the x-axis so that no extra energy is injected into the system. We take the root-mean-squared amplitude $A_{\rm rms} \equiv \langle A^2 \rangle^{1/2}$ of the perturbations to be 10^{-4} or 10^{-3}.

In summary, three input parameters μ_0, Ri, and $v_{\rm max}$ determine our isothermal equilibrium initial conditions (equations 2.24, 2.28, and 2.30).[6] The equilibrium solution for $\mu(z)$ is then perturbed (equation 2.31) by a root-mean-squared fractional amount $A_{\rm rms}$. The parameters of primary interest are μ_0 and Ri. For the remaining parameters $v_{\rm max}$ and $A_{\rm rms}$ we consider three possible combinations: $(v_{\rm max}, A_{\rm rms}) = (0.025c_{\rm s}, 10^{-4})$ for our standard runs; $(0.025c_{\rm s}, 10^{-3})$ to probe larger initial perturbations; and $(0.05c_{\rm s}, 10^{-4})$ to assess the effect of a stronger radial pressure gradient.

Note that specifying μ_0 and Ri (and $v_{\rm max}$, though this last variable is fixed for all of our standard runs) specifies the entire dust and gas vertical profiles, $\rho_{\rm d}(z)$ and $\rho_{\rm g}(z)$, and by extension the bulk height-integrated metallicity, $\Sigma_{\rm d}/\Sigma_{\rm g} \equiv \int \rho_{\rm d} dz / \int \rho_{\rm g} dz$. We do not give an explicit expression for $\Sigma_{\rm d}/\Sigma_{\rm g}$ because it is cumbersome and not particularly revealing. The bulk metallicity is in some sense the most natural independent variable because its value is

---

[6] While our initial conditions are isothermal, the temperature of the flow can change because of adiabatic compression/expansion and because our artificial hyperviscosity dissipates the highest wavenumber disturbances. These temperature changes are fractionally tiny because the flow is highly subsonic.

given by the background disk (for ways in which the bulk metallicity may change, e.g., by radial particle drifts, see CY10). We will plot our results in the space of μ_0, Ri, and $\Sigma_{\rm d}/\Sigma_{\rm g}$, keeping in mind that only two of these three variables are independent.

2.2.3 Code

We use the spectral, anelastic, shearing box code developed by Barranco & Marcus (2006) and modified by B09 to simulate well-coupled gas and dust. The code employs shearing periodic boundary conditions in r, periodic boundary conditions in ϕ, and closed lid boundaries in z; the vertical velocity v_z is required to vanish at the top and bottom of the box ($z = \pm L_z/2$).

Spectral methods approximate the solution to the fluid equations as a linear combination of basis functions. The basis functions describe how the flow varies in space, and the coefficients of the functions are determined at every timestep. For each of the periodic dimensions, a standard Fourier basis is used, while for the vertical direction, Chebyshev polynomials are employed. Whereas in r and ϕ grid points are evenly spaced, the use of Chebyshev polynomials in z has the effect that vertical grid points are unevenly spaced; points are concentrated towards the top and bottom boundaries of the box, away from the midplane where the dust layer resides. Thus to resolve the dust layer vertically, we need to increase the number of vertical grid points N_z by an amount disproportionately large compared to the numbers of radial and azimuthal grid points N_r and N_ϕ. See §2.2.4 for further discussion.

Spectral codes have no inherent grid dissipation; energy is allowed to cascade down to the smallest resolved length scales through nonlinear interactions. To avoid an energy "pile-up" at the highest wavenumbers, we dissipate energy using an artificial hyperviscosity, given in §3.3.3 of Barranco & Marcus (2006).

Simulations satisfy the Courant-Friedrichs-Lewy (CFL) condition which states that the CFL number, defined as the code timestep divided by the shortest advection time across a grid cell, be small. In the shearing coordinates in which the code works, that advection time is the cell dimension divided by the local velocity over and above the Keplerian shear, i.e., orbital velocities are subtracted off before evaluating the CFL number. All simulations reported in this chapter are characterized by CFL numbers less than about 0.1.

2.2.4 Box Size and Numerical Resolution

Our standard box dimensions are $(L_r, L_\phi, L_z) = (6.4, 12.8, 8)z_{\rm max}$ and the corresponding numbers of grid points are $(N_r, N_\phi, N_z) = (32, 64, 128)$. By scaling our box lengths L_i to $z_{\rm max}$ and fixing the numbers of grid points N_i, we ensure that each standard simulation enjoys the same resolution (measured in grid points per physical length) regardless of Ri, μ_0, and $v_{\rm max}$. The vertical extent of the dust layer between $z = \pm z_{\rm max}$ is resolved by 22 grid points (this is less than $[128/(8z_{\rm max})] \times 2z_{\rm max} = 32$ because the Chebyshev-based vertical grid only sparsely samples the midplane). The radial and azimuthal directions are resolved by 10 grid points per $2z_{max}$ length. We choose our resolution in the vertical direction to be greater than that of the horizontal directions because the dust layer has finer scale structure in z: the dust layer becomes increasingly cuspy at the midplane as μ_0 increases. We prescribe the same resolution in the radial and azimuthal directions ($L_\phi/N_\phi = L_r/N_r$); experiments with different resolutions in r and ϕ generated spurious results.

Too small a box size can artificially affect the stability of the dust layer, because a given box can only support modes having integer numbers of wavelengths inside it. Small boxes may be missing modes that in reality overturn the layer. We verify that for all runs in which the dust layer overturns, the KH mode that most visibly disrupts the layer spans more than one azimuthal wavelength. Typically 3–5 wavelengths are discerned across the box.

To more thoroughly test our standard choices for L_i, we study how systematic variations in box length affect how the instability develops. For this test, we adopt a fixed set of physical input parameters, $(Ri, \mu_0, v_{\rm max}) = (0.1, 10, 0.025c_s)$, which should lead to instability (Chiang 2008). Our diagnostic is the time evolution of the vertical kinetic energy at the midplane: $\langle \mu(t) v_z^2(t) \rangle/2$, where the average is over all r and ϕ at fixed $z = 0$ and time t. We vary L_i and N_i in tandem to maintain the same resolution from run to run, thereby isolating the effect of box size. Figure 2.1 shows how doubling one of the box dimensions while fixing the other two alters the time history of $\langle \mu v_z^2 \rangle/2$. Panel (a) demonstrates that our standard choice for $L_z = 8z_{\rm max}$ is sufficiently large because the curves for $L_z = 8z_{\rm max}$ and $L_z = 16z_{\rm max}$ practically overlap. Panels (b) and (c) show that our standard choices for $L_\phi = 12.8z_{\rm max}$ and $L_r = 16z_{\rm max}$ are somewhat less adequate. The peak of the curve for $(L_\phi, N_\phi) = (12.8z_{max}, 64)$ is delayed by two orbits compared to that for $(L_\phi, N_\phi) = (25.6z_{\rm max}, 128)$, and the curve for $(L_r, N_r) = (6.4z_{max}, 32)$ peaks an orbit earlier than that for $(L_r, N_r) = (12.8z_{\rm max}, 64)$. Nevertheless these time differences are small compared to the total time to instability, about 10 orbits. Moreover, the errors point in opposite directions. Thus we expect our choices for L_ϕ and L_r to partially compensate for each other so that any error due to our box size in calculating the time to instability will be less than $\sim$1 orbit.

We test how robust our results are to numerical resolution by re-running a few simulations at twice the normal resolution (doubling N_i while fixing L_i). Results at high resolution are given in §2.3.3. Every simulation is run for at least ten orbits. A typical run performed at our standard resolution takes approximately 2.5 wall-clock hours using 56 processors on the Purdue Steele cluster. A high-resolution run takes about 32 wall-clock hours.

2.3 Results

In our standard simulations, we fix $v_{\rm max}$ and $A_{\rm rms}$ while systematically varying Ri and μ_0 from run to run. Our systematic variations of Ri and μ_0 correspond to systematic variations in $\Sigma_{\rm d}/\Sigma_{\rm g}$; recall that only two of the three parameters Ri, μ_0, and $\Sigma_{\rm d}/\Sigma_{\rm g}$ are independent. For each $\mu_0 \in \{0.3, 1, 3, 10\}$ we adjust Ri until the threshold value $Ri_{\rm crit}$ dividing stable from unstable runs is determined to within 0.1 dex.

Deciding by numerical simulation whether a given dust layer is stable or not is unavoidably subject to the finite duration of the simulation. We define our criteria for deciding

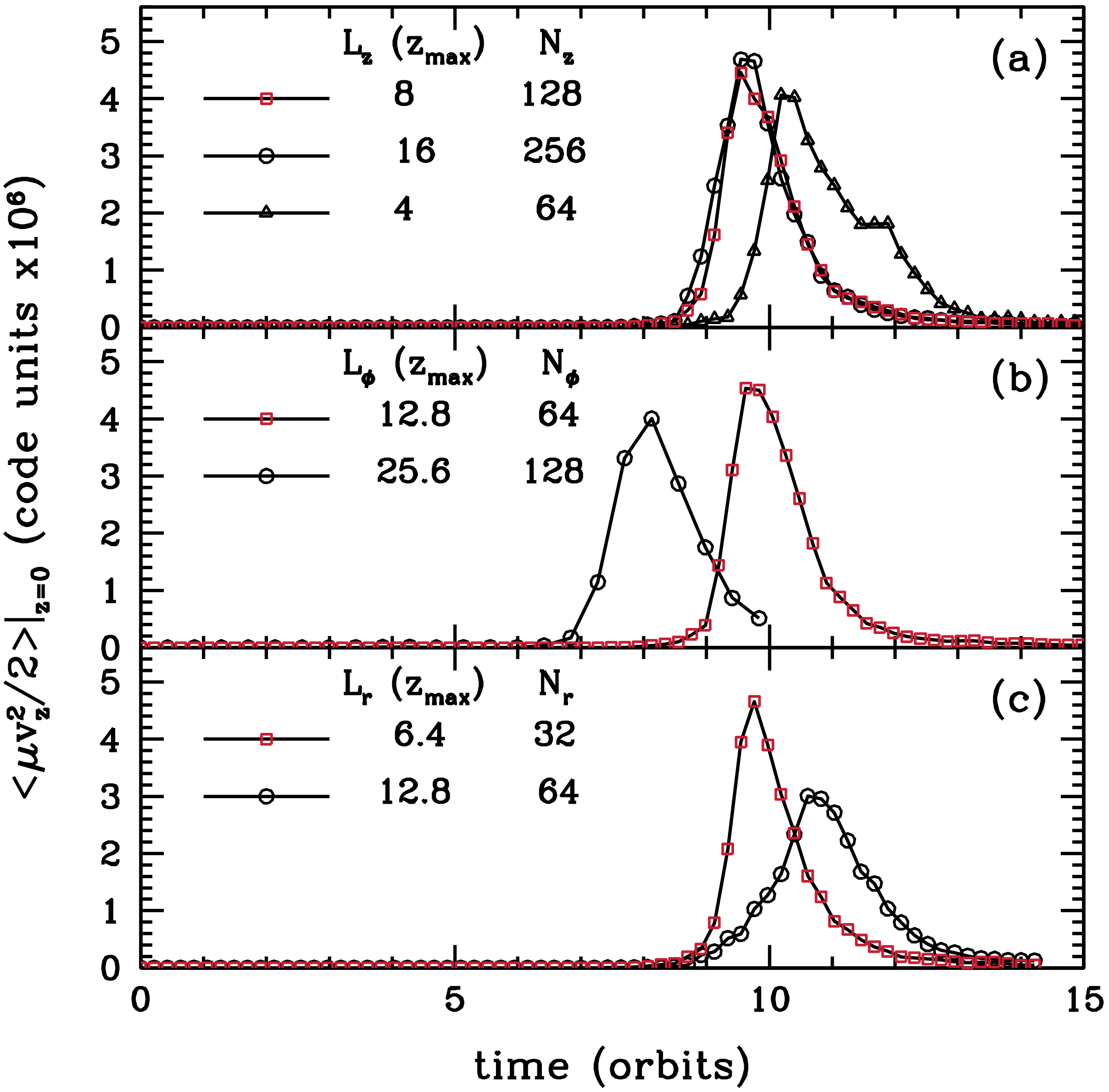

Figure 2.1: Testing box sizes at fixed numerical resolution. For our standard box, $(L_r, L_\phi, L_z) = (6.4, 12.8, 8)z_{\max}$ and $(N_r, N_\phi, N_z) = (32, 64, 128)$. In each panel we vary one box dimension while keeping the other two dimensions fixed at their standard values. In the top panel we vary L_z at fixed resolution N_z/L_z. In the middle and bottom panels, L_ϕ and L_r are varied in turn. All simulations in this figure have $\mu_0 = 10$, $Ri = 0.1$, $v_{\max} = 0.025c_s$, $A_{\rm rms} = 10^{-4}$, and use code units $\rho_{\rm g,ref}(z=0) = \Omega_{\rm K} = H_{\rm g} = 1$. Doubling the box dimensions from our standard values changes when the average vertical kinetic energy peaks by only a few orbits at most. The average $\langle\rangle$ is performed over all r and ϕ at fixed $z = 0$.

stability in §2.3.1. Results are given in §2.3.2 and tested for robustness in §2.3.3.

2.3.1 Criteria for Stability

Stability is assessed by two quantities: the midplane vertical kinetic energy

$$\langle \mu v_z^2 \rangle / 2 \text{ as a function of } t$$

where the average is performed over r and ϕ at fixed $z = 0$ and t, and the dust density profile

$$\langle \mu \rangle \text{ as a function of } z \text{ and } t$$

where the average is performed over r and ϕ at fixed z and t. By definition, in an "unstable" run, $\langle \mu v_z^2 \rangle / 2$ grows exponentially over several orbital periods, and $\langle \mu \rangle$ deviates from its initial value $\langle \mu^\dagger \rangle$ by more than 15%. "Stable" simulations satisfy neither criterion. Some runs are "marginally unstable" in that they satisfy the first but not the second criterion. At the end of the standard ten-orbit duration of a marginally unstable run, we find the kinetic energy continues to rise, suggesting that were the run to be extended for longer than ten orbits, the dust layer would eventually overturn. In every instance where we extend the duration of a marginally unstable run, we verify that this is the case. Thus "marginally unstable" is practically synonymous with "unstable."

Examples of unstable and stable runs are shown in Figure 2.2. In the unstable simulation, after $t \approx 6$ orbits, the kinetic energy rises exponentially. At $t \approx 9$ orbits, the dust layer overturns and the midplane dust-to-gas ratio falls by more than 60%. By contrast, in the stable simulation, after an initial adjustment period lasting $\sim$3 orbits during which the midplane value of $\langle \mu \rangle$ decreases by 10%, the kinetic energy drops by orders of magnitude to a nearly constant value and shows no evidence of further growth.

Figure 2.3 shows the evolution of $|v_i(z)|$ $(i = r, \phi, z)$ and $\langle \mu(z) \rangle$ for the same unstable run of Figure 2.2. The velocity data are sampled at a single (x, y) position at the center of our simulation box. The radial and vertical velocities $|v_r|$ and $|v_z|$, initially zero, grow to become comparable with the shearing velocity $|v_\phi|$. Figure 2.4 displays corresponding snapshots of $\mu(y, z)$, taken at a single radius x near the center of our box. Though the data in Figures 2.3 and 2.4 are sampled at particular radial locations in our box, we verify that the instability develops similarly at all locations—as it should—unlike the ZEUS-based simulations of Chiang (2008).

2.3.2 Stability as a Function of Ri, μ_0, and $\Sigma_{\rm d}/\Sigma_{\rm g}$

Figure 2.5 maps the stable and unstable regions in (Ri, μ_0) space, for fixed $v_{\rm max} = 0.025 c_{\rm s}$ and $A_{\rm rms} = 10^{-4}$. Figures 2.6 and 2.7 portray the same data using alternate but equivalent projections of parameter space: $(Ri, \Sigma_{\rm d}/\Sigma_{\rm g})$ and $(\mu_0, \Sigma_{\rm d}/\Sigma_{\rm g})$, respectively.

These plots demonstrate that there is no unique value of $Ri_{\rm crit}$. Rather $Ri_{\rm crit}$ is a function of μ_0, or equivalently a function of $\Sigma_{\rm d}/\Sigma_{\rm g}$. For bulk metallicities $\Sigma_{\rm d}/\Sigma_{\rm g}$ near the solar value,

$Ri_{\rm crit}$ is found to be close to the classical value of 1/4. But as $\Sigma_{\rm d}/\Sigma_{\rm g}$ decreases below the solar value, $Ri_{\rm crit}$ shrinks to ~0.01 or even lower. A least-squares fit to the four midpoints (evaluated in log space) in Figure 2.5 dividing neighboring stable points (in black) and unstable points (in red or red outlined with black) yields $Ri_{\rm crit} \propto \mu_0^{1.0}$. This same fit, projected into metallicity space, is shown in Figures 2.6 and 2.7; in metallicity space the stability boundary is not a power law.

As Figure 2.7 attests, dust-to-gas ratios μ_0 as high as ~8 can be attained in disks of solar metallicity without triggering a shear instability: see the intersection between the dashed curve fitted to our standard resolution data, and the dotted line representing solar metallicity. This intersection occurs at $\mu_0 \approx 7$. Were we to re-fit the dashed curve using the higher resolution data represented by triangles, the intersection with solar metallicity would occur at μ_0 closer to 8.

A dust-to-gas ratio of $\mu_0 \approx 8$ is within a factor of ~4 of the Toomre threshold for gravitational fragmentation in a minimum-mass disk (CY10; §2.4). We can achieve the Toomre threshold by simply allowing for a gas disk that is ~4× more massive than the minimum-mass nebula. Alternatively we can enrich the disk in metals to increase $\Sigma_{\rm d}/\Sigma_{\rm g}$. Extrapolating the boundary of stability (dashed curve) in Figure 2.7 to higher $\Sigma_{\rm d}/\Sigma_{\rm g}$ suggests that the Toomre threshold $\mu_0 \approx 30$ could be achieved for minimum-mass disks having ~3× the solar metallicity. The sensitivity to metallicity is also exemplified by Figure 2.2. For the same $\mu_0 = 10$, the dust layer based on a near-solar metallicity of $\Sigma_{\rm d}/\Sigma_{\rm g} = 0.013$ overturns, whereas one derived from a supersolar metallicity of $\Sigma_{\rm d}/\Sigma_{\rm g} = 0.030$ remains stable.

2.3.3 Tests at Higher Resolution, Higher $A_{\rm rms}$, and Higher $v_{\rm max}$

We test how robust our determination of $Ri_{\rm crit}$ is to numerical resolution by redoing our simulations for $\mu_0 = 0.3$ and 10 with double the number of grid points in each dimension. The results are overlaid as blue triangles in Figures 2.5, 2.6, and 2.7. At $\mu_0 = 0.3$, increasing the resolution does not change $Ri_{\rm crit}$ from its value of 0.009. At $\mu_0 = 10$, $Ri_{\rm crit}$ shifts downward from 0.3 to 0.2. Although we have not strictly demonstrated convergence of our results with resolution, and although high resolution data at other values of μ_0 are missing, it seems safe to conclude that the slope of the stability boundary in Ri-μ_0 space is close to, but decidedly shallower than, linear.

We also test the sensitivity of our results to $A_{\rm rms}$. Increasing $A_{\rm rms}$ by an order of magnitude to 10^{-3} shifts $Ri_{\rm crit}$ upward by $\lesssim 0.2$ dex at $\mu_0 < 1$, but leaves $Ri_{\rm crit}$ unchanged at larger μ_0 (Figure 2.8). B09 also reported some sensitivity to $A_{\rm rms}$.

Tests where $v_{\rm max}$ was doubled to $0.05c_{\rm s}$ reveal no change in $Ri_{\rm crit}$ (data not shown).

2.4 Summary and Discussion

Where a protoplanetary disk is devoid of turbulence intrinsic to gas, dust particles settle toward the midplane, accumulating in a sublayer so thin and so dense that the dust-gas

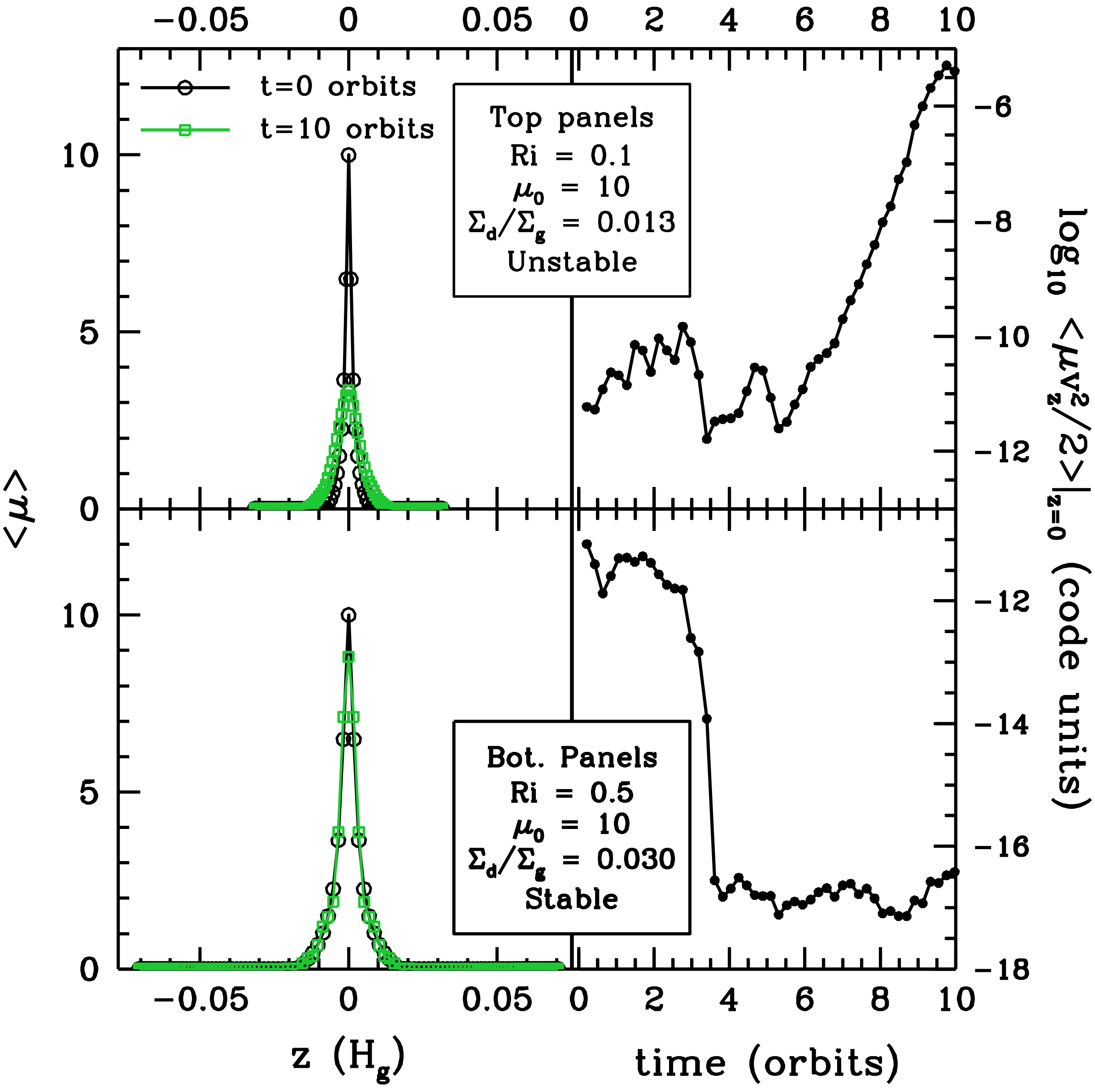

Figure 2.2: Sample unstable (top) and stable (bottom) dust layers. In the unstable case, the layer overturns and mixes dust-rich gas with dust-poor gas, causing the dust-to-gas ratio at the midplane to drop by a factor of ∼3 after 10 orbits (top left). As the instability unfolds, the vertical kinetic energy amplifies exponentially from $t \approx 5$–10 orbits (top right). At fixed μ_0, the layer is stabilized by increasing the Richardson number or equivalently the height-integrated metallicity $\Sigma_\mathrm{d}/\Sigma_\mathrm{g}$. In the stable run, the dust profile changes by less than 15% (bottom left) while the kinetic energy, after dropping precipitously, shows no indication of growing (bottom right). The two runs shown use $v_\mathrm{max} = 0.025 c_\mathrm{s}$ and $A_\mathrm{rms} = 10^{-4}$.

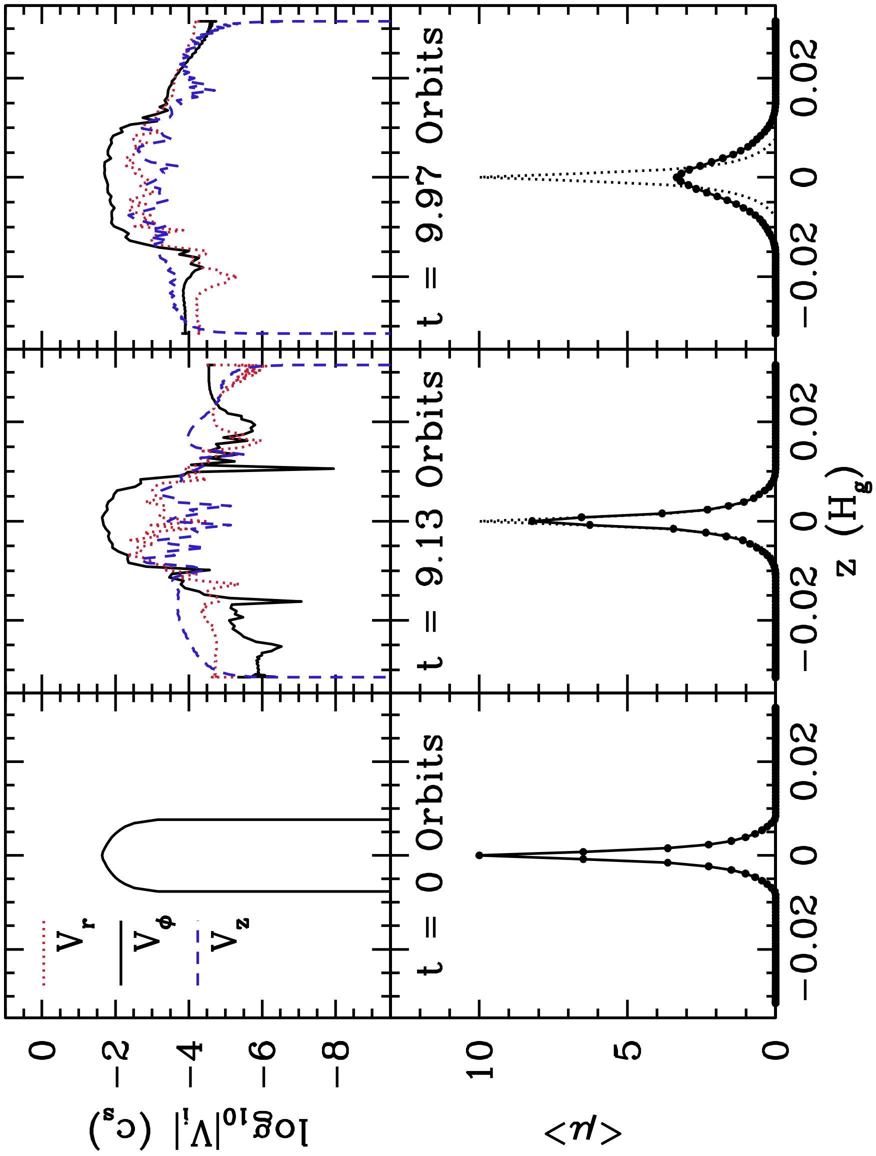

Figure 2.3: Snapshots of absolute values of the three velocity components (top panels) and horizontally averaged dust-to-gas ratio (bottom panels), both as functions of height, at three instants in time. For this unstable run, $(Ri, \mu_0, v_{\rm max}, A_{\rm rms}) = (0.1, 10, 0.025c_{\rm s}, 10^{-4})$. Velocities are taken from a grid point near the middle of the box. The vertical shear $\partial v_\phi/\partial z$ inside the dust layer weakens with time as dust is more uniformly mixed with gas, and as the radial and vertical velocities grow at the expense of the azimuthal velocity.

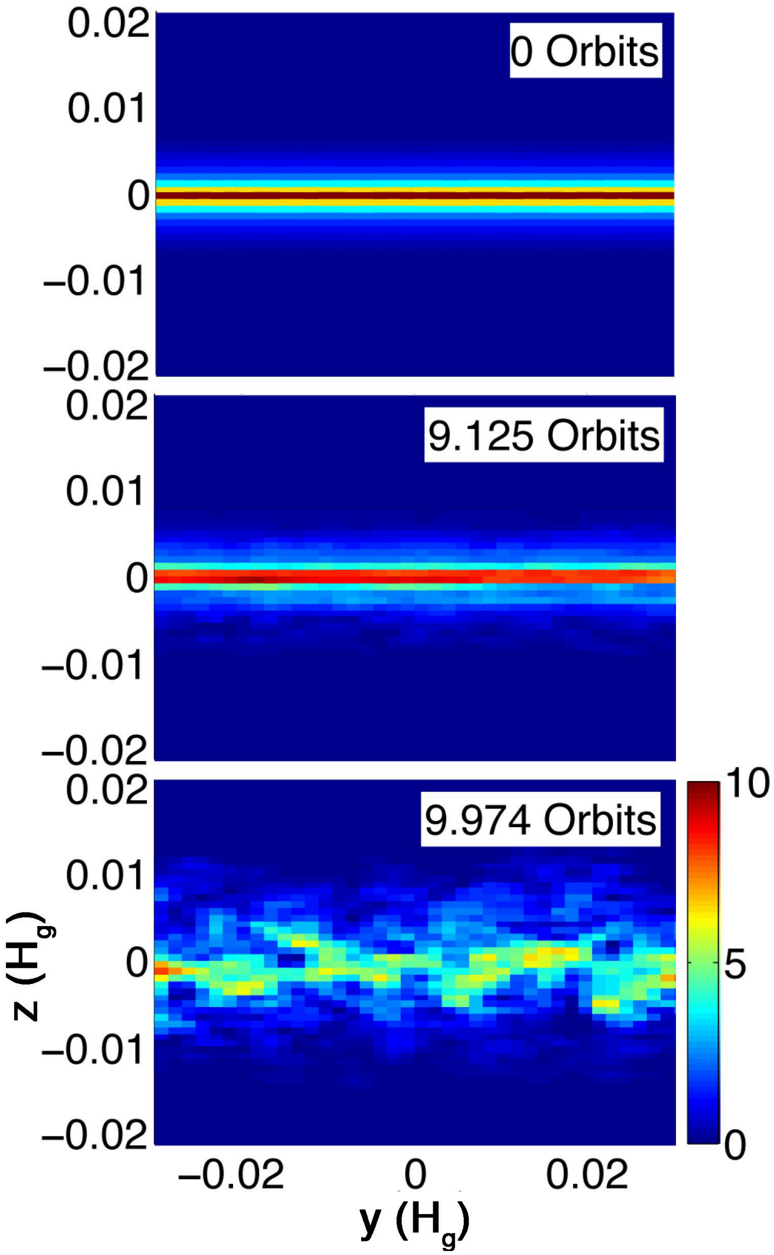

Figure 2.4: Snapshots of $\mu(y, z)$, sampled at $r = R$ ($x = 0$; the central slice of the simulation box) for the same unstable run shown in Figure 2.3. The box size parameters are $(L_r, L_\phi, L_z) = (0.05, 0.1, 0.063)H_\mathrm{g}$, larger than what is shown in the figure, which zooms in for more detail.

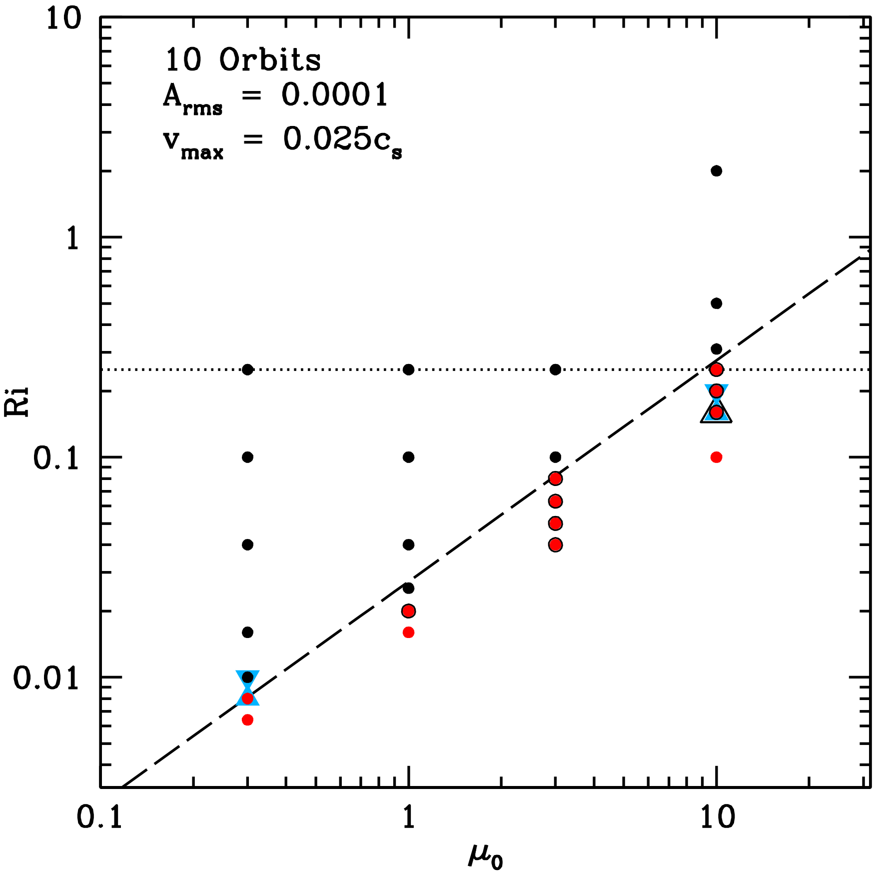

Figure 2.5: Mapping the boundary of stability in the space of initial Ri and μ_0. Red points correspond to unstable dust layers, whose dust-to-gas ratios $\langle\mu\rangle$ change by more than 15%, and whose vertical kinetic energies grow exponentially, within the 10-orbit duration of the simulation. Black points mark stable dust layers satisfying neither criterion. Red points outlined in black signify marginally unstable layers, whose kinetic energies rise but whose dust-to-gas ratios change by less than 15%; these are essentially equivalent to red points without outlines, because every marginally unstable run that we extend beyond 10 orbits eventually becomes fully unstable. Runs performed at twice the standard resolution appear as triangles. Downward pointing triangles symbolize stable runs, upward triangles are unstable, and upward pointing triangles in black outline are marginally unstable. All simulations use $A_{\mathrm{rms}} = 10^{-4}$ and $v_{\mathrm{max}} = 0.025c_{\mathrm{s}}$. There is no unique value for the critical Richardson number separating stable from unstable dust layers. Rather, a least-squares fit to the data from our standard resolution runs yields $Ri_{\mathrm{crit}} \propto \mu^{1.0}$, shown as a dashed line. The classical boundary $Ri_{\mathrm{crit}} = 0.25$ is plotted as a dotted line.

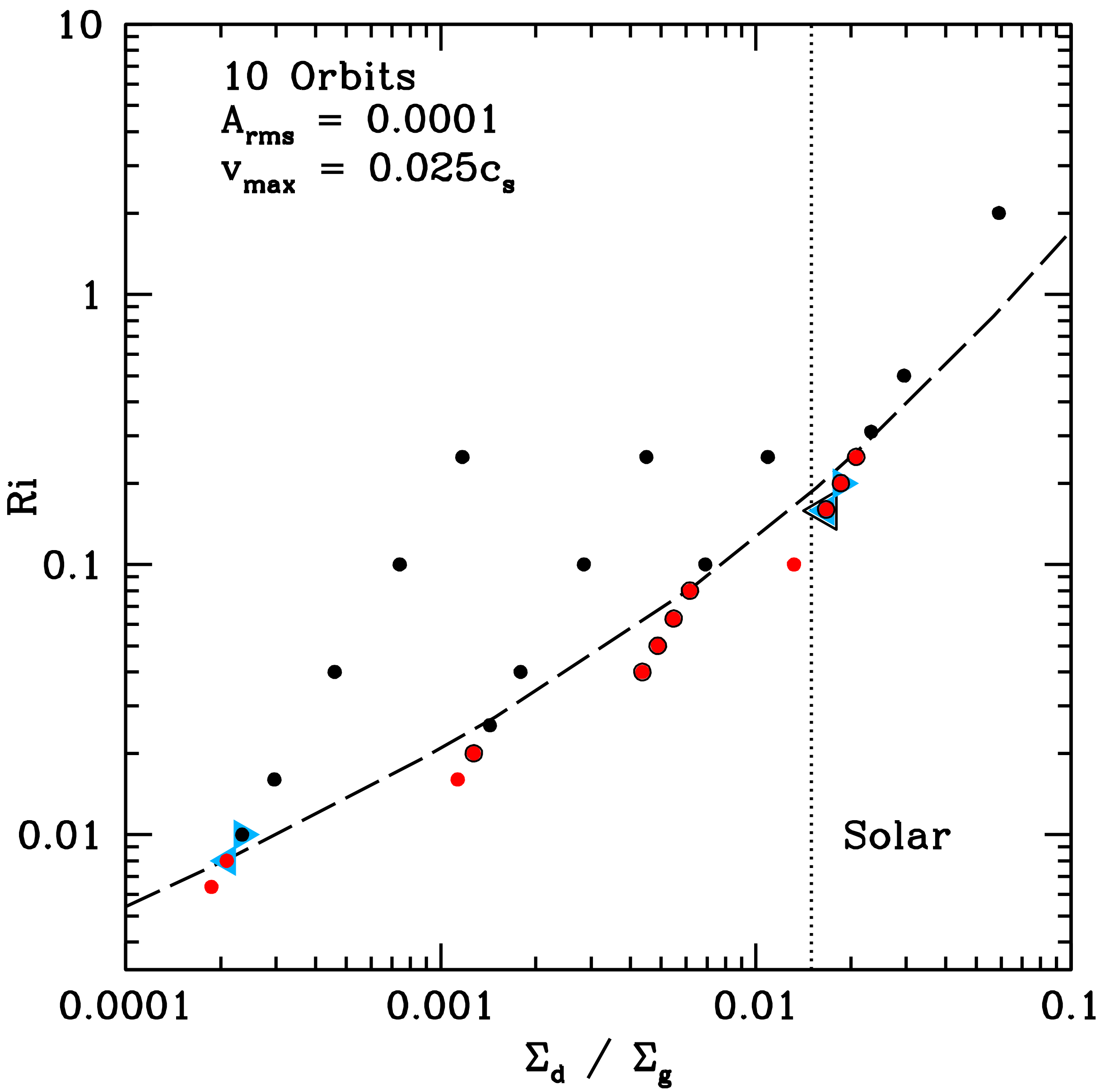

Figure 2.6: Mapping the boundary of stability in the space of initial *Ri* and bulk (height-integrated) dust-to-gas ratio $\Sigma_{\rm d}/\Sigma_{\rm g}$. The data are identical to those in Figure 2.5. The labeling convention is also the same, except that the triangles representing high-resolution runs have adjusted their orientation so that they point towards the stability boundary. The same least-squares fit from Figure 2.5 is projected here as a dashed curve. Solar metallicity $\Sigma_{\rm d}/\Sigma_{\rm g} = 0.015$ (Lodders 2003) is indicated by a dotted line. The critical value $Ri_{\rm crit}$ dividing stable from unstable dusty subdisks trends with metallicity. This trend was only hinted at in the data of C08.

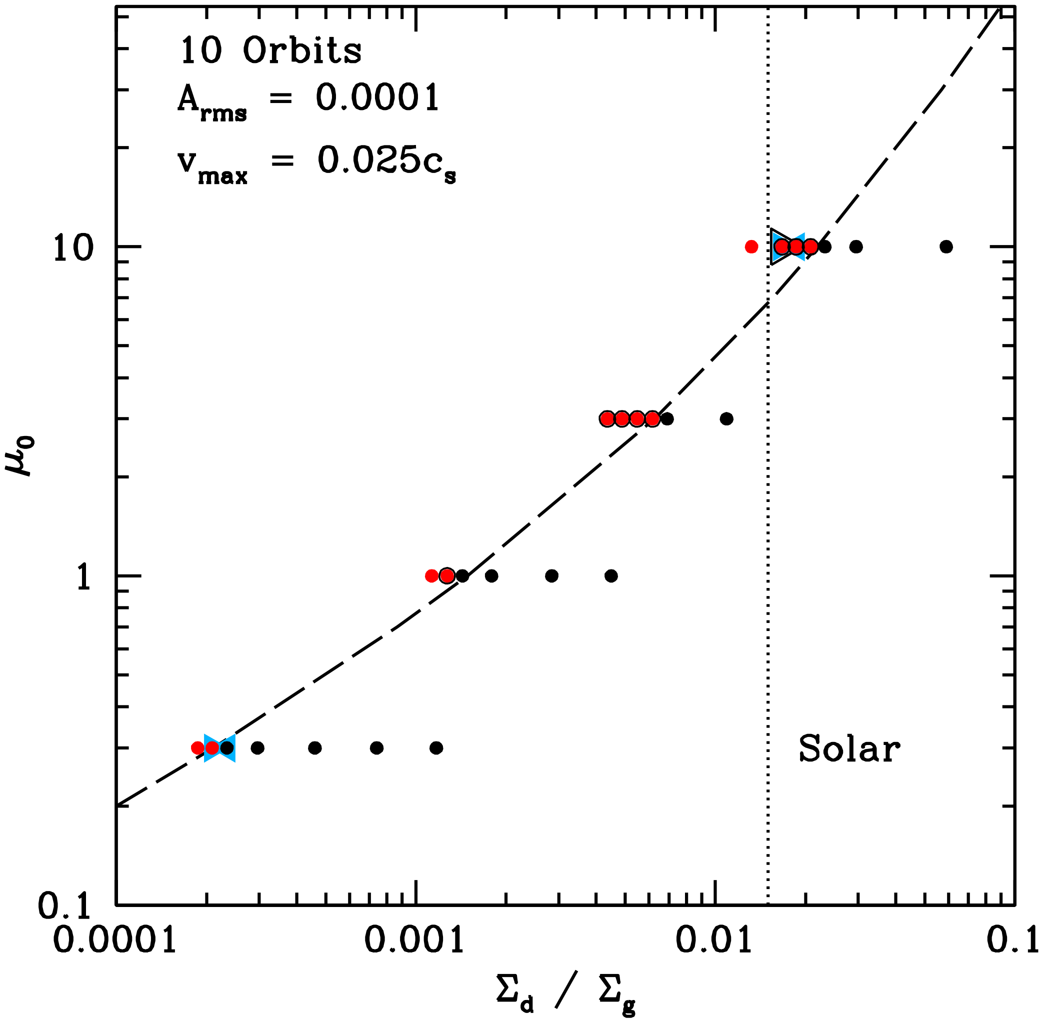

Figure 2.7: Mapping the boundary of stability in the space of midplane dust-to-gas ratio μ_0 and bulk (height-integrated) dust-to-gas ratio Σ_d/Σ_g. The data are identical to those in Figure 2.5. The labeling convention is also the same, except that the triangles representing high-resolution runs have adjusted their orientation so that they point towards the stability boundary. The same least-squares fit from Figure 2.5 is projected here as a dashed curve. Solar metallicity $\Sigma_d/\Sigma_g = 0.015$ (Lodders 2003) is indicated by a dotted line. A minimum-mass solar nebula requires $\mu_0 \approx 30$ for gravitational instability to ensue on a dynamical time (CY10). Extrapolating the boundary of stability to $\mu_0 \approx 30$ suggests that metallicities roughly ∼3 times solar would be required for dynamical gravitational instability in a minimum-mass disk. The required degree of metal enrichment would be proportionately less in more massive disks.

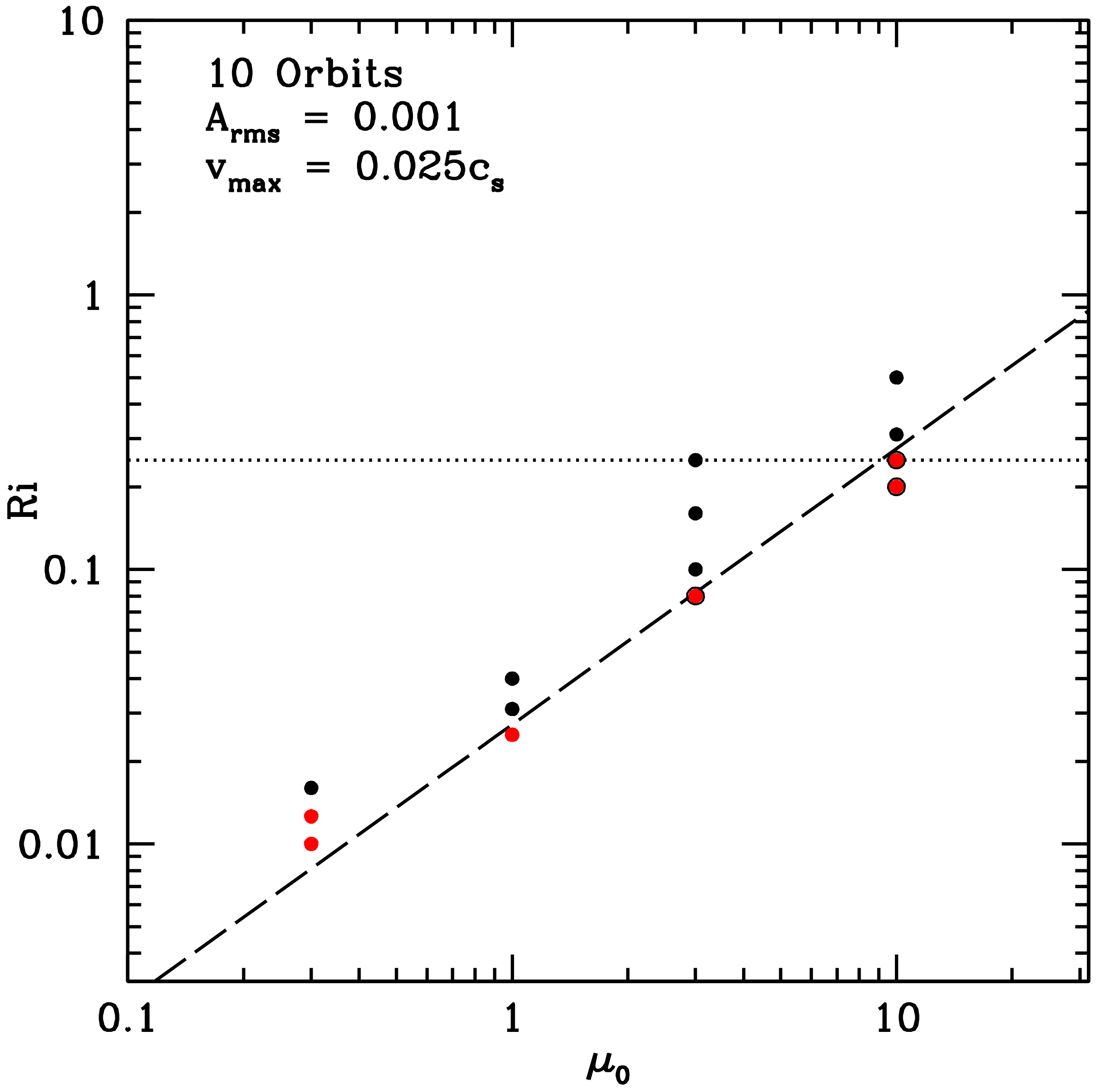

Figure 2.8: How the stability boundary changes with stronger initial perturbations. This figure is the same as Figure 2.5, except that all data correspond to $A_{\rm rms} = 10^{-3}$. For comparison with $A_{\rm rms} = 10^{-4}$, the same best-fit line of Figure 2.5 is reproduced here. Not much changes, except that $Ri_{\rm crit}$ shifts upward by 0.2 dex at $\mu_0 = 0.3$.

mixture becomes unstable. If the first instability to manifest is self-gravitational, dust particles are drawn further together, possibly spawning planetesimals. If instead the layer is first rendered unstable by a Kelvin-Helmholtz-type shearing instability (KHI), the resultant turbulence prevents dust from settling further, pre-empting gravitational collapse. In this chapter we investigated the conditions which trigger the KHI, hoping to find a region of parameter space where the KHI might be held at bay so that planetesimals can form by self-gravity.

A fundamental assumption underlying our work is that turbulence intrinsic to gas can, in some regions of the disk, be neglected. There is some consensus that near disk midplanes, in a zone extending from ∼1 to at least ∼10 AUs from the parent star, gas may be too poorly ionized to sustain magnetohydrodynamic turbulence (Ilgner & Nelson 2006; Bai & Goodman 2009; Turner et al. 2010a). Presumably if the magnetorotational instability (e.g., Balbus 2009) cannot operate at the midplane, disk gas there is laminar—pending the uncertain ability of magnetically active surface layers to stir the disk interior (e.g., Turner et al. 2010a), or the discovery of a purely hydrodynamic form of turbulence (Lithwick 2009). To get a sense of how laminar disk gas must be to permit dust sublayers to form, Chiang & Youdin (2010) compared the height to which dust particles are stirred in an "alpha"-turbulent disk to the thickness of the sublayer (2.6). They estimated that the former is smaller than the latter when the dimensionless turbulent diffusivity $\alpha \lesssim 3 \times 10^{-4} \Omega_{\rm K} t_{\rm stop} (r/{\rm AU})^{4/7}$ for $t_{\rm stop} < \Omega_{\rm K}^{-1}$. To place this requirement in context, α values for magnetically active zones are typically quoted to be greater than $\sim 10^{-3}$. Whether magnetically dead zones are sufficiently passive for dust to settle into sublayers remains an outstanding question.

Modulo this concern, we studied the stability of dust layers characterized by spatially constant Richardson numbers Ri using a three-dimensional, spectral, anelastic, shearing box code (Barranco & Marcus 2006) that models gas and dust as two perfectly coupled fluids (Barranco 2009). We found that stability is not characterized by a single critical Richardson number. Rather the value of $Ri_{\rm crit}$ distinguishing layers that overturn from those that do not is a nearly linear function of the midplane dust-to-gas ratio μ_0 (Figure 2.5). Dust-rich sublayers having $\mu_0 \approx 10$ have $Ri_{\rm crit} \approx 0.2$—near the canonical value of 1/4—while dust-poor sublayers having $\mu_0 \approx 0.3$ (still orders of magnitude dustier than well-mixed gas and dust at solar abundances) have $Ri_{\rm crit}$ as low as 0.009.

Previous studies (e.g., Sekiya 1998; Youdin & Shu 2002; Youdin & Chiang 2004) assumed a universal critical Richardson number of 1/4. This popular assumption seems correct only for dust-rich layers having μ_0 so large they are on the verge of gravitational instability. For less dusty midplanes, the assumption appears to be incorrect. Our numerical results are roughly consistent with those of Chiang (2008), who also found evidence that $Ri_{\rm crit}$ decreases with decreasing μ_0. Comparing his Table 2 with our Figure 2.5 shows that his constraints on $Ri_{\rm crit}$ are, for the most part, compatible with those presented here, for the range $\mu_0 \approx 0.3$–10 where our respective data overlap. Our findings supersede those of Chiang (2008) insofar as we have explored parameter space more finely and systematically, at greater and more uniform resolution, with numerical methods better suited for subsonic flows.

Our results turn out to be consistent with the classical Richardson criterion—which states only that $Ri < 1/4$ is necessary, not sufficient, for instability—even though the criterion as derived by Miles (1961) applies only to two-dimensional flows, which our dust layers are not. Our simulations demonstrate that the criterion can still serve as a useful guide for assessing stability in disks having bulk metallicities ranging from subsolar to slightly supersolar values—with the proviso that the actual Richardson number dividing KH-stable from KH-unstable flows, while $< 1/4$, is generally not equal to $1/4$.

Why isn't the Richardson criterion for instability sufficient in rotating dust disks? The criterion considers the competition between the destabilizing vertical shear and the stabilizing influence of buoyancy, which causes fluid parcels to oscillate about their equilibrium positions at the Brunt-Väisälä frequency. However, there exists another stabilizing influence, ignored by the Richardson number, provided by the radial Kepler shear (Ishitsu & Sekiya 2003). In the limit $\mu_0 \ll 1$, the Brunt frequency (2.4) becomes negligible relative to the Kepler shearing frequency (2.7), suggesting stability now depends on the competition between the destabilizing vertical shear and stabilizing radial Kepler shear. We expect the flow to be stable as long as the Kepler shear can wind up unstable eigenmodes to higher radial wavenumbers before their amplitudes grow large enough to trigger nonlinear effects. This suggests that we replace the Richardson number with a "shearing number," defined by analogy as the square of the ratio of the Kepler shearing frequency to the vertical shearing frequency:

$$Sh \equiv \frac{|\partial\Omega/\partial \ln r|^2}{(\partial v_\phi/\partial z)^2} \propto \left(\frac{\Delta z}{\Delta v_\phi}\right)^2 \propto Ri\frac{1+\mu_0}{\mu_0} \tag{2.32}$$

where we have used (2.3) and (2.6). By assuming Sh is constant for marginally stable dust profiles, we arrive at the relation

$$Ri_{\rm crit} \propto \mu_0 \ \text{ for } \ \mu_0 \ll 1\,. \tag{2.33}$$

What is surprising is that this trend, although expected to hold only for $\mu_0 \ll 1$, appears to hold approximately for all μ_0, according to our simulation results in Figure 2.5. For $\mu_0 \gtrsim 1$, we would have expected from (2.32) that $Ri_{\rm crit}$ asymptote to a constant; but it does not. Our higher resolution runs do suggest the stability curve slightly flattens at $\mu_0 \approx 10$, but such deviations seem too small to be fully explained using arguments relying purely on the shearing number.

To explain the observed trend, we might co-opt the methods of Ishitsu & Sekiya (2003), who linearized and numerically integrated the 3D equations of motion for the dust layer. For their particular choice of background vertical density profile, they solved for the maximum growth factors for the most unstable KH modes (see also Knobloch & Spruit 1985 who considered the axisymmetric problem). We would need to replace their assumed profile with our profiles having spatially constant Ri. Perhaps our numerically determined stability curve $Ri_{\rm crit}(\Sigma_{\rm d}/\Sigma_{\rm g})$ corresponds to a locus of fixed maximum growth factor.

Gravitational instability occurs on a dynamical time when the dust layer's Toomre $Q \approx M/[2\pi r^3 \rho_{\rm g}(1+\mu_0)]$ reaches unity (Toomre 1964; Goldreich & Lynden-Bell 1965). For $\rho_{\rm g}$ given by the minimum-mass solar nebula, this occurs when $\mu_0 \approx 30$, fairly independently of r

(Chiang & Youdin 2010). Of course in more massive gas disks (greater $\rho_{\rm g}$), the requirement on μ_0 is proportionately lower. Figure 2.7 shows that for disks having bulk metallicities $\Sigma_{\rm d}/\Sigma_{\rm g}$ equal to the solar value of 0.015, the dusty sublayer can achieve $\mu_0 \approx 8$ before it becomes KH unstable. Taken at face value, such a marginally KH-stable subdisk, embedded in a gas disk having $30/8 \approx 4$ times the mass of the minimum-mass solar nebula, would undergo gravitational instability on the fastest timescale imaginable, the dynamical time. The case that planets form from disks several times more massive than the minimum-mass solar nebula is plausible (e.g., Goldreich et al. 2004b; Lissauer et al. 2009).

An alternate way of crossing the Toomre threshold is to allow the bulk metallicity $\Sigma_{\rm d}/\Sigma_{\rm g}$ to increase above the solar value of 0.015. Extrapolating the boundary of stability in Figure 2.7 to $\mu_0 \approx 30$ suggests that metallicities roughly ∼3 times solar would be required for dynamical gravitational instability in a minimum-mass disk. There are several proposed ways to achieve supersolar metallicities in some portions of the disk, among them radial pileups (Youdin & Shu 2002) or dissipative gravitational instability (Ward 1976; Coradini et al. 1981; Ward 2000; Youdin 2005a; Youdin 2005b; see also the introduction of Goodman & Pindor 2000).

None of the ways we have outlined for achieving gravitational instability rely on the streaming instability or turbulent concentration of particles, mechanisms that we have criticized in §2.1.3. Nevertheless our scenarios may be too optimistic because all our dust profiles are predicated on the assumption of a spatially constant Ri. This assumption tends to generate strong density cusps at the midplane that might not be present in reality. In the next chapter we will relax the assumption of spatially constant Ri and measure the maximum μ_0 attainable, as a function of metalllicity $\Sigma_{\rm d}/\Sigma_{\rm g}$, by simulating explicitly the settling of dust towards the midplane.

Acknowledgments

We thank Daniel Lecoanet, Eve Ostriker, Prateek Sharma, Jim Stone, and Yanqin Wu for discussions. An anonymous referee provided a thoughtful and encouraging report that helped to place our work in a broader context. This research was supported by the National Science Foundation, in part through TeraGrid resources provided by Purdue University under grant number TG-AST090079.

Chapter 3

Forming Planetesimals by Gravitational Instability: II. How Dust Settles to its Marginally Stable State

[1] Dust at the midplane of a circumstellar disk can become gravitationally unstable and fragment into planetesimals if the local dust-to-gas ratio $\mu_0 \equiv \rho_{\rm d}/\rho_{\rm g}$ is sufficiently high. We simulate how dust settles in passive disks and ask how high μ_0 can become. We implement a hybrid scheme that alternates between a 1D code to settle dust and a 3D shearing box code to test for dynamical stability. This scheme allows us to explore the behavior of small particles having arbitrarily short stopping times in gas: $t_{\rm stop} \ll \Omega_{\rm K}^{-1}$, where $\Omega_{\rm K}$ is the Kepler orbital frequency. Dust settles until instabilities at the top and bottom faces of the dust layer threaten to overturn the entire layer. In this state of marginal stability, $\mu_0 = 2.9$ for a disk whose bulk (height-integrated) metallicity $\Sigma_{\rm d}/\Sigma_{\rm g}$ is solar—thus μ_0 increases by more than two orders of magnitude from its well-mixed initial value of $\mu_{0,\rm init} = \Sigma_{\rm d}/\Sigma_{\rm g} = 0.015$. For a disk whose bulk metallicity is 4× solar ($\mu_{0,\rm init} = \Sigma_{\rm d}/\Sigma_{\rm g} = 0.06$), the marginally stable state has $\mu_0 = 26.4$. These maximum values of μ_0, which depend on the background radial pressure gradient, are so large that gravitational instability of small particles is viable in disks whose bulk metallicities are just a few ($\lesssim 4$) times solar, without recourse to the streaming instability which relies on large particles having $t_{\rm stop} \gtrsim 0.1\Omega_{\rm K}^{-1}$. Our result supports earlier studies that assumed that dust settles until the Richardson number Ri is spatially constant. Our simulations are free of this assumption but provide evidence for it within the boundaries of the dust layer, with the proviso that Ri increases with $\Sigma_{\rm d}/\Sigma_{\rm g}$ in the same way that we found in the previous chapter. Because increasing the dust content decreases the vertical shear and increases stability, the midplane μ_0 increases with $\Sigma_{\rm d}/\Sigma_{\rm g}$ in a faster than linear way, so fast that modest enhancements in $\Sigma_{\rm d}/\Sigma_{\rm g}$ can spawn planetesimals directly from small particles.

---

[1] Large portions of this chapter have been previously published as Lee, A. T., Chiang, E., Asay-Davis, X., Barranco, J. 2010. Forming Planetesimals by Gravitational Instability. II. How Dust Settles to its Marginally Stable State. ApJ, 725, 1938.

3.1 Introduction

Dust can settle quickly in gaseous protoplanetary disks. In a passive (non-turbulent) nebula, a particle's vertical height z above the midplane obeys

$$\ddot{z} = -\dot{z}/t_{\rm stop} - \Omega_{\rm K}^2 z \tag{3.1}$$

where the first term on the right-hand side accounts for gas drag, and the second term accounts for stellar gravity when $z \ll r$, the cylindrical radius. Here $\Omega_{\rm K}$ is the Keplerian orbital frequency and

$$t_{\rm stop} \equiv \frac{m v_{\rm rel}}{F_{\rm D}} \tag{3.2}$$

is the momentum stopping time of a particle of mass m moving at speed $v_{\rm rel}$ relative to gas. Expressions for the drag force $F_{\rm D}$ can be found in Adachi et al. (1976) and Weidenschilling (1977). We are interested in small, well-coupled particles having stopping times much smaller than the dynamical time: $\tau_{\rm s} \equiv \Omega_{\rm K} t_{\rm stop} \ll 1$. Spherical particles of radius s and internal density $\rho_{\rm s}$ that experience Epstein drag ($F_{\rm D} \propto v_{\rm rel}$ so that $t_{\rm stop}$ is a constant) settle to the midplane at terminal velocity $-\Omega_{\rm K}^2 z t_{\rm stop}$ in a time

$$t_{\rm settle} \sim \frac{1}{\Omega_{\rm K}\tau_{\rm s}} \sim 10^3 \left(\frac{0.1\,{\rm cm}}{s}\right)\left(\frac{1\,{\rm g\,cm^{-3}}}{\rho_{\rm s}}\right)\left(\frac{F}{1}\right)\,{\rm yr}\,. \tag{3.3}$$

For this and all other numerical evaluations in this chapter, we use a background disk that is F times more massive than the minimum-mass nebula of Chiang & Youdin (2010, hereafter CY10; see Appendix A). For such a disk $t_{\rm settle}$ is nearly independent of stellocentric distance. The assumption that particles are spherical may not be too bad, given that fractal aggregates of grains are expected to compactify as they collide with one another (Dominik & Tielens 1997; Dullemond & Dominik 2005; Ormel et al. 2007).

For millimeter-sized particles, the settling time $t_{\rm settle}$ is much shorter than the disk lifetime, measured in Myr (Hillenbrand 2005; Hernández et al. 2008). By comparison, micron-sized and smaller particles stay suspended at least one scale height above the midplane as long as the gas disk is present. To the extent that collections of particles of different sizes tend to place their mass at the upper end of the size distribution and their surface area at the lower end, we can expect most of the solid mass in disks to sediment out into a thin sublayer, leaving behind the smallest of grains to absorb incident starlight in a flared disk atmosphere. On the whole this picture is consistent with observed disk spectral energy distributions of T Tauri stars, although some models hint that large grains might remain lofted up in a disk two gas scale heights thick (D'Alessio et al. 2006). Settling can only proceed when and where disk turbulence dies, in regions where gas is insufficiently dense to sustain gravitoturbulence (Gammie 2001) and too poorly ionized to be magnetorotationally unstable (Gammie 1996). Our current understanding of disk turbulence easily admits such passive regions. Turner et al. (2010b) found in numerical simulations that even when their disk surface layers were magnetorotationally unstable, grains at the midplane settled much as they would in a laminar flow. Recently Perez-Becker & Chiang (2010) estimated that practically the entire disk would be immune to the magnetorotational instability, in part because polycyclic aromatic hydrocarbons quench the ion density.

In passive disk regions, how far down do dust particles settle? When particles are small enough not to be affected by aerodynamic streaming instabilities (e.g., Bai & Stone 2010a), we expect them to settle until the dust density gradient $\partial\rho_{\rm d}/\partial z$ becomes so large, and the consequent vertical shear in orbital velocity $\partial v_\phi/\partial z$ so strong, that the sublayer is on the verge of overturning by a Kelvin-Helmholtz-type instability (KHI; Weidenschilling 1980). An order-of-magnitude estimate of the minimum layer thickness can be derived using the Richardson number

$$Ri \equiv \frac{-(g/\rho)(d\rho/dz)}{(dv_\phi/dz)^2}\,, \tag{3.4}$$

which if less than some critical value $Ri_{\rm crit}$ may signal that the layer is KH unstable (e.g., Drazin & Reid 2004). Here g is the vertical gravitational acceleration and $\rho = \rho_{\rm d} + \rho_{\rm g}$ is the total density of dust plus gas. In Lee et al. (2010, hereafter Paper I, also Chapter 2), we found that $Ri_{\rm crit}$ increases with $\Sigma_{\rm d}/\Sigma_{\rm g}$, the ratio of dust to gas surface densities, a.k.a. the bulk metallicity. For disks of bulk solar metallicity, we determined empirically that $Ri_{\rm crit} \approx 0.2$.

To translate the Richardson number (3.4) into a critical dust layer thickness, first recognize that the orbital velocity v_ϕ depends on the local dust-to-gas ratio $\mu \equiv \rho_{\rm d}/\rho_{\rm g}$ according to

$$v_\phi = \Omega_{\rm K} r\left(1 - \frac{\eta}{\mu+1}\right) \tag{3.5}$$

in the inertial frame, where

$$\begin{aligned}\eta &\equiv \frac{-(1/\rho_{\rm g})\partial P/\partial r}{2\Omega_{\rm K}^2 r}\\ &\approx \frac{1}{2}\left(\frac{c_{\rm s}}{\Omega_{\rm K} r}\right)^2 \approx \frac{1}{2}\left(\frac{H_{\rm g}}{r}\right)^2\\ &\approx 8\times10^{-4}\left(\frac{r}{\rm AU}\right)^{4/7}\end{aligned} \tag{3.6}$$

is a dimensionless measure of the strength of the background radial pressure gradient $\partial P/\partial r$, with gas scale height $H_{\rm g}$ and sound speed $c_{\rm s}$ (e.g., Nakagawa et al. 1986). When $\partial P/\partial r < 0$, pressure provides extra support against radial stellar gravity and so drives the gas to move on slower than Keplerian orbits. The orbital velocity depends on μ as in (3.5) because dust-laden gas, weighed down by the extra inertia of solids, is accelerated less by the radial pressure gradient than is dust-free gas, and so must hew more closely to Keplerian rotation. Call the critical layer height $\Delta z_{\rm Ri}$ for which $Ri = Ri_{\rm crit}$, and assume the midplane gas-to-dust ratio $\mu_0 \gtrsim 1$ (above the layer $\mu \ll 1$). Then evaluating equation (3.4) with the approximations

$g \approx -\Omega_{\rm K}^2 \Delta z_{\rm Ri}$ (no self-gravity), $\rho^{-1}\partial\rho/\partial z \sim -1/\Delta z_{\rm Ri}$, and $\partial v_\phi/\partial z \sim -\eta\Omega_{\rm K} r/\Delta z_{\rm Ri}$, we find

$$\begin{aligned}
\Delta z_{\rm Ri} &\sim Ri_{\rm crit}^{1/2}\eta r \\
&\sim \frac{1}{2} Ri_{\rm crit}^{1/2}\frac{H_{\rm g}}{r}H_{\rm g} \\
&\sim 5\times 10^{-3}\left(\frac{Ri_{\rm crit}}{0.2}\right)^{1/2}\left(\frac{r}{\rm AU}\right)^{2/7} H_{\rm g}\,. \qquad (3.7)
\end{aligned}$$

Equation (3.7) indicates the dust layer could be quite thin, subtending on the order of 1% of the gas scale height. Is this thin enough for the dust to self-gravitate and hopefully fragment into planetesimals? One can compare the midplane density to the "Toomre density" required for the disk to undergo gravitational instability on the dynamical time $\Omega_{\rm K}^{-1}$ (Safronov 1969; Goldreich & Ward 1973). As reviewed by CY10, the Toomre density is[2]

$$\rho_{\rm Toomre} \approx \frac{M_*}{2\pi r^3} \approx 10^{-7}\left(\frac{r}{\rm AU}\right)^{-3} {\rm g\,cm^{-3}} \qquad (3.8)$$

where the numerical evaluation is for a central stellar mass M_* equal to $1M_\odot$. Now the actual midplane (subscript 0) density is

$$\rho_0 = \rho_{\rm g0} + \rho_{\rm d0} = 2.7\times 10^{-9} F(1+\mu_0)\left(\frac{r}{\rm AU}\right)^{-39/14} {\rm g\,cm^{-3}} \qquad (3.9)$$

which means the midplane dust-to-gas ratio must be

$$\mu_{0,\rm Toomre} \approx 34\left(\frac{1}{F}\right)\left(\frac{M_*}{M_\odot}\right)\left(\frac{r}{\rm AU}\right)^{-3/14} \qquad (3.10)$$

for the midplane density to match the Toomre density. By comparison, in our crude model of a dust sublayer whose height above the midplane cannot be smaller than $\Delta z_{\rm Ri}$, the midplane dust-to-gas ratio cannot exceed

$$\mu_{0,\rm Ri} \sim \frac{\Sigma_{\rm d}/(2\Delta z_{\rm Ri})}{\rho_{\rm g0}} \sim 1\left(\frac{\Sigma_{\rm d}/\Sigma_{\rm g}}{0.015}\right)\left(\frac{0.2}{Ri_{\rm crit}}\right)^{1/2}\left(\frac{r}{\rm AU}\right)^{-4/7} \qquad (3.11)$$

which is nominally smaller than $\mu_{0,\rm Toomre}$ by more than an order of magnitude. Here the bulk (height-integrated) metallicity $\Sigma_{\rm d}/\Sigma_{\rm g}$ is normalized to solar abundance (Lodders 2003), assuming all metals have condensed into grains.

For many years the fact that $\mu_{0,\rm Ri}$ falls short of $\mu_{0,\rm Toomre}$ was believed to rule out the formation of planetesimals by collective effects, self-gravitational or otherwise (e.g., Weidenschilling & Cuzzi 1993). But there are more ways to achieve the Toomre density than

---

[2] Strictly speaking, the Toomre criterion for gravitational instability is derived for two-dimensional disks characterized by surface densities, not volume densities (Toomre 1964; Goldreich & Lynden-Bell 1965). To derive our Toomre volume density, we assign a half-thickness to the disk equal to $c/\Omega_{\rm K}$, where c is the velocity dispersion of the dust + gas mixture. This assignment is not rigorous; see CY10.

vertical settling. A dissipative form of gravitational instability can, in principle, collect particles radially into overdense rings even when self-gravity is weaker than stellar tidal forces (Ward 1976; Ward 2000; Coradini et al. 1981; Youdin 2005a; for a simple explanation of the instability, see the introduction of Goodman & Pindor 2000). It is not clear whether this instability, which operates over lengthscales and timescales longer than those characterizing the Toomre instability by at least a factor of $(\mu_{0,\mathrm{Toomre}}/\mu_0)^2$, can compete with other effects that seek to rearrange dust and gas (e.g., Youdin 2005a).

Another alternative is to invoke larger dust particles that are more weakly coupled to gas; these can clump by the aerodynamic streaming instability (SI; Youdin & Goodman 2005; Johansen et al. 2009; Bai & Stone 2010a; Bai & Stone 2010c). In their 3D numerical simulations, Johansen et al. (2009) reported that particles having $\tau_\mathrm{s} = 0.1$–0.4—corresponding to sizes of a few centimeters at $r = 5$ AU if $F = 1$, and larger sizes if $F > 1$—concentrated so strongly by aerodynamic effects that planetesimals effectively hundreds of kilometers across coalesced within just a few orbits. To obtain this result, Johansen et al. (2009) initialized their simulations by placing the bulk of the disk's solid mass in particles approaching decimeters in size. Bai & Stone (2010a) greatly expanded the range of τ_s modeled and found similar results for their 3D simulations: in the highly turbulent states driven by the SI, instantaneous densities exceeded the Roche density[3] when the disk's solids were all composed of particles having $\tau_\mathrm{s} = 0.1$–1 and the bulk height-integrated metallicity was about twice solar; see run R10Z3-3D in their Figure 5. For this same run, the time-averaged dust-to-gas ratio at the midplane was $\sim$12, a factor of a few less than the Toomre threshold; see their Figure 4 and compare with our equation (3.10). By contrast, when half or more of the disk's solid mass had $\tau_\mathrm{s} < 0.1$, or when disks had smaller metallicities, their simulated densities fell short of the Roche and Toomre densities by more than an order of magnitude. Note that we are quoting from the 3D simulations of Bai & Stone (2010a).

Given how sensitive the SI is to the existence of marginally coupled particles (centimeter to meter sized for $taus \sim 0.1$–1, $r \sim 1$–30 AU, and order unity F), whether enough such particles actually exist in protoplanetary disks for the SI to play a dominant role in planetesimal formation remains an open and delicate issue. Appeal is often made to observed spectral energy distributions and images of T Tauri disks at centimeter wavelengths; these suggest that much of the solid mass is in millimeter to centimeter sized particles (e.g., D'Alessio et al. 2001; Wilner et al. 2005). Larger sized particles are plausibly also present but are not inferred for want of data probing the disk at longer wavelengths. One problem concerns how quickly $\tau_\mathrm{s} \gtrsim 0.1$ particles can be grown, and how they can survive orbital decay by gas drag. In the 3D simulations of Bai & Stone (2010a), the SI clumped particles strongly enough for self-gravity to be significant when $\tau_\mathrm{s} \gtrsim 0.1$ particles comprised more than half of the disk's solid mass. It is unclear whether particle-particle sticking can build up such a population before it is lost to the star by gas drag. This concern is ameliorated by enhancements in particle density (pileups) that may occur as particles drift radially inward (Youdin & Shu 2002; Youdin & Chiang 2004), and by the reduction of drift speeds brought about by multiple

---

[3] The Roche density is that required for a fluid satellite to be gravitationally bound against tidal forces exerted by a central body. It is greater than the Toomre density by a factor of $\sim 7\pi$.

particle sizes (Bai & Stone 2010a; see their Figure 8).

Regardless of which scenario nature prefers—particle concentration by the streaming instability; dissipative gravitation into rings; or dynamical collapse of a vertically settled sublayer, which is the subject of this chapter—all the proposed ways of forming planetesimals depend on knowing how far down dust settles and what maximum dust-to-gas ratios μ_0 can be attained at the midplane. Our order-of-magnitude estimate in equation (3.11) requires testing. Among the most realistic studies of particle settling are those by Johansen et al. (2009) and Bai & Stone (2010a), both of which concentrated on the SI. Johansen et al. (2009) reported that super-centimeter sized particles settled into sublayers in which the midplane-averaged μ_0 ranged from 0.6 to 9.0 as the bulk metallicity ranged from 1 to 3× solar. Bai & Stone (2010a) found that the highest $\tau_{\rm s}$ particles settled the most, driving turbulence that lofted smaller $\tau_{\rm s}$ particles to greater heights. They argued that in their simulations, all of which were characterized by $\max \tau_{\rm s} \geq 0.1$, particles were so strongly stirred by the SI that the KHI never manifested.

To complement these studies, we would like to understand the settled equilibrium states of disks composed entirely of small particles, well but not perfectly coupled to gas ($\max \tau_{\rm s} \ll 1$), isolated from the complicating effects of the streaming instability but not other instabilities like the KHI. Previous attempts in this regard relied on assumed forms for the density profile of settled dust. Barranco (2009) presumed the dust density profile was Gaussian in shape, and did not seek to determine the maximum value of μ_0 *per se.* Chiang (2008) and Paper I, following Sekiya (1998) and Youdin & Shu (2002), assumed the dust density profile had a spatially constant Richardson number. We found in Paper I that under this assumption, in a disk of bulk solar metallicity ($\Sigma_{\rm d}/\Sigma_{\rm g} = 0.015$), the sublayer could remain KH stable for μ_0 as high as 8—a value that is nearly an order of magnitude higher than our crude estimate in (3.11), and as such lowers the hurdle to forming planetesimals by gravitational instability. Nevertheless the true density profile to which small dust grains relax, in the limit of small but non-zero stopping time $\tau_{\rm s}$, remains only guessed at.

This work makes some headway towards finding the fabled marginally stable state, using an approach that is free of the popular but untested assumption that dust settles until the Richardson number equals a constant everywhere. We allow dust grains to fall until they are stopped by whatever instabilities they self-generate. In the calculations presented here we assume that dust begins well mixed with gas in a Gaussian density profile, and then follow the dust into whatever non-Gaussian distribution it seeks to relax. Although we try only a Gaussian initial profile, our method accommodates arbitrary initial conditions.

At the heart of our approach lie two codes. The first code is in one dimension (z) and computes the vertical drift of dust grains at their terminal velocities. Though incapable of deciding whether the density profiles it generates are prone to the KHI (or any other instability), the 1D code can evolve dust profiles for the entire settling time $t_{\rm settle}$, which can be arbitrarily long for arbitrarily small grains. The task of assessing stability is reserved for the second code: the spectral, anelastic, shearing box code of Barranco (2009) which treats gas and dust in the perfectly coupled $\tau_{\rm s} = 0$ limit. Though incapable of allowing dust to settle out of gas, the 3D code accounts for the complicated interplay of vertical shearing and rotational effects to decide whether a given dust layer overturns from the KHI (or some other instability). It tests dynamical stability by running for dozens of dynamical times $t_{\rm dyn} = \Omega_{\rm K}^{-1}$. Our procedure involves alternating between these two codes: allowing dust to settle over some fraction of the settling timescale $t_{\rm settle}$ using the 1D code; passing the results of the 1D code to the 3D code and allowing the dust profile to relax dynamically over timescales $t_{\rm dyn}$; passing the results of the 3D code back to the 1D code for further sedimentation on the settling timescale; and so on, back and forth, until the midplane dust-to-gas ratio stops increasing, at which point the marginally stable state is identified.

In §3.2 we describe our method in full. Results are presented in §3.3, extended in §3.4, and summarized and discussed in §3.5.

3.2 Method

As sketched in §3.1, to find the marginally stable state to which small dust grains relax, we alternate between two codes: a 1D code that tracks how dust drifts toward the midplane on the settling timescale $t_{\rm settle}$, and a 3D shearing box code developed by Barranco (2009) that allows dusty gas to stabilize on the dynamical timescale $t_{\rm dyn} \ll t_{\rm settle}$. The 3D code integrates the anelastic fluid equations for perfectly coupled dust and gas using a spectral method. It includes a background radial pressure gradient to drive a vertical shear. Details about the 3D code are in Barranco & Marcus (2006), Barranco (2009), and Paper I.

Dust and gas are initially well mixed with a spatially constant density ratio: $[\rho_{\rm d}(z)/\rho_{\rm g}(z)]_{\rm init} \equiv \mu_{\rm init} = $ constant. We set $\mu_{\rm init}$ equal to either solar metallicity ($\mu_{\rm init} = 0.015$; Lodders 2003) or four times solar metallicity ($\mu_{\rm init} = 0.06$). To determine the initial form of the dust density profile $\rho_{\rm d}(z)$, we solve the equation for vertical hydrostatic equilibrium where gas is assumed to be initially isothermal:

$$\frac{c_{\rm s}^2}{\rho_{\rm g} + \rho_{\rm d}} \frac{\partial \rho_{\rm g}}{\partial z} = -\Omega_{\rm K}^2 z\,, \tag{3.12}$$

whence

$$\rho_{\rm d} = \mu_{\rm init} \rho_{\rm g0} \exp\left[-\frac{(1+\mu_{\rm init})z^2}{2H_{\rm g}^2}\right] \tag{3.13}$$

for constants $\mu_{\rm init}$, a characteristic initial height $H_{\rm g} \equiv c_{\rm s}/\Omega_{\rm K}$, and the midplane gas density $\rho_{\rm g0}$.

Equation (3.13) defines the initial dust profile used by the 1D code, whose grid extends from $z = 0$ to $z = 3H_{\rm g}$. The 1D Lagrangian code uses particles to track the motion of dust mass. Each particle represents the same amount of dust mass. Any dust density profile $\rho_{\rm d}(z)$ can be converted into particle positions and back again. The closer particles are spaced, the greater is $\rho_{\rm d}$.

Starting with equation (3.13), we proceed as follows:

1) 1D code: Initialize positions of dust particles and establish hydrostatic equilibrium for the gas.

Given $\rho_{\rm d}(z)$, calculate the positions of ~60,000 particles in the 1D code. Also determine the hydrostatic gas density $\rho_{\rm g}(z)$ at each particle's position by solving equation (3.12).[4]

2) 1D code: Settle dust particles by one timestep Δt.

By equating the vertical gravitational force $\propto z$ to the Epstein drag force $F_{\rm D} \propto \rho_{\rm g} v_{\rm rel}$ (e.g., Weidenschilling 1977), assign terminal velocities

$$v_{\rm rel} \propto \frac{z}{\rho_{\rm g}} \tag{3.14}$$

to each particle. Advect each particle vertically downward by a distance $v_{\rm rel}\Delta t$, where Δt is chosen small enough that particles do not overtake one another. Note that the coefficient of proportionality on the right hand side of (3.14), which depends on quantities such as the disk mass parameter F and grain properties s and $\rho_{\rm s}$, does not affect the shapes of the density profiles generated so long as it is the same for all particles.

Bin particle positions and recalculate $\rho_{\rm d}(z)$.

3) 1D code: Repeat (1) + (2) until the midplane dust density $\rho_{\rm d}(0)$ rises by 30%.

4) 1D → 3D code: Insert results of the 1D code into the 3D code and run the 3D code for ten orbits. Let $z_{\rm max}$ be the position of the highest particle in the 1D code, and set the dimensions of the shearing box in the 3D code to be $(L_r, L_\phi, L_z) = (1.455, 2.91, 4)z_{\rm max}$, resolved by $(N_r, N_\phi, N_z) = (32, 64, 128)$ gridpoints.[5] Initialize the 3D code by assigning $\mu(z)$, as calculated by the 1D code, to each horizontal gridpoint (r, ϕ).[6] The 3D code initializes the remaining variables—velocity, gas density, temperature, and enthalpy—to ensure dynamical equilibrium; see section 2.2 of Paper I.

The background radial pressure gradient is parameterized by the variable $v_{\rm max}$:

$$-\frac{1}{\rho_{\rm g}}\frac{\partial P}{\partial r} = 2\Omega_{\rm K} v_{\rm max}\,. \tag{3.15}$$

---

[4] Actually the calculation of $\rho_{\rm g}(z)$ can be neglected to good approximation, as the hydrostatic gas density deviates only slightly from a Gaussian throughout the evolution. Even when $\mu \gg 1$ near the midplane, $\Delta\rho_{\rm g}/\rho_{\rm g} \sim (z/H_{\rm g})^2\mu$, which for our parameters remains much less than unity. In fact, the gas density is practically constant once the dust falls to $z \lesssim 0.1H_{\rm g}$.

[5] These choices imply that every $z_{\rm max}$ length in the r and ϕ directions is resolved by 22 grid points. In the z direction we need $N_z = 128$ points to achieve comparable resolution because the vertical grid differs from the horizontal grid (see section 2.4 of Paper I). As discussed in §3.4.1, we test the robustness of our results to box size by using bigger boxes as the marginally stable state is approached.

[6] The transfer of $\mu(z)$ from the 1D code to the 3D code involved some smoothing because the vertical grid for the 3D code is ~10× coarser than that of the 1D code. We captured all features of the 1D dust profile $\rho_{\rm d}(z)$ to within ~10% for $z \lesssim 0.8z_{\rm max}$. Fractional errors generally increased away from the midplane and were largest at $z_{\rm max}$ where the dust content goes to zero.

We fix $v_{\max} = 0.025c_{\rm s}$ for all simulations.[7] Physically, $v_{\max} \sim c_{\rm s}^2/(\Omega_{\rm K} r)$ represents the difference in azimuthal velocity between pressure-supported dust-free gas and a strictly Keplerian flow.

Before running the 3D code, perturb $\mu(r, \phi, z)$ by an amount

$$\Delta\mu(r, \phi, z) = A(r, \phi)\mu(z)[\cos(\pi z/2z_{\max}) + \sin(\pi z/2z_{\max})], \tag{3.16}$$

where $A(r, \phi)$ is a random variable constructed in Fourier space (see the discussion following equation 31 of Paper I). Fix the root-mean-square of the perturbations to be $A_{\rm rms} \equiv \langle A^2 \rangle^{1/2} = 10^{-3}$.

Run the 3D code for ten orbits.

5) 3D code: Assess stability. Extend simulations beyond ten orbits as necessary to make this assessment. Label the dust profile "KH-unstable" if the horizontally averaged dust-to-gas ratio at the midplane

$$\langle \mu(z = 0) \rangle \text{ as a function of } t$$

decreases by more than 15%. Otherwise, monitor the horizontally averaged vertical kinetic energy at the midplane:

$$\langle \mu v_z^2(z = 0) \rangle/2 \text{ as a function of } t.$$

If $\langle \mu v_z^2 \rangle/2$ monotonically decreases or levels off, label the dust profile "KH-stable." If $\langle \mu v_z^2 \rangle/2$ is increasing towards the end of the simulation, extend the integration an additional ten orbits and re-assess stability. Repeat step (5) as necessary.

6) If "KH-unstable," stop. Identify the last KH-stable dust profile, generated in the iteration just previous to that of the KH-unstable simulation, as the "marginally stable state."[8]

7) If "KH-stable," pass results of the 3D code back to the 1D code and return to step (1). Fit a polynomial $\mu_{\rm poly}(z)$ to the final, horizontally averaged dust-to-gas ratio $\langle \mu(z) \rangle$ as calculated by the 3D code. Adjust the order of the polynomial to capture all features of the profile. If one polynomial is insufficient, use two to create a piecewise function. Convert $\mu_{\rm poly}(z)$ to $\rho_{\rm d}(z)$ by assuming the gas profile to be Gaussian (see footnote 4): $\rho_{\rm d}(z) = \mu_{\rm poly}(z) \cdot \rho_{\rm g0} \exp[-z^2/(2H_{\rm g}^2)]$. Using this $\rho_{\rm d}(z)$, return to step (1) for the next iteration.

---

[7] This value of $v_{\max}$ coincides with the standard value from Paper I. Technically the minimum-mass disk model in Appendix A gives $v_{\max}/c_{\rm s} = 0.036$. See §3.5.2 and Appendix A.1 for how our results might depend on $v_{\max}$.

[8] This marginally stable state will be superseded by the marginally stable state identified under an improved scheme in §3.4.2.

3.3 Results

To orient the reader, in Figure 3.1 we show results obtained from the 1D code only. The dust is initially well mixed with gas at solar metallicity ($\mu_{\rm init} = 0.015$). As dust settles and the midplane dust-to-gas ratio μ_0 increases, sharp cusps appear at the edge of the dust layer where particles pile up vertically. Pileups occur because particle speeds $|v_{\rm rel}| \propto z/\rho_{\rm g}$ decrease with decreasing height. The slowdown is greatest at large height where the gas density gradient is steepest.

Unlike us, Garaud & Lin (2004) did not find vertical pileups because they chose their initial dust profile to have a scale height equal to $0.1H_{\rm g}$. Thus where their dust was located, the gas density gradient was too low to produce pileups. We verified this by inserting their initial profile into our 1D code.

The shapes of the settled dust profiles $\mu(z)$ and their relative spacing in time are independent of the dust internal density $\rho_{\rm s}$, dust particle size s, and the scaling parameter F for disk mass. Changing these parameters only alters the absolute physical time elapsed (equation 3.3). Relative time is tracked by the dimensionless parameter $f \equiv t/t_{\rm settle}$, labeled on this and many subsequent figures.

Below we compare these 1D-only results to those that include the full 3D dynamics. The solar metallicity case is described in §3.3.1. The metal-rich case ($\mu_{\rm init} = 0.06$) is presented in §3.3.2.

3.3.1 Solar Metallicity

Figure 3.2 traces the evolution of dust that starts well mixed with gas at solar metallicity. Plotted are several KH-stable curves from the 3D code resulting from step (5) of our procedure. For ease of comparison with the purely 1D results, the relative timestamps in Figure 3.2, measured by f, coincide with those in Figure 3.1. The leftmost curve at $f = 1.0$ represents the marginally stable state identified using our standard procedure. This state achieves a midplane dust-to-gas ratio of $\mu_0 = 2.45$, about an order of magnitude below the value required for gravitational instability (equation 3.10). In §3.4 we extend our procedure to see if we might achieve still higher dust-to-gas ratios.

Comparing Figures 3.1 and 3.2, we see that the pileups at the edges of the dust layer do not survive in the dynamical 3D code. By $f \approx 0.44$, the pileups are nearly gone. At this point, the vertical extent of the dust layer $z_{\rm max}$ has shrunk to $\sim 0.1H_{\rm g}$, and local gradients in $\rho_{\rm g}$ are too small for appreciable pileups to re-form.

The instability that eliminates the pileups is likely related to the Rayleigh-Taylor instability (RTI), triggered by heavy fluid lying on top of lighter fluid, and we will refer to it henceforth as such. The RTI originates locally at the edges of the dust layer. By contrast, the midplane is relatively stable (at least until the marginally stable state is reached). Another way of seeing this is to note that midplane dust-to-gas ratios μ_0 in Figures 3.1 and 3.2 agree to within 25%. Closer examination reveals that those in Figure 3.2 are consistently higher. This suggests that the RTI transfers some of the dust in the pileups to the midplane.

Figure 3.3 confirms this transfer mechanism. The top middle panel shows that over the course of a 20-orbit-long 3D simulation (iteration #6, occurring at a time $f = 0.31$, out of a total of 19 iterations), dust is redistributed from the layer's edges to the midplane, raising μ_0 by about 20%. Note that the effect of the RTI has been to transport dust toward the midplane, not to higher altitudes. The RTI is confined to where dust is unstably stratified (increasing total density in the direction opposite to gravity).

Compare this behavior with that in the top row of Figure 3.4, which documents a later iteration, #16. The top middle panel shows that an instability has occurred near the edges of the dust layer. Dust is redistributed to higher, not lower, altitudes. The midplane is not affected. The instability at this relatively late stage of settling is probably driven by the vertical shear associated with strong density gradients at the edges of the layer, and we will refer to it henceforth as the Kelvin-Helmholtz instability (KHI). As a result of the KHI, gradients in density and velocity are reduced.

The marginally stable state identified using our standard procedure is displayed in Figure 3.5. The bottom panels show that during the last iteration #19, the usual 30% increase in the midplane μ_0 (left bottom) results in a KH-unstable profile (middle bottom). In the top panels, we re-do iteration #19, this time incrementing μ_0 by only 10% (left top). The resultant profile is KH stable (middle and right top panels), and has $\langle \mu_0 \rangle = 2.45$. In §3.4.2, we modify our standard procedure and extend it to later times to achieve still higher dust-to-gas ratios in stable flows.

The μ-profiles in Figures 3.2–3.5 betray oscillations just inside the edges of the dust layer. We believe these ripples are artificial because when each first appears, it spans only a few grid points of the 3D code: see the $f = 0.054$ profile of Figure 3.2, which shows two nascent ripples. The features probably arise because the truncated Chebyshev series used to model the flow in z has too few terms to adequately capture the steep vertical density gradient (Gibbs 1898). Originating in the 3D code, the ripples are then amplified as mini-pileups in the 1D code. We could have tried to smooth away these oscillations by reducing the order of our polynomial fit (step 7 of our procedure), but chose instead to retain all features of the dust profile generated by both codes to minimize bias. In any case the oscillations are eventually erased by instabilities during the later stages of settling (Figure 3.2). In and of themselves the oscillations do not appear to introduce instabilities, which as discussed above are triggered instead by smooth density gradients—realistically computed—at the boundaries of the layer (top rows of Figures 3.3 and 3.4).

3.3.2 Metal-Rich Case: 4 × Solar Metallicity

Figure 3.6 follows the evolution of dust that is initially well mixed with gas at 4× solar metallicity. It shares the same timeline as Figures 3.1 and 3.2. Thus the last profile marked $f = 1.1$ in Figure 3.6 is attained at a time 10% later than that marked $f = 1.0$ in the other figures. This last profile is the marginally stable state identified using our standard procedure, for the case of supersolar metallicity. It achieves a midplane dust-to-gas ratio of $\mu_0 = 20.3$—large enough to exceed the Toomre threshold in a disk that has twice the gas content of the minimum-mass solar nebula ($F = 2$ in equation 3.10).

The evolution of the metal-rich disk over the course of a total of 21 iterations—some of which are sampled in Figures 3.7–3.9—is similar to that of the solar metallicity disk, with two notable differences. When the unstably stratified pileups of dust collapse (iteration #4, shown in the top row of Figure 3.7), enough dust is transferred to the midplane that μ attains an appreciable maximum there. This bump contrasts with the nearly flat profile seen for the solar metallicity run (Figure 3.3), and persists at least through iteration #16 (Figure 3.8). A second difference is that in every KH-stable simulation following iteration #13, the vertical kinetic energy, although it eventually levels off, ends orders of magnitude higher than where it began (Figure 3.8, and top row of Figure 3.9). Some currents and/or turbulence appear to be sustained as the state of marginal stability is approached.

Related to this second point, we should acknowledge that our standard procedure ignores whatever velocities are present at the end of a given 3D simulation when initializing the velocities of the subsequent 3D simulation. That is, with every iteration, velocities are set anew according to equation (3.5), with vertical and radial velocities reset to zero. The assumption we make in our standard procedure is that whatever velocities are maintained in a KH-stable layer do not stop dust from settling at the local terminal velocity $v_{\rm rel}$. A crude attempt at relaxing this assumption is made in §3.4.2.

3.4 Extensions

In §3.4.1 we test the robustness of our results against the size of our simulation box. In §3.4.2 we modify the procedure of §3.2, pushing to still higher dust-to-gas ratios at the midplane and revising our identification of marginally stable states.

3.4.1 Bigger Box Runs

Box size can artificially affect stability because a given box can only support modes having an integer number of azimuthal wavelengths inside it. Thus too small a box may be missing modes that would otherwise destabilize the layer. To assess whether our box size is too small, we redo the 3D simulations of our standard marginally stable states (iteration #19 of the solar metallicity case and iteration #21 of the metal-rich case), quadrupling simultaneously the azimuthal box size L_ϕ and the number of grid points N_ϕ. By increasing both in tandem, we maintain the same resolution N_ϕ/L_ϕ as that of our standard runs. The results are plotted as dotted lines in the top panels of Figures 3.5 and 3.9. For both the solar metallicity and metal-rich cases, the bigger box runs still yield stable layers, just as the standard box runs do. We conclude that our standard box sizes are probably adequate.

This conclusion is a bit surprising when we compare our standard box size to our findings in Paper I. We found in Paper I that those KH modes that most visibly disrupted the dust layer had azimuthal wavelengths between $2.6 z_{\rm max}$ and $4.3 z_{\rm max}$. Our standard choice here for azimuthal box size is $L_\phi = 2.91 z_{\rm max}$, which at face value means that we are only resolving one

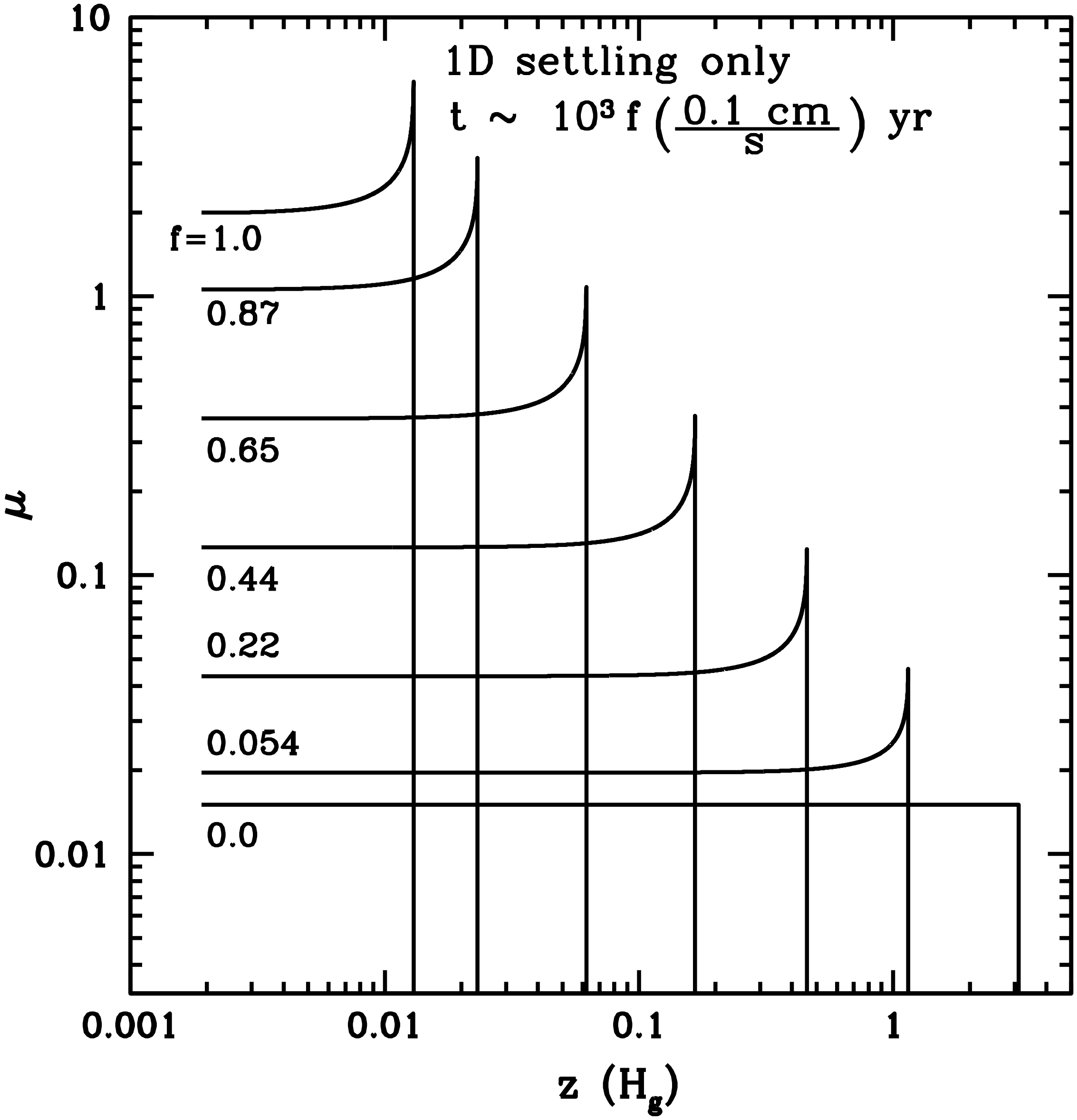

Figure 3.1: Snapshots of settling dust computed with the 1D code only. Plotted is the dust-to-gas ratio as a function of height at various instants of time. Relative timestamps are assigned by the non-dimensional parameter f; see the inset equation for the absolute elapsed time, which assumes $\rho_s = 1$ g/cm^3 and $F = 1$ (equation 3.3). Note that the shapes of these profiles and their relative spacing in time are independent of the absolute elapsed time, and thus independent of ρ_s, F, and s. At $f = 0$, dust begins well mixed at solar metallicity ($\mu = 0.015$; Lodders 2003). By $f = 0.054$, a pileup has formed at the dust layer's edge. The pileup forms because particle speeds $\propto z/\rho_g$ decrease with decreasing height, and the gradient of ρ_g is steepest at the edge.

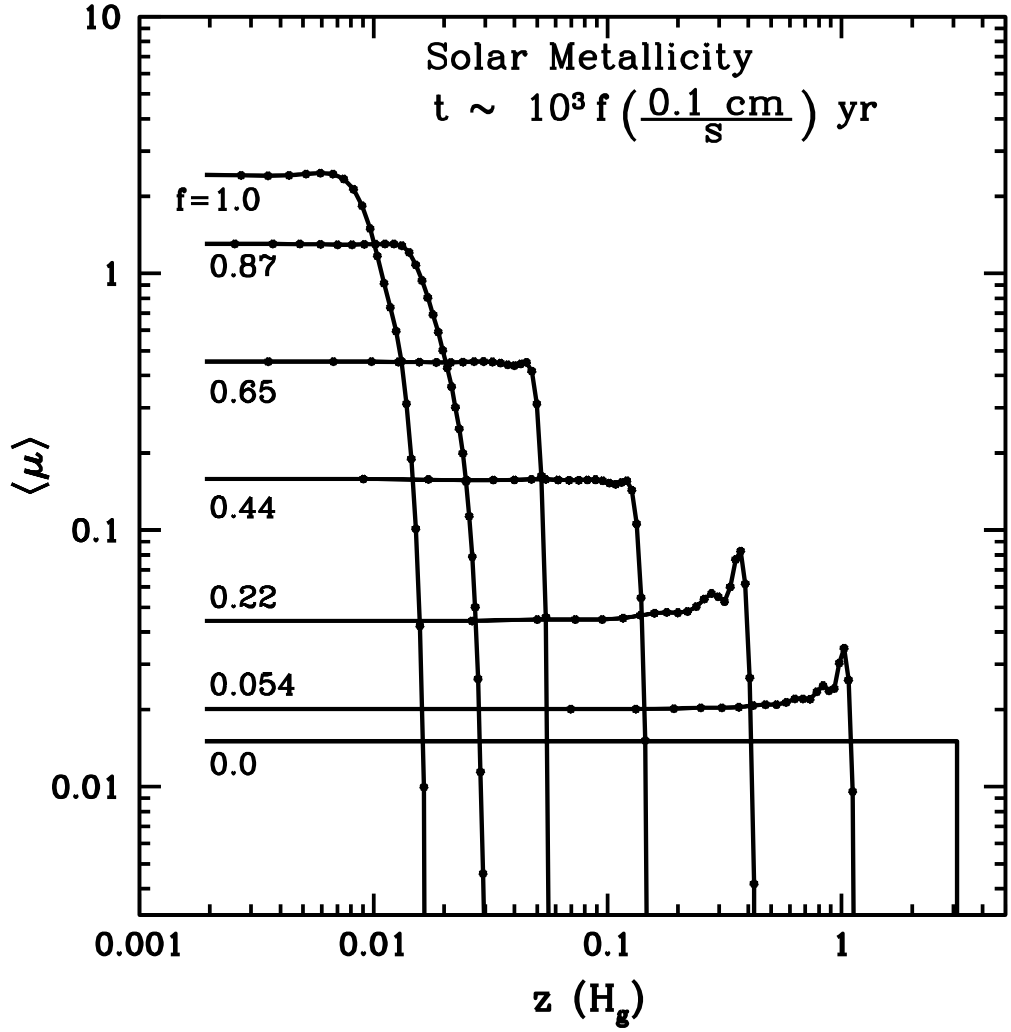

Figure 3.2: Snapshots of settling dust computed using the standard procedure of §3.2 which combines the 1D and 3D codes, for the case of bulk solar metallicity. Elapsed time is marked by f; plotted values coincide with those in Figure 3.1. The shapes of the profiles and their relative spacing in time do not depend on the absolute elapsed time; they are independent of ρ_s, F, and s. Dust begins well mixed with gas at $\mu = 0.015$ and ends in the marginally stable state with midplane $\mu_0 = 2.45$. Vertical gridpoints from the 3D code are plotted as dots. In comparison to the purely 1D results of Figure 3.1, the pileup at the layer's edge is smoothed away, probably by the Rayleigh-Taylor instability, between $f = 0.22$ and $f = 0.65$. Except for transferring some dust at altitude to the midplane, the instability leaves the midplane relatively unaffected, which until $f = 1.0$ evolves much as it does in Figure 3.1.

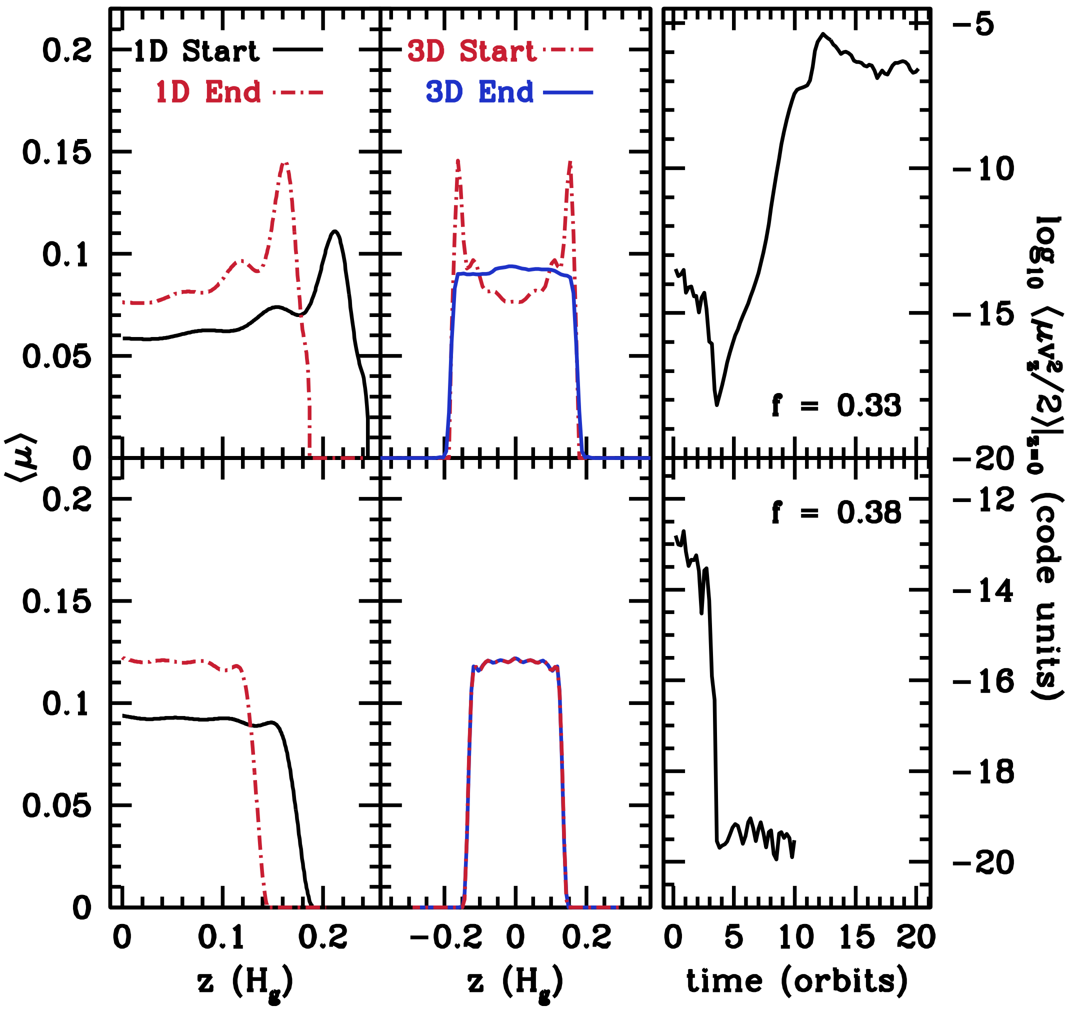

Figure 3.3: Two successive iterations of our procedure of §3.2, for the case of bulk solar metallicity. From left to right, the panels show a starting dust profile (black curve) settled by the 1D code until its midplane μ_0 increases by 30% (red dot-dashed curve). This settled curve is then passed to the 3D code and evolved (blue curve) until it stabilizes (rightmost panel showing how the vertical kinetic energy at the midplane eventually levels off). Top panels show iteration #6 of 19 (equivalently $f = 0.33$ on the timeline of Figure 3.2). The unstably stratified pileups collapse around $t \sim 11$ orbits, increasing the midplane dust content by ~20% (top middle). Bottom panels show iteration #7 ($f = 0.38$) which begins where iteration #6 leaves off—except that the kinetic energy of the flow is reset to a low value (bottom right versus top right panels), and the slight asymmetry in $\langle\mu\rangle$ about $z = 0$ (top middle panel, blue curve) is dropped upon fitting a polynomial only to $z \geq 0$ (bottom left, black curve). The oscillations in the μ-profiles are artifacts of having too few basis functions in z. They did not seem to introduce instability, which always occurred instead at the edges of the dust layer where gradients were steepest and realistically computed.

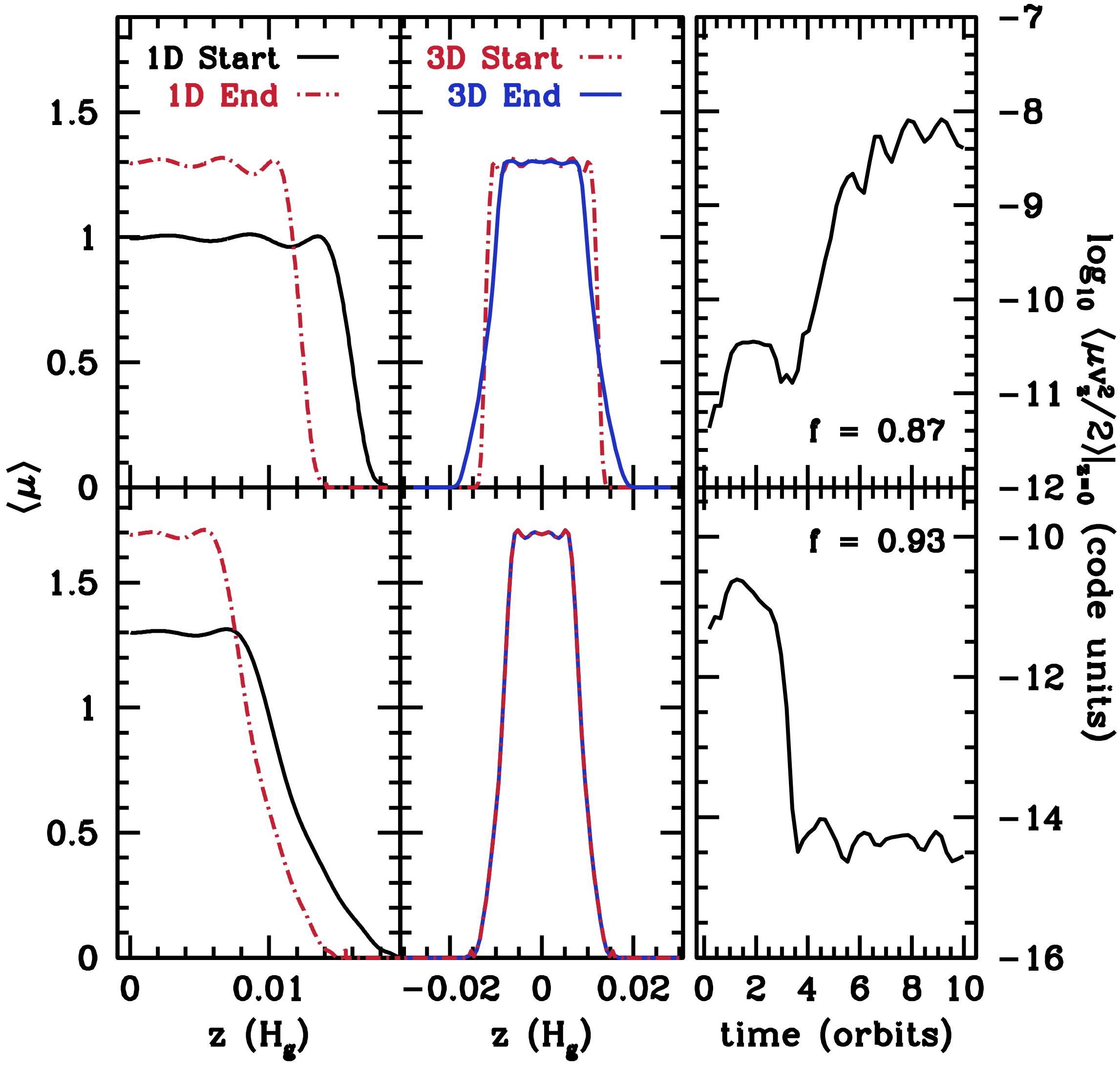

Figure 3.4: Similar to Figure 3.3 but showing iterations #16 (top) and #17 (bottom) out of a total of 19, for the case of bulk solar metallicity. In iteration #16, dusty gas at the layer's edges mixes with dust-poor gas at higher altitudes (top middle), probably by the KHI. The subsequent evolution during iteration #17 shows no sign of instability after 10 orbits.

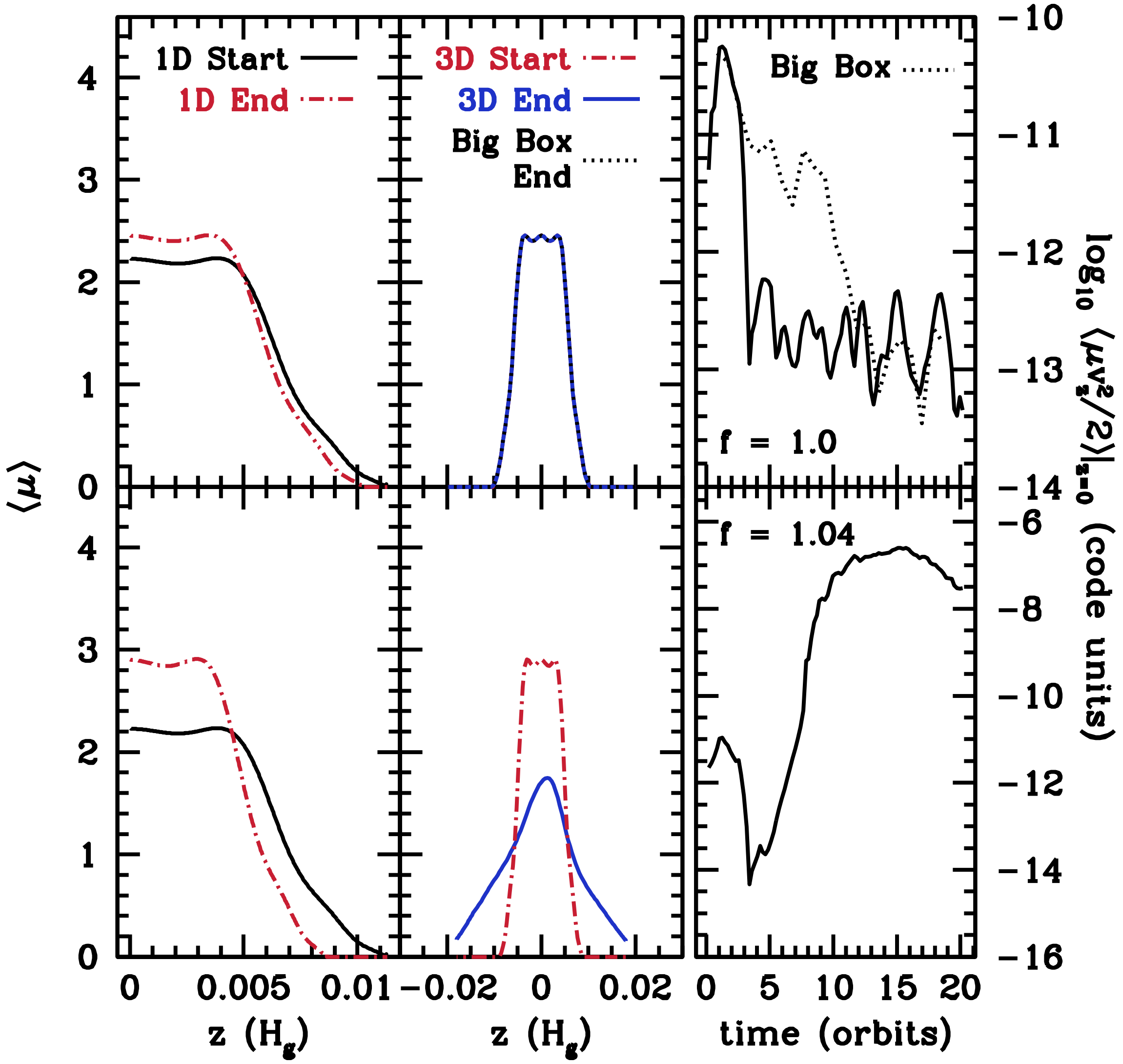

Figure 3.5: Similar to Figures 3.3 and 3.4 but showing the last couple iterations (#19a and #19b) which provisionally identify the marginally stable state for the case of bulk solar metallicity. Increasing the midplane dust content from iteration #18 by 30% (bottom panels) leads to a KH-unstable profile, while an increase of 10% preserves KH stability (top panels). Quadrupling L_ϕ and N_ϕ simultaneously (dotted lines) does not change our answer. The marginally stable state in the top panels is refined according to a modified procedure in §3.4.2.

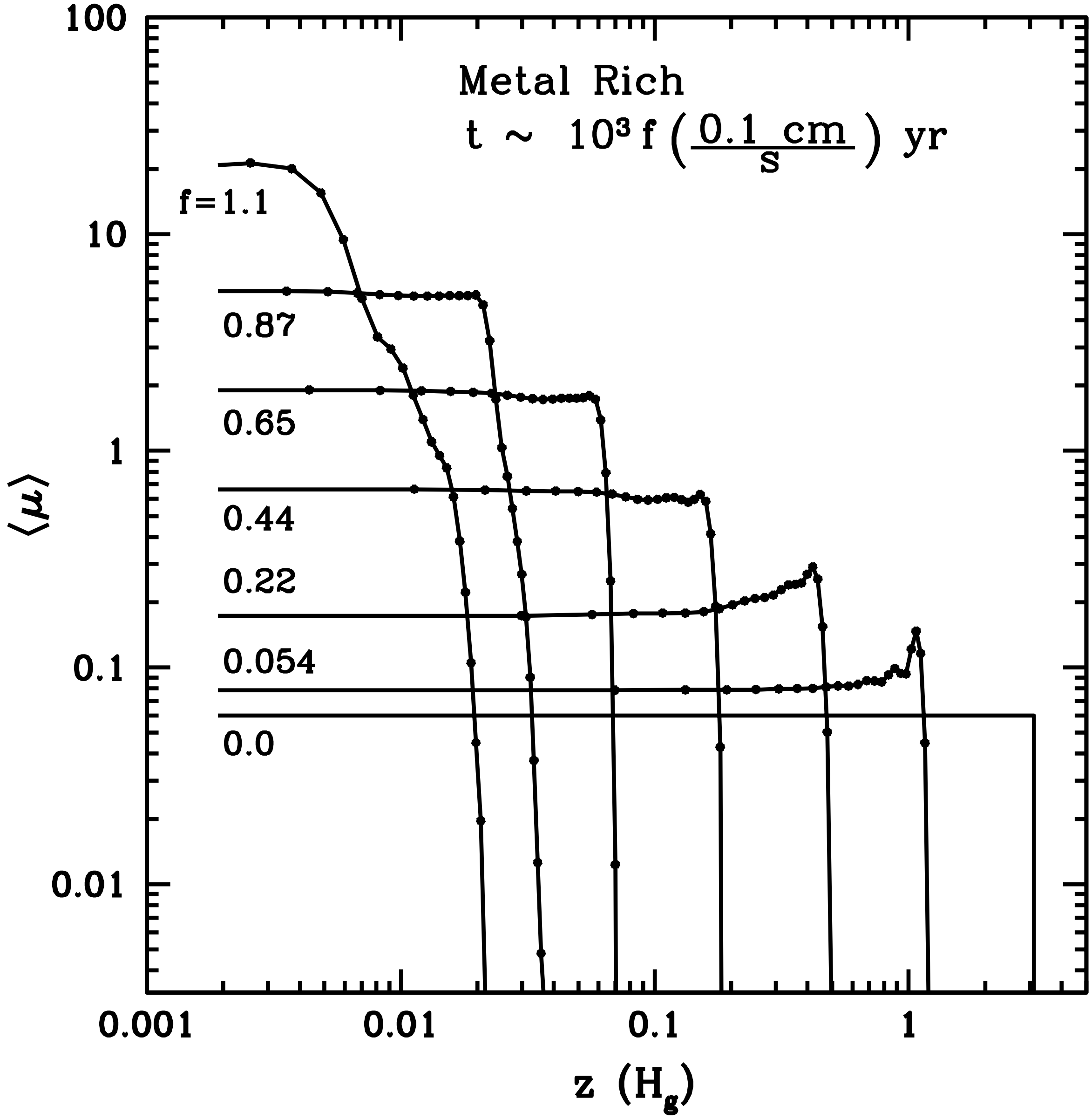

Figure 3.6: Snapshots of settling dust computed with the full procedure of §3.2 which combines the 1D and 3D codes, for the case of 4× bulk solar metallicity. Elapsed time is marked by f, measured on the same timeline characterizing Figures 3.1 and 3.2. The shapes of the profiles and their relative spacing in time do not depend on the absolute elapsed time; in this sense the evolution is not sensitive to ρ_s, F, and s. Vertical gridpoints from the 3D code are plotted as dots. Dust begins well mixed with gas at $\mu = 0.06$ and ends in the marginally stable state with midplane $\mu_0 = 20.3$. The midplane density in this last state already exceeds the threshold for Toomre instability in a disk with twice the gas content of the minimum mass solar nebula (equation 3.10 with $F = 2$).

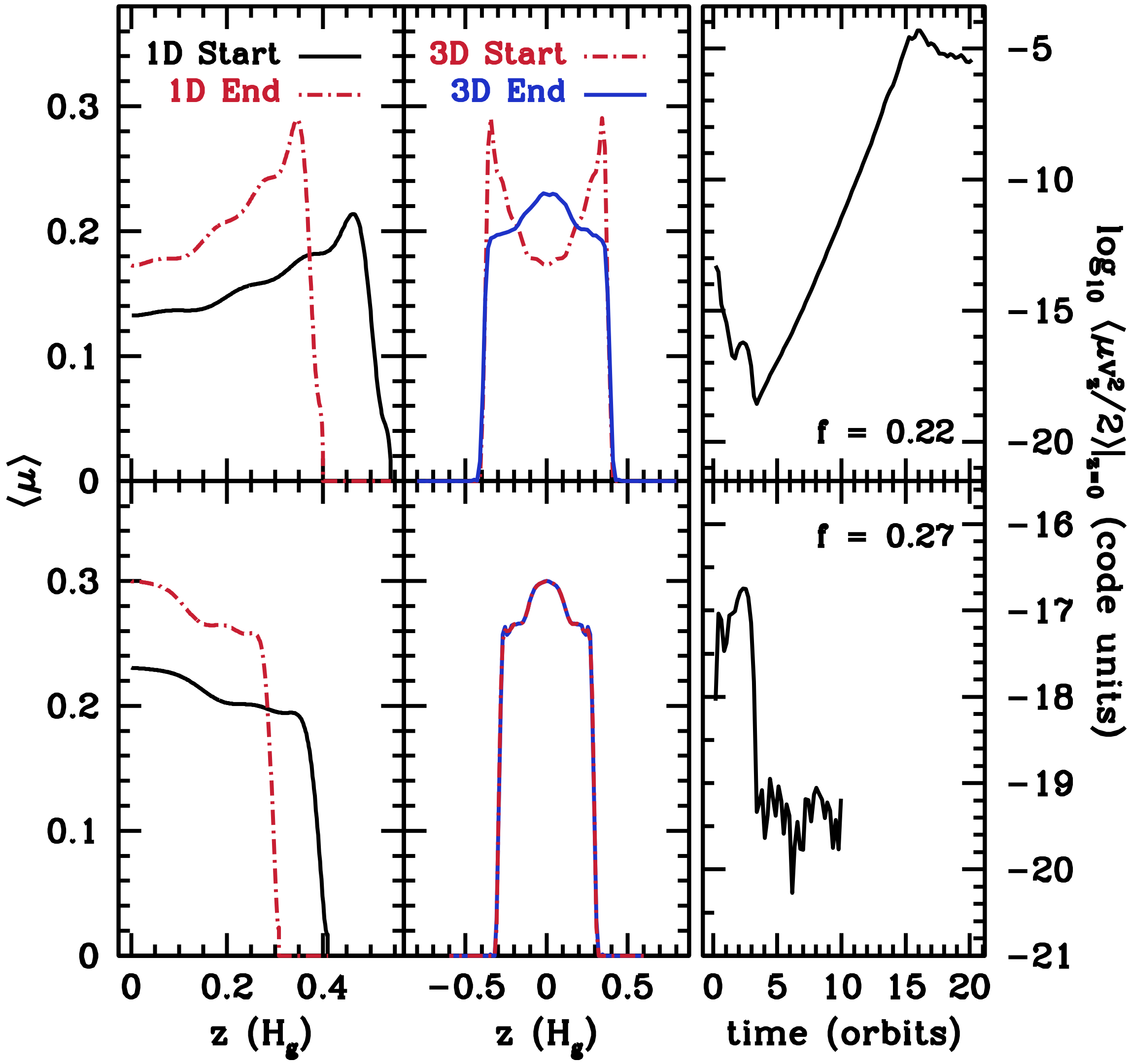

Figure 3.7: Two successive iterations of our procedure outlined in §3.2, for the case of 4× solar metallicity. From left to right, the panels show a starting dust profile (black curve) settled by the 1D code until its midplane μ_0 increases by 30% (red dot-dashed curve). This settled profile is then passed to the 3D code and evolved (blue curve) until it stabilizes (rightmost panel showing how the vertical kinetic energy at the midplane eventually levels off). Top panels show iteration #4 of 21 (equivalently $f = 0.22$ on the timeline of Figure 3.6). When the unstably stratified pileups collapse, they increase the dust content of the midplane by ∼35% (top middle panel). The resultant dust profile, settled further in iteration #5 (bottom panels), remains free of instabilities after 10 orbits.

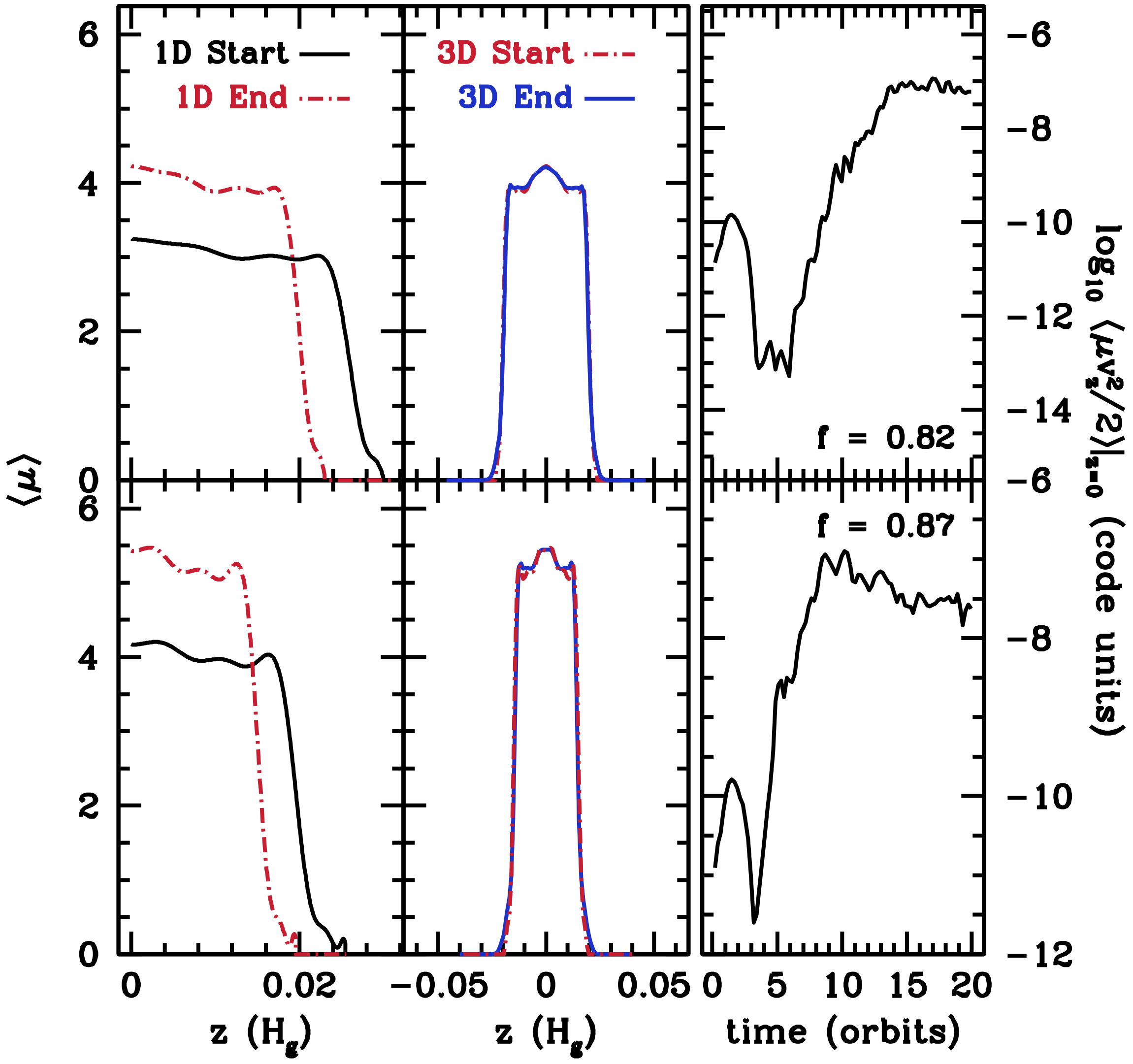

Figure 3.8: Similar to Figure 3.7 but showing iterations #15 and #16 out of a total of 21 for the case of 4× solar metallicity. Shown are two examples of KH-stable profiles whose midplane vertical kinetic energies end orders of magnitude above their starting values. Every 3D simulation starting with iteration #13 in the metal-rich case shows this kind of sustained motion even though the density profiles may be KH stable according to our criterion.

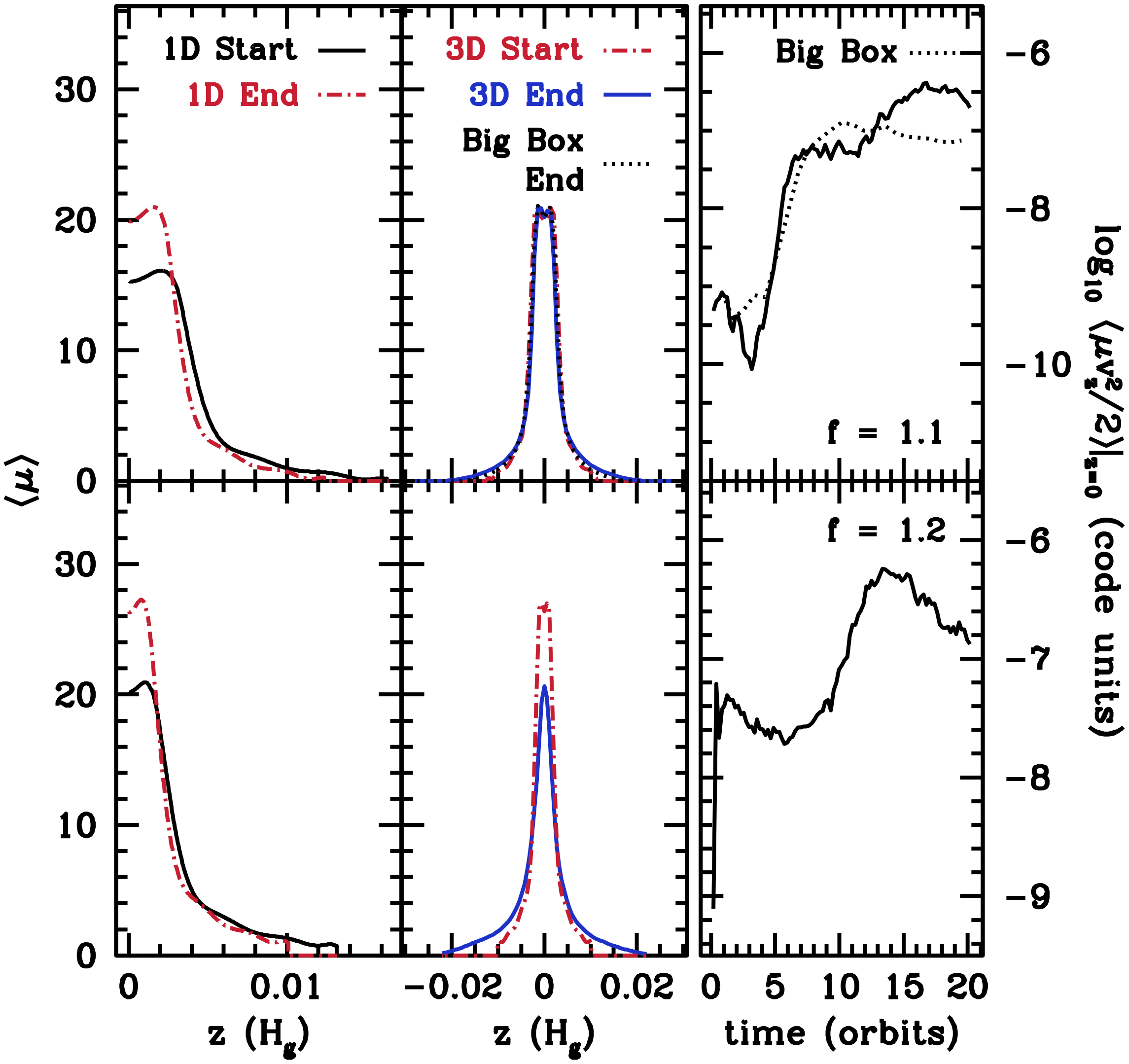

Figure 3.9: Similar to Figures 3.7 and 3.8 but showing iteration #21 in the top panels, in which the marginally stable state is found for the case of 4× solar metallicity according to our standard procedure. The midplane $\langle \mu_0 \rangle = 20.3$, corresponding to a midplane density that exceeds that required for gravitational instability in a disk having twice the gas content of the minimum mass solar nebula. The same dust profile inserted into a shearing box four times as wide in the azimuthal direction as our standard box and having four times as many azimuthal grid points yields qualitatively the same result (dotted line). Settling still further according to our standard procedure results in KH instability (bottom panels), but in §3.4.2 we experiment with a modified procedure that tries to hold off KH instability for a while longer.

wavelength at best of an important mode. However, this simple comparison may not be fair. In Paper I we studied dust layers characterized by a spatially constant Richardson number. The vertical density profiles there differ somewhat from those derived here. In particular the profiles in Paper I have steep density gradients everywhere, whereas here density gradients are steep only at the edges of the layer. When a layer in Paper I became KH unstable, it seemed to do so everywhere at once, whereas here instability always originates at the edges. We have verified that this is true even for the final iterations leading to our identification of the marginally stable state. Obviously these edges have vertical thicknesses that are smaller than that of the layer as a whole. Since the most unstable azimuthal wavelength of the KHI is expected to be of order the vertical thickness of the shearing layer (e.g., Chandrasekhar 1981), it seems that our standard box sizes here, though small compared to our box sizes in Paper I, permit us to resolve several azimuthal wavelengths of those modes that are most able to disrupt the thin edges.

3.4.2 Refining the Marginally Stable State Using a Modified Settling Procedure

Using our standard procedure of §3.2, we can only provisionally identify marginally stable dust profiles. The identification is provisional because by settling all dust particles at their local terminal velocities $v_{\rm rel}$, we wind up with edges so unstable that they also destabilize the midplane. In reality, dust particles at the edge may stop settling because they attain a state of marginal stability first, remaining lofted up by the gas motions they stir up locally, and leaving dust particles near the midplane free to keep settling. In other words, marginal stability may be reached sequentially, starting from the edges and ending at the midplane. Our standard procedure does not allow for this kind of gradual evolution because the 1D code settles all dust particles at their local terminal speeds regardless of their location. In this sense our standard procedure is too blunt because it does not allow for slower settling at the stirred-up edges and faster settling at the quiescent midplane. True marginally stable states should have midplane dust-to-gas ratios even higher than the maximum ones displayed in Figures 3.2, 3.5, 3.6, and 3.9.[9]

To remedy this shortcoming, we modify our procedure by applying a weighting function $W(z)$ to each dust particle's settling velocity. Starting with a KH-stable state near the end of our standard sequence of iterations, we continue to let dust particles drift to the midplane in the 1D code, not at $v_{\rm rel}(z)$ but at the weighted velocity $W(z)v_{\rm rel}(z)$. We use either a Fermi function

$$W(z) = \frac{1}{1+\exp[(z-z_{50})/z_{\rm w}]}, \tag{3.17}$$

described by two parameters z_{50} and $z_{\rm w}$, or a Gaussian

$$W(z) = \exp(-z^2/2z_{\rm w}^2), \tag{3.18}$$

---

[9] Another reason our dust profiles underestimate actual dust-to-gas ratios is because we neglect vertical self-gravity, which enhances stability by increasing the Brunt frequency (Sekiya 1998; Youdin & Shu 2002).

described by a single parameter $z_{\rm w}$. The choice of weighting function is somewhat arbitrary; it depends on the shape of the dust profile to be settled and is made case-by-case according to considerations outlined below.

We start with the KH-stable profile in iteration #18 of our solar metallicity sequence (black solid curve in top left panel of Figure 3.10). The dust layer is characterized by a "core" from $z = 0$ to $z \approx 0.5z_{\rm max}$ over which $\langle\mu\rangle$ is fairly constant, and an "edge" from $z \approx 0.5z_{\rm max}$ to $z = z_{\rm max}$ over which the dust content drops to zero. Because the instabilities that threaten to disrupt the layer originate in the edge and not the core, we seek a weighting function $W(z)$ that slows the downward drift of dust in the edge but not in the core. At the same time $W(z)$ should not be so strongly weighted toward the midplane that the core disconnects from the edge and opens a local gap in dust content. We find upon experimentation that a Gaussian does not have enough flexibility to meet these requirements for this particular iteration. However a Fermi function with $z_{50} = 0.005H_{\rm g}$—corresponding approximately to the boundary of the core—and $z_{\rm w} = 0.05z_{50}$ seems to work well (blue dashed curve in top left panel of Figure 3.10). We use this weighting function to settle the dust profile until its midplane μ_0 increases by 30% to a value of 2.9 (red dot-dashed curve). This settled layer remains KH stable (top middle and right panels of Figure 3.10)—unlike the layer settled without the weighting function (bottom panels of Figure 3.5).

In the new profile settled with our modified procedure, the distinction between the core and the edge is no longer so sharp. Thus to settle this new profile even further, a simple Gaussian for the weighting function seems to suffice. Choosing $z_{\rm w} = 0.0025H_{\rm g} \approx 0.25z_{\rm max}$, we attempt to increase the midplane μ_0 yet again by 30%, but find the resultant profile to be KH unstable (bottom panels of Figure 3.10).

Similar results are obtained for the metal-rich case (Figure 3.11). Using Gaussian weighting functions we are able to push the midplane dust-to-gas ratio μ_0 to a new record of 26.4, which is 30% greater than the value attained using our unweighted standard procedure.

3.5 Summary and Discussion

To form rocky planets and gas giant cores, dust must amass in a circumstellar disk. In the classic scenario for forming planetesimals, dust settles vertically toward the midplane into an ever thinner and denser layer that eventually becomes gravitationally unstable (Safronov 1969; Goldreich & Ward 1973). Toomre's criterion for gravitational instability on a dynamical time is satisfied for midplane dust-to-gas ratios $(\rho_{\rm d}/\rho_{\rm g})_0 \equiv \mu_0 \gtrsim \mu_{0,\rm Toomre}$, where $\mu_{0,\rm Toomre} \approx 34$ for a minimum-mass nebula orbiting a solar-mass star (equation 3.10; note that $\mu_{0,\rm Toomre}$ is nearly independent of disk radius). For comparison, in a disk of well-mixed dust and gas at solar abundance, $\mu_0 \approx 0.015$ (Lodders 2003). Whether dust can accumulate until its density increases by more than three orders of magnitude depends on how turbulent the ambient gas is. Even supposing that gas in certain regions of the disk is not intrinsically turbulent (e.g., because it is too weakly ionized to support the magnetorotational instability), the dust itself may excite turbulence in gas by a Kelvin-Helmholtz-type shearing

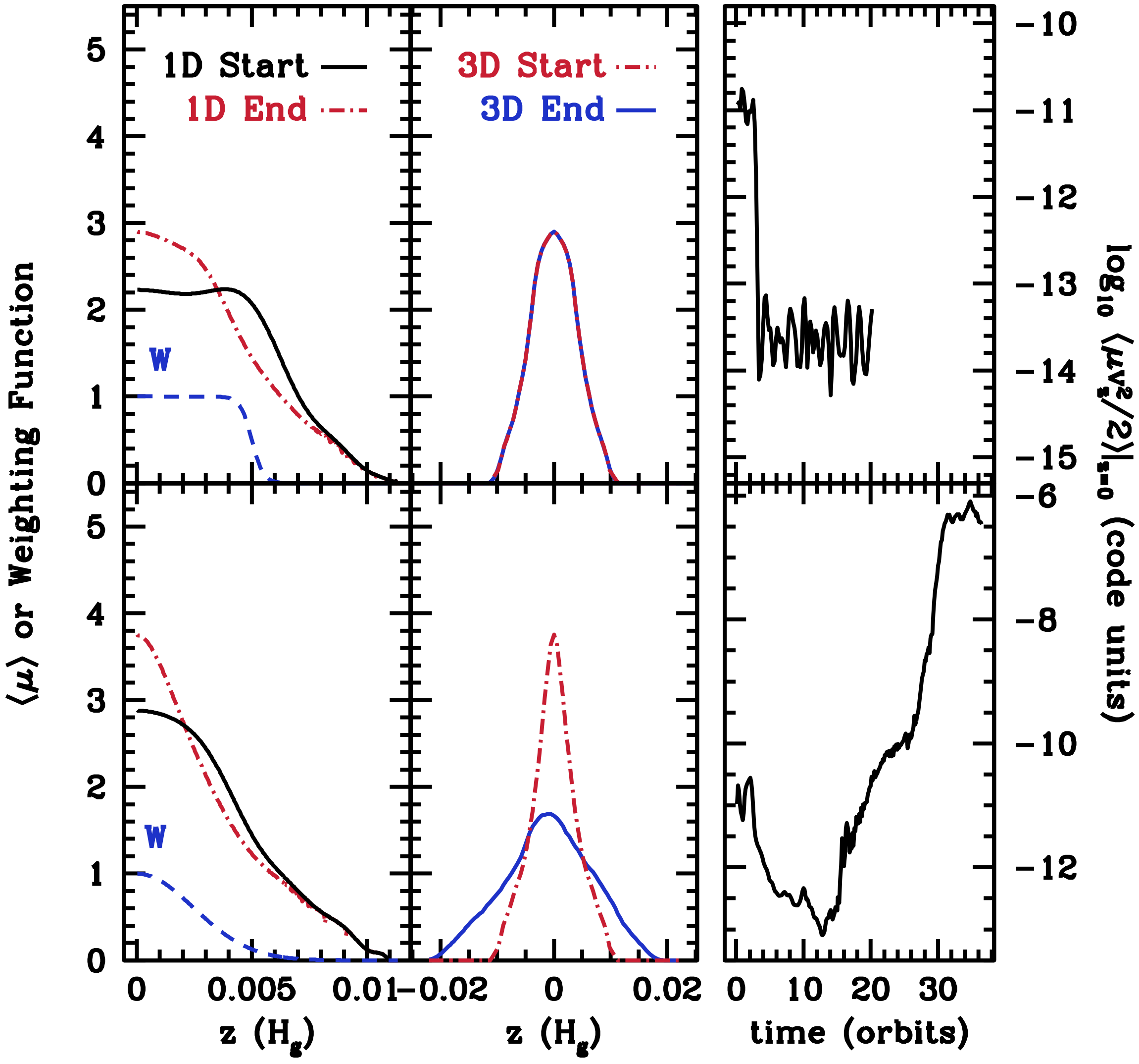

Figure 3.10: Extended settling simulations using the modified procedure of §3.4.2, for the case of bulk solar metallicity. We start with iteration #18 from our standard procedure (black curve, top left). A Fermi weighting function with $z_{50} = 0.005H_\mathrm{g}$ and $z_\mathrm{w} = 0.05z_{50}$ (equation 3.17, labeled '*W*' at top left) allows dust near the midplane to settle more than dust at higher altitude. The settled profile attains a midplane $\langle\mu_0\rangle = 2.9$ and is KH stable (top middle and right panels). Further settling, this time using a Gaussian weighting function with $z_\mathrm{w} = 0.0025H_\mathrm{g}$, results in KH instability (bottom panels). Although the modified procedure enables us to settle beyond the last stable state identified using our standard procedure, the gains are not large enough to reach the Toomre density in solar metallicity disks.

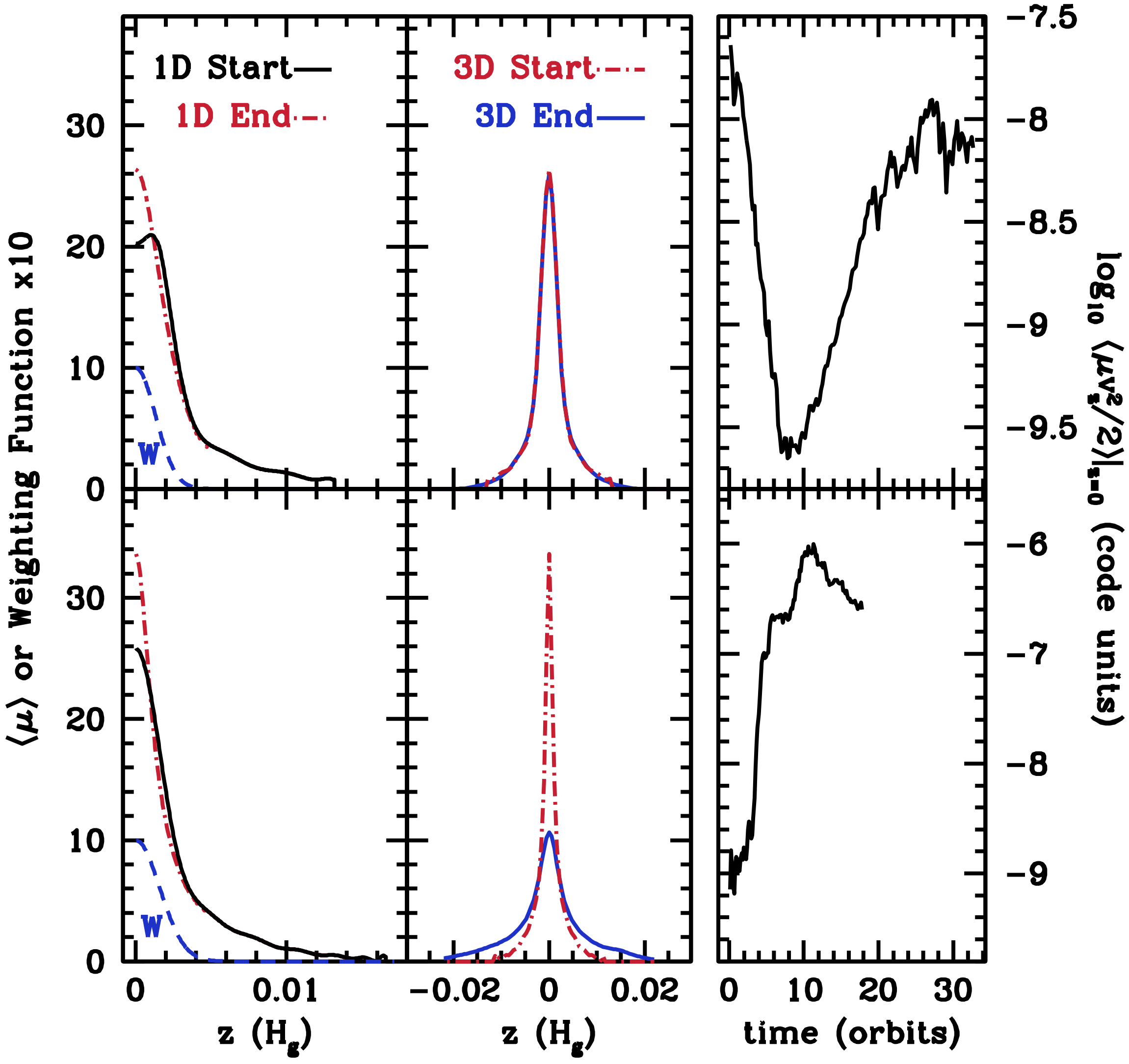

Figure 3.11: Extended settling simulations using the modified procedure of §3.4.2, for the case of 4× bulk solar metallicity. We start with iteration #21 from our standard procedure (black curve, top left). A Gaussian weighting function with $z_{\rm w} = 0.00132H_{\rm g} = z_{\rm max}/12$ is used to settle preferentially the innermost layers, which achieve a maximum $\langle\mu_0\rangle = 26.4$ and remain KH stable (top middle and right panels). Although further gains in μ_0 did not materialize (bottom panels, using a Gaussian of $z_{\rm w} = 0.00127H_{\rm g}$), μ_0 is already high enough that gravitational instability is viable in a disk having $\sim$1–2 times the gas content of the minimum-mass solar nebula.

instability (KHI). The KHI is triggered when the velocity gradient between dust-rich gas at the midplane and dust-poor gas at altitude becomes too large. Barring gravitational instability, dust should settle to a state that is marginally stable against the KHI. The question of whether gravitational instability is viable translates into the question of whether the state that is marginally stable to the KHI has $\mu \gtrsim \mu_{0,\mathrm{Toomre}}$.

In this chapter, we sought out such marginally stable states by numerical simulation. Starting with dust well mixed with gas at either bulk solar or supersolar metallicity, we allowed dust to settle vertically until dynamical instabilities prevented the midplane density from increasing further. We tracked the approach to the marginally stable state using a combination of a 1D settling code and a 3D shearing box code, working in the limit that particles are small enough not to excite streaming instabilities. All the instabilities that afflicted our dust layer originated at the layer's edges, where dust density gradients were steepest. We found evidence for two kinds of instabilities: the usual KHI driven by vertical shear, and the Rayleigh-Taylor instability (RTI) driven by the weight of piled-up dust. These instabilities were mostly confined to the dust layer's top and bottom surfaces, leaving dust near the midplane free to settle but occasionally speeding up the accumulation of solids by transferring dust from pile-ups downward. The midplane density stopped increasing when the dust layer became so thin that instabilities at the edges threatened to overturn the entire layer.

Using our standard procedure of §3.2, we attained maximum dust-to-gas ratios of $\mu_0 \approx 2.45$ and 20.3 for the cases of solar and 4× solar bulk metallicity, respectively (Figures 3.5 and 3.9). These values are lower limits because in our standard procedure dust particles at the layer's top and bottom faces keep settling until they excite instabilities so vigorous that dust at the midplane is stirred up. In reality, dust at the layer's edges may reach a state of marginal stability and stop settling, leaving dust near the midplane free to settle further. We modified our procedure in §3.4.2 to try to account for this effect, reaching $\mu_0 \approx 2.9$ and 26.4 for the two metallicity cases (Figures 3.10 and 3.11). These values are still lower limits because our simulations omit self-gravity. But the correction for self-gravity should be small for the solar metallicity disk, on the order of 10% ($\sim 2.9/34$). For our supersolar metallicity disk, the correction for self-gravity might be on the order of unity ($\sim 26.4/34$)—although it might also be much higher, as Sekiya (1998) and Youdin & Shu (2002) showed that vertical self-gravity can yield a singularity in μ_0.

We conclude that a minimum-mass disk of bulk (height-integrated) solar metallicity orbiting a solar-mass star cannot form planetesimals by self-gravity alone: even neglecting turbulence intrinsic to gas, the KHI would force the midplane dust density to fall short of the Toomre density by about an order of magnitude. Our results make clear what changes to the circumstellar environment would be needed for self-gravity to prevail. To attain the Toomre density in a minimum-mass gas disk, the bulk metallicity would need to be enhanced over solar by a factor of a few $\lesssim 4$. Naturally, the required degree of metal enrichment would be lower in a disk whose total mass (gas plus dust) were greater than that of the minimum-mass solar nebula.

3.5.1 How Spatially Constant is the Richardson Number?

In Paper I, as in previous works (Sekiya 1998; Youdin & Shu 2002; Youdin & Chiang 2004), all dust profiles were assumed to have spatially constant Richardson numbers Ri. The dust profiles we have computed are free of this assumption, whose validity we can now test.

We calculate $Ri(z)$ for our marginally stable states, derived under both standard (§3.2) and modified (§3.4.2) procedures. To compute the numerator (Brunt frequency) of Ri in equation (3.4), we use the horizontally averaged dust-to-gas ratio $\langle\mu(z)\rangle$, computing derivatives using centered differences and assuming the gas to obey a Gaussian density profile (see footnote 4). To compute the denominator (vertical shearing rate) of Ri, we also use $\langle\mu(z)\rangle$, inserting it into equation (3.5) and computing therefrom the velocity derivative. Of course we could also compute the denominator more directly by using the simulation output itself for v_ϕ—this alternative approach turns out to give identical results for the solar metallicity disk, but for the metal-rich disk the $Ri(z)$ so generated varies much more erratically. As noted in §3.3.2, the metal-rich disk sustains gas motions well above those we put in as seed noise. These motions are not strong enough to overturn the dust layer but they are large enough to render Ri highly variable, both in time and space. By not using the simulation data for $v_\phi(z)$ and relying instead on the better behaved $\langle\mu(z)\rangle$, we effectively average Ri in time and space.

Results for the solar metallicity runs are shown in Figure 3.12. We plot Ri only where $\langle\mu\rangle$-gradients are large enough to be computed reliably—thus we avoid regions closest to the midplane. Although we find that Ri is not a strict constant, it varies only between 0.1 and 0.3 within a large fraction of the edges of the dust layer—precisely where instabilities, presumably shear-driven, have rearranged dust. This result compares favorably with Paper I, where we found that a solar metallicity disk has a critical Richardson number of $Ri_{\rm crit} \approx 0.2$.

Evidence for a constant Ri within the edges of the dust layer is even stronger for the metal-rich disk, as shown in Figure 3.13. Here Ri hovers near 0.5 over much of the edges—again consistent with Paper I. See Figure 6 of that paper; admittedly the curve for $Ri_{\rm crit}(\Sigma_{\rm d}/\Sigma_{\rm g})$ in Paper I needs to be extrapolated to the supersolar metallicity considered here, $\Sigma_{\rm d}/\Sigma_{\rm g} = 0.06$.

The $Ri(z)$ profiles in Figures 3.12 and 3.13 differ from those in Figure 3 of Bai & Stone (2010a); see the dashed curves corresponding to their 3D simulations, all of which include marginally aerodynamically coupled particles. These differences support their arguments that their simulations were not afflicted by the KHI.

In summary, the assumption made in other studies that well coupled particles settle into a profile for which Ri is spatially constant does not appear too bad. The caveat is that we have not tested this assumption for those regions closest to the midplane, as they could not relax by our method before being disrupted by instabilities at the edges. This is an area where more work can be done; see §3.5.3. Another caveat, supported independently by Paper I, is that the critical value of Ri to which dust relaxes is not unique, but increases with bulk metallicity $\Sigma_{\rm d}/\Sigma_{\rm g}$. For a solar metallicity disk, $Ri_{\rm crit} \approx 0.2$, but for one having 4× solar metallicity, $Ri_{\rm crit} \approx 0.5$. This trend has not yet been explained.

3.5.2 The Super-Linear Relation Between Maximum Dust-to-Gas Ratio μ_0 and Bulk Metallicity $\Sigma_{\rm d}/\Sigma_{\rm g}$

The degree to which *Ri* is constant is related to the scaling between the maximum midplane dust-to-gas ratio μ_0 and bulk metallicity $\Sigma_{\rm d}/\Sigma_{\rm g}$. Naively it might be expected that μ_0 scales linearly with $\Sigma_{\rm d}/\Sigma_{\rm g}$—a greater total amount of metals simply yields a proportionately dustier midplane—and indeed a linear relation is implied by our order-of-magnitude estimate in equation (3.11). But we did not find a linear trend in our simulations. We find instead that the relation is super-linear: a factor of 4 increase in $\Sigma_{\rm d}/\Sigma_{\rm g}$ results in a factor of 9.1 increase in maximum μ_0 (26.4 versus 2.9).

A super-linear trend is predicted by theories assuming a constant Richardson number (Sekiya 1998; Youdin & Shu 2002). The large gain in midplane density afforded by a comparatively modest increase in bulk metallicity is partly what makes planetesimal formation by gravitational instability so attractive. Increasing $\Sigma_{\rm d}/\Sigma_{\rm g}$ does more than just increase the total amount of metals in the disk—it also helps to stabilize it, by decreasing the vertical shear $\partial v_\phi/\partial z$. In the limit $\mu_0 \sim (\Sigma_{\rm d}/\Sigma_{\rm g})H_{\rm g}/\Delta z \gg 1$, where Δz is the dust layer thickness, we have from equation (3.5):

$$\frac{\partial v_\phi}{\partial z} \sim \frac{\eta\Omega_{\rm K} r/\mu_0}{\Delta z} \sim \frac{\eta\Omega_{\rm K} r}{H_{\rm g}}\frac{1}{\Sigma_{\rm d}/\Sigma_{\rm g}}$$

which decreases with increasing $\Sigma_{\rm d}/\Sigma_{\rm g}$. By comparison the Brunt frequency $[(g/\rho)\partial\rho/\partial z]^{1/2} \sim [(\Omega_{\rm K}^2\Delta z/\mu)\mu/\Delta z]^{1/2} \sim \Omega_{\rm K}$ hardly changes with $\Sigma_{\rm d}/\Sigma_{\rm g}$. Thus the Richardson number increases as $\Sigma_{\rm d}/\Sigma_{\rm g}$ increases; the enhanced stability allows Δz to decrease; and thus $\mu_0 \propto \Sigma_{\rm d}/\Delta z$ scales super-linearly with $\Sigma_{\rm d}$.

The above order-of-magnitude relations show qualitatively how a super-linear trend follows from the decrease in vertical shear, and the consequent increase in stability, brought about by an increase in bulk metallicity. However, these simple relations are not enough to quantify the super-linear trend because Δz appears to have cancelled out of both $\partial v_\phi/\partial z$ and the Brunt frequency. This difficulty is avoided in a more formal derivation of the relation between μ_0 and $\Sigma_{\rm d}/\Sigma_{\rm g}$, made under the assumption of constant *Ri*, as described in Appendix A.1.

We note further that μ_0 scales with the inverse of the radial pressure gradient parameter $v_{\rm max}/c_{\rm s}$ (equivalently η) in the same super-linear way as for $\Sigma_{\rm d}/\Sigma_{\rm g}$. The smaller is $v_{\rm max}/c_{\rm s}$, the greater is the maximum μ_0 attainable; the relation between these quantities is derived under the assumption of constant *Ri* in Appendix A.1. Thus we expect our numerical results for $\max \mu_0$ (2.9, 26.4) to depend sensitively on our assumed value of $v_{\rm max}/c_{\rm s} = 0.025$. (Bai & Stone (2010c) also reported that the degree of clumping caused by the streaming instability increased strongly with decreasing $v_{\rm max}/c_{\rm s}$.)

3.5.3 Future Directions

With each iteration of our standard procedure we allowed dust particles to settle at their full terminal velocities, regardless of gas motions evinced in previous iterations. We tried to account for these gas motions in a modified procedure by reducing settling velocities at altitude where dust may have already attained marginal stability. Settling velocities were reduced by weighting functions chosen by eye. This modified procedure enabled us to extend the settling sequence by one iteration but no more. Other weighting functions might allow the sequence to be extended further. Introducing weighting functions earlier in the sequence (rather than at the end of our standard procedure, as we have done), and increasing the midplane density by a smaller increment with each iteration (less than the 30% increment that we have adopted), would allow for a more gradual evolution and possibly permit the midplane to reach still greater densities.

Such a program would be straightforward to pursue but would be subject to arbitrariness in the forms of the weighting functions. A more direct approach would be to abandon our hybrid 1D+3D scheme and upgrade the 3D code to allow for a non-zero aerodynamic stopping time $t_{\rm stop}$ for dust. Then both settling and stability could be tracked within a single 3D code. Similar codes have been written (e.g., Johansen et al. 2009; Bai & Stone 2010b), but their application has been focussed on the streaming instability, on particles having $\Omega_{\rm K} t_{\rm stop} \gtrsim 0.1$ and (model-dependent) sizes upwards of decimeters. By contrast, we are interested in the possibility that even the smallest particles, for which $0 < \Omega_{\rm K} t_{\rm stop} \ll 1$, undergo gravitational instability. The problem of settling small particles may be coupled to the problem of settling big ones. Even if marginally coupled particles comprise only a minority of the disk's solid mass, the turbulence they induce by the streaming instability may prevent smaller particles from settling into the thin layers required for gravitational instability (Bai & Stone 2010a). Quantifying what is meant by "minority" remains a forefront issue. An efficient scheme for numerically simulating this problem would combine the anelastic methods we have adopted (so that the code timestep is not limited by the sound-crossing time) with an implicit particle integrator like the kind devised by Bai & Stone (2010b) (so that the code timestep is not limited by $t_{\rm stop}$).

Adding self-gravity would be another improvement. As noted at the beginning of §3.5, vertical self-gravity is expected to increase the maximum dust-to-gas ratio by of order 10% for the case of bulk solar metallicity. For the case of a few × supersolar metallicity, the magnitude of the correction is uncertain. It is probably at least of order unity, but might be much more, given the appearance of an infinite density cusp in those solutions of Sekiya (1998) and Youdin & Shu (2002) that account for vertical self-gravity. At the same time, self-gravity might actually limit maximum dust-to-gas ratios if the fluid becomes gravito-turbulent without producing collapsed objects (Gammie 2001).

3.5.4 Connection to Observations and The Need For Supersolar Metallicity

Observations have unveiled several trends between stellar properties and the likelihood of planet occurrence. Among the most well-known is the positive correlation between the occurrence rate of giant planets and the host star metallicity [Fe/H] (Gonzalez 1997; Santos et al. 2004; Fischer & Valenti 2005). Johnson et al. (2010) provided a comprehensive analysis, using a sample of 1266 local stars to ask whether the trend with metallicity persists across all stellar masses M_*. The answer is contained in their Figure 2. The need for supersolar metallicity is clear for M dwarfs ($0.2 < M_*/M_\odot < 0.7$), where the average metallicity of the planet-hosting stars is [Fe/H] = 0.4. Metal-rich stars presumably once carried metal-rich disks, and so the planet-metallicity correlation for M dwarfs supports our results, and those of others (Sekiya 1998; Youdin & Shu 2002; Lee et al. 2010; see also Johansen et al. 2009; Bai & Stone 2010a) that planetesimals form much more readily in metal-rich environments. In particular the data for M dwarfs indicate that a mere factor of $10^{0.4} = 2.5$ increase in metallicity above solar substantially increases the probability of planet occurrence. This is consistent with our finding of a super-linear trend between maximum dust-to-gas ratio and bulk metallicity (§3.5.2 and Appendix A.1).

However, the planet-metallicity correlation weakens systematically with increasing stellar mass (Johnson et al. 2010). For A stars ($1.4 < M_*/M_\odot < 2.0$), the correlation is arguably not present. This calls into question the need for supersolar metallicities to form planetesimals. The observations of Johnson et al. (2010) might still be reconciled with gravitational instability if more massive stars host more massive disks, although disk mass would have to scale with stellar mass in a faster than linear way to lower the threshold Toomre density (equation 3.10). The possibility also remains that the observations are not actually a direct or sensitive probe of the theory. The observations concern stellar metallicity, which might at best correlate with the global metallicity of the disk, integrated over both disk height and disk radius. By comparison, theory concerns the local metallicity $\Sigma_\mathrm{d}/\Sigma_\mathrm{g}$, integrated over height but not radius. This local metallicity (not to be confused with the local dust-to-gas ratio μ) can evolve substantially from its global value, as a consequence of radial particle drifts and photoevaporation (e.g., CY10).

Rather than look to their parent stars for evidence for local disk enrichment, we can look to the planets themselves. If planetesimals can only form in metal-enriched environments, we expect that the resultant planets will also be metal-enriched. Guillot et al. (2006) computed the bulk metallicities of the first nine extrasolar planets discovered to be transiting, all of which are hot Jupiters. The results are listed in Table 3.1, together with the modeled bulk metallicities of Jupiter and Saturn. All eleven are indeed metal-enriched, by factors ranging from 2–47 relative to the Sun, and 2–20 relative to their host stars. One caveat behind these results is that models of hot Jupiter interiors are subject to the uncertainty over the extra source of internal heat responsible for their unexpectedly large radii (see, e.g., Batygin & Stevenson 2010, who also describe a promising solution). To inflate planetary radii, Guillot et al. (2006) included in each hot Jupiter model an additional source of power equal to 0.5%

Table 3.1: Metallicities of Extrasolar Planets (Guillot et al. 2006) and Solar System Gas Giants (Guillot 2005).

Name	M_{planet} ($M_\oplus$)	$M_{\mathrm{Z}}{}^{a}$ ($M_\oplus$)	Z_{planet} ($M_{\mathrm{Z}}/M_{\mathrm{planet}}$)	$Z_{\mathrm{planet}}/Z_\odot{}^{b}$	$[\mathrm{Fe/H}]_*$	Z_{planet}/Z_*
HD209458	210	20	0.095	6.35	0.02	6.06
OGLE-TR-56	394	120	0.304	20.3	0.25	11.418
OGLE-TR-113	429	70	0.163	10.9	0.15	7.7
OGLE-TR-132	350	105	0.3	20	0.37	8.531
OGLE-TR-111	168	50	0.297	19.84	0.19	12.81
OGLE-TR-10	200	10	0.05	3.33	0.28	1.75
TrES-1	238	50	0.21	14.0	0.06	12.2
HD149026	114	80	0.70	46.78	0.36	20.42
HD189733	365	30	0.082	5.479	-0.03	5.87
Jupiter	318	10–42	0.03–0.13	2.0–8.8	0	2.0–8.8
Saturn	95.2	15–30	0.16–0.32	11–21	0	11–21

[a] The metal content for each listed extrasolar planet was derived from a model of a planetary interior that includes an additional energy source at the planet's center whose power equals 0.5% of the incident stellar luminosity.

[b] The solar metallicity $Z_\odot$ is taken to be 0.015 (Lodders 2003).

of the received stellar irradiation (Guillot & Showman 2002). The bulk metallicities inferred from the models depend on the details of this extra energy source.

Acknowledgments

We thank Xue-Ning Bai, John Johnson, Eve Ostriker, Jim Stone, and Neal Turner for discussions, and Tristan Guillot for the data in Table 3.1. Xue-Ning Bai, Anders Johansen, Jim Stone, and Andrew Youdin provided valuable feedback on a draft version of this chapter. This research was supported by the National Science Foundation, in part through TeraGrid resources provided by Purdue University under grant number TG-AST090079. A.T.L. acknowledges support from an NSF Graduate Fellowship.

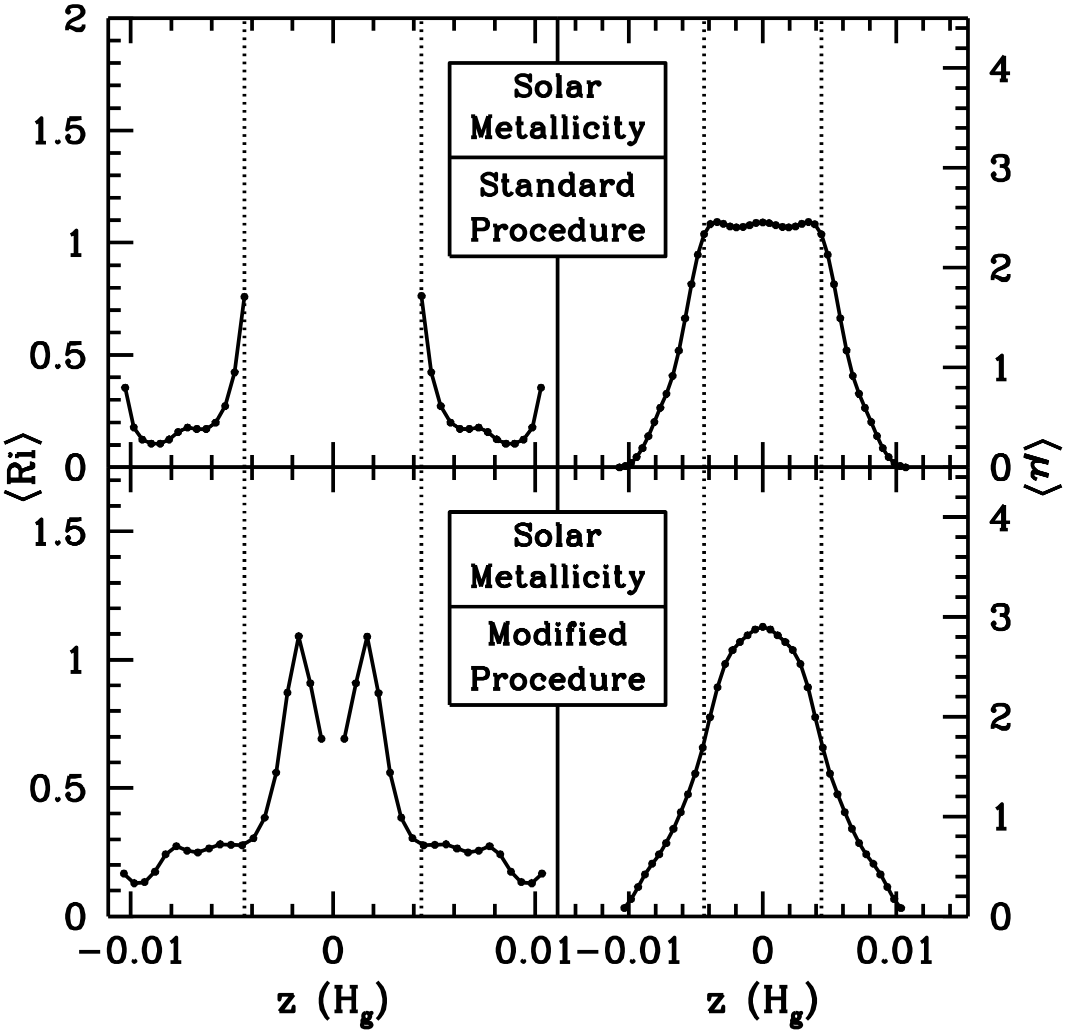

Figure 3.12: Richardson numbers Ri from the marginally stable profile of our standard procedure (top left) and from the marginally stable profile of our modified procedure (bottom left), for the case of solar metallicity. Vertical dotted lines separate the "core" from the "edges" in the standard profile (top right); these dotted lines are extended into the bottom panels for reference. We plot Ri everywhere except where density gradients are too small to compute reliably; thus we avoid the entire core region of the standard profile, and the midplane of the modified profile. Over most of the edges—those layers outside the dotted lines which have directly experienced instability, almost certainly related to the KHI—the Richardson number varies between ∼0.1–0.3. Thus, the traditional assumption that dusty sublayers relax to states of spatially constant Ri receives some empirical support from this figure.

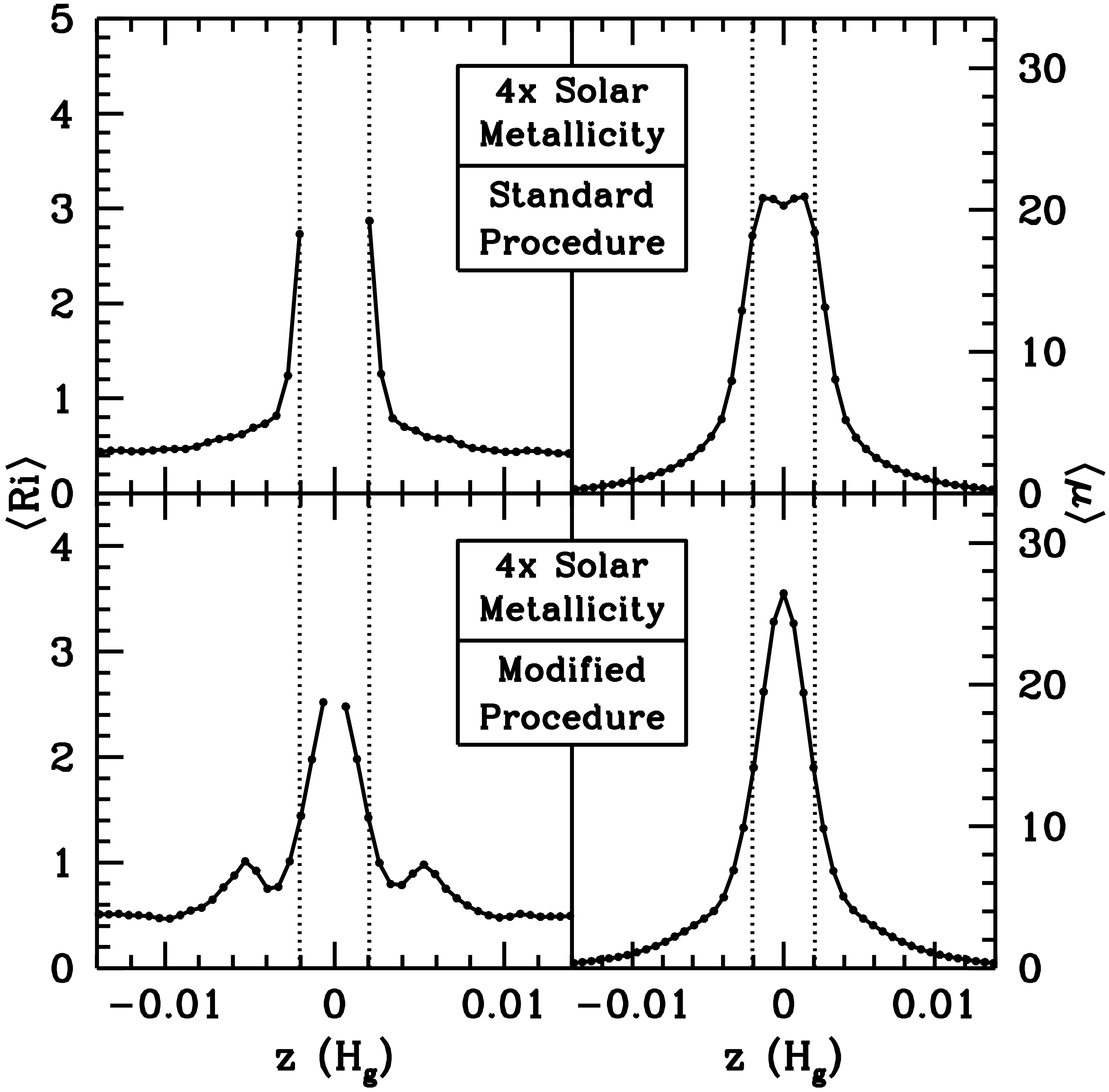

Figure 3.13: Same as Figure 3.12 except for the case of 4× bulk solar metallicity. Here the evidence that layers relax to states of spatially constant Ri is even stronger than for the case of solar metallicity. Moreover, the value to which Ri tends in this metal-rich case is higher than for the solar metallicity case: 0.4–0.6 here, versus 0.1–0.3 in Figure 3.12. This trend of increasing Ri with increasing bulk metallicity Σ_d/Σ_g is the same trend independently identified in Paper I (see Figure 6 of that paper). In the bottom panels showing the marginally stable profile identified using our modified procedure, the bumps near $z \approx \pm 0.005 H_g$ are probably artificial, a reflection of our imposed weighting function $W(z)$.

Chapter 4

Bondi-Hoyle Accretion in an Isothermal Magnetized Plasma

[1] In regions of star formation, protostars and newborn stars will accrete mass from their natal clouds. These clouds are threaded by magnetic fields with a strength characterized by the plasma β—the ratio of thermal and magnetic pressures. Observations show that molecular clouds have $\beta \lesssim 1$, so magnetic fields have the potential to play a significant role in the accretion process. We have carried out a numerical study of the effect of large-scale magnetic fields on the rate of accretion onto a uniformly moving point particle from a uniform, non-self-gravitating, isothermal gas. We consider gas moving with sonic Mach numbers of up $\mathcal{M} \approx 45$, magnetic fields that are either parallel, perpendicular, or oriented 45° to the flow, and β as low as 0.01. Our simulations utilize adaptive mesh refinement in order to obtain high spatial resolution where it is needed; this also allows the boundaries to be far from the accreting object to avoid unphysical effects arising from boundary conditions. Additionally, we show our results are independent of our exact prescription for accreting mass in the sink particle. We give simple expressions for the steady-state accretion rate as a function of β and $\mathcal{M}$ for the parallel and perpendicular orientations. Using typical molecular cloud values of $\mathcal{M} \sim 5$ and $\beta \sim 0.04$ from the literature, our fits suggest a $0.4\ M_\odot$ star accretes $\sim 4 \times 10^{-9}\ M_\odot$/year, almost a factor of two less than accretion rates predicted by hydrodynamic models. This disparity can grow to orders of magnitude for stronger fields and lower Mach numbers. We also discuss the applicability of these accretion rates versus accretion rates expected from gravitational collapse, and under what conditions a steady state is possible. The reduction in the accretion rate in a magnetized medium leads to an increase in the time required to form stars in competitive accretion models, making such models less efficient than predicted by Bondi-Hoyle rates. Our results should find application in numerical codes, enabling accurate subgrid models of sink particles accreting from magnetized media.

---

[1] Large portions of this chapter have been previously published as Lee, A. T., Cunningham, A., McKee, C., Klein, R. 2014. Bondi-Hoyle Accretion in an Isothermal Magnetized Plasma. ApJ, 783, 50.

4.1 Introduction

Accretion is ubiquitous in astrophysics. With examples including protostellar accretion from molecular clouds, mass transfer between binary companions, and gas falling onto a supermassive black hole in the center of galactic nuclei, understanding how (or whether) a gravitating source gathers mass has received much attention over the past century. In the case of star formation, considerable study has been given to understanding the process of accretion from a background medium. Knowing how much mass a star can accrete from its natal cloud will help elucidate, for example, whether the final mass of the star is determined primarily through gravitational collapse (e.g., Shu 1977) or through post-collapse accretion (e.g., Bonnell et al. 1997, 2001). Mass accretion also could play a role in the dynamics of stars in clusters. If the accretor is moving relative to the background gas, then the accretion of mass and momentum will be non-spherical, and this may play a role in the radial redistribution of objects in stellar clusters (Lee & Stahler 2011).

Several physical processes exist for transferring mass from the cloud to the surface of a (proto)star. In core-collapse models (Shu 1977), a dense core's self-gravity induces global gravitational collapse, resulting in supersonic infall either directly onto the stellar surface or into a surrounding centrifugally supported disk. Material that ultimately ends up on the star comes from a local gravitationally bound region of the parent molecular cloud. If the core is not collapsing directly onto the star+disk, or if the core is exhausted and the star is moving through the more tenuous regions of the cloud, another accretion mechanism is at play. Here the local gas initially unbound to the star can be captured and subsequently accreted. The self-gravity of this local gas is negligible relative to the gravity of the star itself. Such accretion is often called Bondi accretion when the star is stationary or Bondi-Hoyle(-Lyttleton) accretion when the star is moving relative to the background gas, named after the pioneering investigators (Hoyle & Lyttleton 1939; Bondi & Hoyle 1944; Bondi 1952).

The primary goal of this work is concerned with understanding the steady-state mass accretion rate for Bondi-Hoyle accretion when the background gas is an isothermal plasma pervaded by a magnetic field. In particular, we seek to construct an interpolation formula that reproduces both known analytic and numerical results as well as the steady-state accretion rates we will obtain via numerical simulations. In our work and these previous works, the effects of stellar winds and outflows are neglected. We begin this study by summarizing some of the known results in the next section. From there, we propose new interpolation formulas for the mass accretion rate of magnetized Bondi-Hoyle flow. This function will have two free parameters, which we fix by fitting to numerical simulations. Section 4.3 discusses our methodology and numerical convergence studies of the numerical code. Section 4.4 presents the numerical results and the results of fitting our proposed interpolation formulas to the simulation data. In Section 4.5 we discuss the applicability of such steady-state models in regions of star formation. Section 4.6 concludes this work with a summary and discussion. How these models can be implemented in sub-grid and sink particle algorithms is discussed in Appendix B.

4.2 Mass Accretion Rates

4.2.1 Known Results

The study of steady-state accretion from an initially uniform background medium has enjoyed many analytical and numerical studies. Edgar (2004) gives a nice pedagogical review of some of the earlier work. Hoyle & Lyttleton (1939) first solved the problem for a point particle of mass M_* moving through a collisionless fluid at (hypersonic) speed v_0. Matter was focused into a vanishingly thin wake and accreted through a spindle downstream of the accretor. The accretion rate was

$$\dot{M}_{\rm HL} = 4\pi r_{\rm HL}^2 \rho_0 v_0 = \frac{4\pi G^2 M_*^2 \rho_0}{v_0^3} \ , \tag{4.1}$$

for the far-field mass density ρ_0. Associated with v_0 is the characteristic radius

$$r_{\rm HL} \equiv \frac{GM_*}{v_0^2} \ , \tag{4.2}$$

which measures the dynamic length scale within which gravity wins over the inertia of the gas. In the opposite limit of stationary or subsonic motion, the thermal pressure exceeds the ram pressure of the gas by a factor of $\sim (c_{\rm s}/v_0)^2$ for sound speed $c_{\rm s}$. Bondi (1952) analytically solved the problem for a stationary accretor, arriving at

$$\begin{aligned} \dot{M}_{\rm B} &= 4\pi\lambda r_{\rm B}^2 \rho_0 c_{\rm s} = \frac{4\pi\lambda G^2 M_*^2 \rho_0}{c_{\rm s}^3} & (4.3) \\ &= 1.02 \times 10^{-6} \left(\frac{M_*}{0.4\ M_\odot}\right)^2 \left(\frac{n_0}{10^4\ {\rm cm}^{-3}}\right) \left(\frac{T}{10\ {\rm K}}\right)^{-3/2} \ M_\odot\ {\rm yr}^{-1} \ , & (4.4) \end{aligned}$$

where we have defined the Bondi radius

$$r_{\rm B} \equiv \frac{GM_*}{c_{\rm s}^2} = 9.0 \times 10^{16} \left(\frac{M_*}{0.4\ M_\odot}\right) \left(\frac{T}{10\ {\rm K}}\right)^{-1} \ {\rm cm} \ . \tag{4.5}$$

In our numerical evaluations, we have normalized the temperature T to 10 K, the number density n_0 to 10^4 particles per cm^3, and masses to the solar mass $M_\odot$. The mass density is related to the number density by $\rho_0 = (2.34 \times 10^{-24}\ \text{grams}) \cdot n_0$. The symbol λ is a function of the adiabatic index γ ($\lambda = \exp(3/2)/4 \approx 1.12$ for an isothermal gas, $\gamma = 1$).

Both limits then established, Bondi proposed his venerable Bondi-Hoyle interpolation formula that connects the stationary and hypersonic regimes:

$$\dot{M}_{\rm BH} = \frac{4\pi\rho_0 r_{\rm B}^2 c_{\rm s}}{(1+\mathcal{M}^2)^{3/2}} = \frac{\dot{M}_{\rm B}/\lambda}{(1+\mathcal{M}^2)^{3/2}} \ , \tag{4.6}$$

where we have introduced the sonic Mach number $\mathcal{M} \equiv v_0/c_{\rm s}$. The characteristic velocity for Bondi-Hoyle accretion is

$$v_{\rm BH} = (c_{\rm s}^2 + v_0^2)^{1/2} \ , \tag{4.7}$$

and the corresponding Bondi-Hoyle radius is

$$r_{\rm BH} = \frac{GM_*}{v_{\rm BH}^2} = \frac{r_{\rm B}}{1+\mathcal{M}^2} \ . \tag{4.8}$$

We will see in Section 4.4 that magnetic fields reduce the accretion rate below these values. Furthermore, for the fiducial values of n_0 and T and for $M_* > 0.4M_\odot$, Bondi accretion is not in a steady state (Section 4.5).

We shall express all accretion rates in terms of the Bondi accretion rate in two equivalent forms. For example, the Bondi-Hoyle accretion rate will be written as

$$\dot{M}_{\rm BH} = \phi_{\rm BH} \cdot 4\pi\lambda r_{\rm BH}^2 \rho_0 v_{\rm BH} = \phi_{\rm BH} \left(\frac{c_{\rm s}}{v_{\rm BH}}\right)^3 \dot{M}_{\rm B} \ . \tag{4.9}$$

This first form emphasizes the underlying physical parameters, and we have introduced a correction factor $\phi_{\rm BH} = \phi_{\rm BH}(\mathcal{M})$, which will be of order unity. The second form is

$$\dot{M}_{\rm BH} = \left(\frac{c_{\rm s}}{v_{\rm BH,eff}}\right)^3 \dot{M}_{\rm B} \ . \tag{4.10}$$

Here, the effective Bondi-Hoyle velocity $v_{\rm BH,eff}$ is an interpolation formula; the rationale for introducing the second form will become clear below.

Simulations have shown that Bondi's interpolation formula ($\phi_{\rm BH} = 1/\lambda$) can be in error by several ten's of percent (Shima et al. 1985; Ruffert 1994). These authors, among others, have considered the non-isothermal case as well and have proposed two-dimensional interpolation formulas (in $\mathcal{M}$ and γ) to match simulation results. Typically such formulas are monotonically decreasing functions of both $\mathcal{M}$ and γ and agree well the simulations. A complication is that Ruffert (1994, 1996) has shown that accretion rates do not decrease monotonically as $\mathcal{M}$ increases, but instead increase near $\mathcal{M} \sim 1$ and then asymptote to $\dot{M}_{\rm HL}$. For the isothermal case, we have found that

$$\phi_{\rm BH} = \frac{(1+\mathcal{M}^2)^{3/2}[1+(\mathcal{M}/\lambda)^2]^{1/2}}{1+\mathcal{M}^4} \ , \tag{4.11}$$

corresponding to

$$\mathcal{M}_{\rm BH} \equiv \frac{v_{\rm BH,eff}}{c_{\rm s}} = \frac{(1+\mathcal{M}^4)^{1/3}}{[1+(\mathcal{M}/\lambda)^2]^{1/6}} \ , \tag{4.12}$$

and agrees with the numerical results of Ruffert (1996, for $\gamma = 1.01$) and those reported below with a maximum error of 27%. Observe that $\phi_{\rm BH} \to 1$ as $\mathcal{M} \to 0$ and $\phi_{\rm BH} \to 1/\lambda$ for $\mathcal{M} \gg 1$. This function is plotted in Figure 4.1.

Other numerical studies of accretion have studied the role of additional physics like radiation pressure (Milosavljević et al. 2009), turbulence (Krumholz et al. 2006), turbulence and magnetic fields in spherically symmetric accretion (Shcherbakov 2008), the presence of a disk (Moeckel & Throop 2009), or thermal instabilities (Gaspari et al. 2013). Our finding is that Equation (4.9) with $\phi_{\rm BH}$ given by Equation (4.11) is a reasonable measure of the accretion rate for isothermal Bondi-Hoyle accretion when additional physics do not play an appreciable role in the dynamics of the gas near the accretor.

One physical effect that could play an important role in the gas dynamics is a global magnetic field. In star-forming regions, there is ample evidence that molecular clouds are pervaded by magnetic fields (Crutcher 1999; McKee & Ostriker 2007), whose strength (i.e., its ability to influence dynamics) can be characterized by the plasma β, the ratio of the thermal pressure to the magnetic pressure:

$$\beta \equiv \frac{\rho c_{\rm s}^2}{B^2/8\pi} = 2\left(\frac{c_{\rm s}}{v_{\rm A}}\right)^2 = 2\left(\frac{\mathcal{M}_{\rm A}}{\mathcal{M}}\right)^2 , \tag{4.13}$$

for magnetic field amplitude B. We have also introduced the Alfvén Mach number $\mathcal{M}_{\rm A} = v_0/v_{\rm A}$, the ratio of the gas velocity to the Alfvén velocity, $v_{\rm A} = B/\sqrt{4\pi\rho}$. Observations of the Zeeman effect, linear polarization emission of dust, and the Chandrasekhar-Fermi method (see the review of Crutcher 2012) have suggested molecular clouds have β values of most order unity, but more have $\beta < 0.1$ (e.g., Crutcher 1999).

Cunningham et al. (2012) have studied accretion from a magnetized, isothermal, static medium. For the case in which thermal pressure is negligible (the low-β limit) they argued that gas would collapse along the field lines from a distance $\sim r_{\rm B}$ above and below the point mass, and would then accrete from $\sim$ an Alfvén radius,

$$r_{\rm A} \equiv \frac{GM_*}{v_{\rm A}^2} = \frac{c_{\rm s}^2}{v_{\rm A}^2}\, r_{\rm B} \tag{4.14}$$

at velocity $\sim v_{\rm A}$. In our notation,

$$\dot{M}_{\rm A} = \phi_{\rm A} \cdot 4\pi\lambda r_{\rm B} r_{\rm A} \rho_0 v_{\rm A} \ . \tag{4.15}$$

As a result, the accretion rate varies as $v_{\rm A}^{-1} \propto \beta^{1/2}$,

$$\dot{M}_{\rm A} = \phi_{\rm A}\left(\frac{c_{\rm s}}{v_{\rm A}}\right)\dot{M}_{\rm B} = \phi_{\rm A}\left(\frac{\beta}{2}\right)^{1/2}\dot{M}_{\rm B} \ . \tag{4.16}$$

Cunningham et al. (2012) expressed the accretion rate as

$$\dot{M}_{\rm A} = \left(\frac{\beta}{\beta_{\rm ch}}\right)^{1/2}\dot{M}_{\rm B} \ , \tag{4.17}$$

where $\beta_{\rm ch}$ is a numerical factor; this corresponds to $\phi_{\rm A} = (2/\beta_{\rm ch})^{1/2}$. They estimated $\beta_{\rm ch} \approx 5$, so that $\phi_{\rm A} = 0.63$. From here, they generalized this to include a finite temperature (the "Alfvén-Bondi" case). By writing $v_{\rm A}$, $r_{\rm A} \rightarrow v_{\rm AB}$, $r_{\rm AB}$ in Equation (4.15), the accretion rate becomes

$$\dot{M}_{\rm AB} = \phi_{\rm AB} \cdot 4\pi\lambda r_{\rm B} r_{\rm AB} \rho_0 v_{\rm AB} = \phi_{\rm AB}\left(\frac{c_{\rm s}}{v_{\rm AB}}\right)\dot{M}_{\rm B} \ , \tag{4.18}$$

$$= \left(\frac{c_{\rm s}}{v_{\rm AB,eff}}\right)\dot{M}_{\rm B} \ , \tag{4.19}$$

where $v_{\rm AB} \equiv (c_{\rm s}^2 + v_{\rm A}^2)^{1/2}$ and $r_{\rm AB} = GM_*/v_{\rm AB}^2$.[2] The effective Alfvén-Bondi velocity, $v_{\rm AB,eff}$, can be chosen to provide an interpolation formula between the Alfvén and Bondi cases that agrees best with the numerical simulations; Cunningham et al. (2012) adopted

$$v_{\rm AB,eff} \equiv \left[c_{\rm s}^n + \left(\frac{\beta_{\rm ch}}{2}\right)^{n/2} v_{\rm A}^n\right]^{1/n} = \left[1 + \left(\frac{\beta_{\rm ch}}{\beta}\right)^{n/2}\right]^{1/n} c_{\rm s} \ , \tag{4.20}$$

which gives

$$\dot{M}_{\rm AB} = \left[1 + \left(\frac{\beta_{\rm ch}}{\beta}\right)^{n/2}\right]^{-1/n} \dot{M}_{\rm B} \ . \tag{4.21}$$

They found that $n = 0.42$ and $\beta_{\rm ch} = 5.0$ gave agreement with their numerical results to within 5% for $\beta \geq 0.01$.

4.2.2 Alfvén-Bondi-Hoyle Accretion

We wish to extend the work of Cunningham et al. (2012) to the case in which the accreting mass is moving through a magnetized ambient medium. Our primary interest is in star-forming regions, which are approximately isothermal because the dust and the molecules can efficiently radiate the energy supplied by compression; we therefore assume that the gas is isothermal.[3] The magnetic flux in stars is orders of magnitude less than that in the gas from which they form, so most of of the magnetic flux in the accreting gas decouples from the gas and accumulates in the vicinity of the protostar (Zhao et al. 2011). As a result, even in cases where the thermal pressure ($\sim \rho c_{\rm s}^2$) or ram pressure ($\sim \ \rho v^2$) initially control the dynamics of the gas near the accretor, accretion can redistribute magnetic flux so that the magnetic pressure ($\propto B^2$) eventually dominates the dynamics near the accreting object. In steady-state Bondi accretion from a magnetized gas, Cunningham et al. (2012) found that even if β was initially > 1, a steady-state was reached when the gas within $\sim r_{\rm AB}$ of the accretor had $\beta \approx 1$. In a steady-state flow where there is relative motion between the gas and the accretor (i.e., magnetized Bondi-Hoyle accretion), we anticipate that if $\mathcal{M}_{\rm A} \ll \min(1, \mathcal{M})$, the inertia of the gas will play a small role in setting the steady-state accretion rate, so $\dot{M}$ will be well approximated by the Alfvén-Bondi result (Eq. 4.18). If instead $\mathcal{M}_{\rm A} \gg 1$, the inertia of the gas is able to drag away most of the magnetic flux so that the the magnetic field is not dominant anywhere and the accretion rate should approach the non-magnetized Bondi-Hoyle limit. We wish to develop an approximate analytic expression for the rate of accretion by a point mass moving at a constant speed through a uniform, isothermal, magnetized medium by further generalizing the above known results. Our expression will

---

[2] Note that Equation (16) of Cunningham et al. (2012) has a typo, one factor of $r_{\rm AB}$ should be written as $r_{\rm B}$ instead.

[3] As we discuss in Section 4.6, our results should also be applicable to the central regions of active galactic nuclei.

therefore also include parameters n and $\beta_{\rm ch}$, which we can then adjust to best reproduce the results of our simulations of magnetized Bondi-Hoyle flow.

We generalize Equation (4.18) by replacing $r_{\rm B}$ with $r_{\rm BH}$, $r_{\rm AB}$ with $r_{\rm ABH}$, and $v_{\rm AB}$ with

$$v_{\rm ABH} \equiv (c_{\rm s}^2 + v_0^2 + v_{\rm A}^2)^{1/2} \ . \tag{4.22}$$

Here, and throughout the remainder of the chapter, c_s is the isothermal sound speed. With

$$r_{\rm ABH} \equiv \frac{GM_*}{v_{\rm ABH}^2} \ , \tag{4.23}$$

the accretion rate is then

$$\dot{M}_{\rm ABH} = \phi_{\rm ABH} \cdot 4\pi\lambda r_{\rm BH} r_{\rm ABH} \rho_0 v_{\rm ABH} = \phi_{\rm ABH} \left(\frac{c_{\rm s}^3}{v_{\rm BH}^2 v_{\rm ABH}} \right) \dot{M}_{\rm B} \tag{4.24}$$

$$= \left(\frac{c_{\rm s}^3}{v_{\rm BH,eff}^2 v_{\rm ABH,eff}} \right) \dot{M}_{\rm B}. \tag{4.25}$$

Equations (4.24–4.25) do not take into account the orientation of the flow relative to the ambient magnetic field. In Section 4.4.2 below, we find that we need different interpolation formulas for the cases where the flow is parallel and perpendicular to the magnetic field. For the parallel case, we generalize $v_{\rm AB,eff}$ to $v_{{\rm ABH},\parallel,{\rm eff}}$ with

$$\frac{v_{{\rm ABH},\parallel,{\rm eff}}}{c_{\rm s}} = \left[\left(\frac{v_{\rm BH,eff}}{c_{\rm s}} \right)^n + \left(\frac{\beta_{\rm ch}}{\beta} \right)^{n/2} \right]^{1/n} \ . \tag{4.26}$$

The accretion rate in this case is

$$\dot{M}_\parallel = \frac{1}{\mathcal{M}_{\rm BH}^2} \left\{ \mathcal{M}_{\rm BH}^n + \left(\frac{\beta_{\rm ch}}{\beta} \right)^{n/2} \right\}^{-1/n} \dot{M}_{\rm B} \ , \tag{4.27}$$

where $\mathcal{M}_{\rm BH}$ was defined in Equation (4.12). Observe that this expression reduces to Bondi accretion for $\mathcal{M} = \beta^{-1} \to 0$, to Bondi-Hoyle accretion if $\beta \to \infty$, and to Alfvén-Bondi accretion for $\mathcal{M} = 0$ and arbitrary β. The factor $\phi_{{\rm ABH},\parallel}$ in this case is then

$$\phi_{{\rm ABH},\parallel} = \phi_{\rm BH}^{2/3} \cdot \left(\frac{v_{\rm ABH}}{v_{{\rm ABH},\parallel,{\rm eff}}} \right) \ ; \tag{4.28}$$

it is also plotted in Figure 4.1.

For flows perpendicular to the field, we obtain better agreement with our simulations with the less-intuitive interpolation

$$\frac{v_{{\rm ABH},\perp,{\rm eff}}}{c_{\rm s}} \equiv \max \left[\frac{v_{\rm BH,eff}}{c_{\rm s}}, \frac{v_{{\rm ABH},\parallel,{\rm eff}}}{v_{\rm BH,eff}} \right] \tag{4.29}$$

in Equation (4.25). One can readily verify that this has the correct limits for Bondi, Bondi-Hoyle, and Alfvén-Bondi accretion. We then obtain

$$\dot{M}_\perp = \frac{1}{\mathcal{M}_{\rm BH}^2} \min \left\{ \frac{1}{\mathcal{M}_{\rm BH}} \ , \ \mathcal{M}_{\rm BH} \left[\mathcal{M}_{\rm BH}^n + \left(\frac{\beta_{\rm ch}}{\beta} \right)^{n/2} \right]^{-1/n} \right\} \dot{M}_{\rm B} \ . \tag{4.30}$$

An immediate interesting result of this formulation—that is born out in our simulations, see §4.2—is for the particular case of highly supersonic flow with an Alfvén Mach number $\gtrsim 1$, the accretion rate for the perpendicular case reduces to

$$\dot{M}_\perp = \frac{\dot{M}_{\rm B}/\lambda}{\mathcal{M}^3} = \dot{M}_{\rm HL} \quad (\mathcal{M} \gg 1, \ \mathcal{M}_{\rm A} \geq 1) \ , \tag{4.31}$$

even when $\mathcal{M}_{\rm A} \approx 1$.

If the point mass is moving through a medium at an angle θ with respect to the magnetic field, we approximate the accretion rate by

$$\dot{M} \simeq \dot{M}_{\|} \cos^2 \theta + \dot{M}_\perp \sin^2 \theta \ . \tag{4.32}$$

Indeed, we confirm for one of our simulations that when $\theta = 45°$, the resulting accretion rate is decently approximated by the average of the two limiting rates. If the orientation changes randomly in time, the proposed average accretion rate is

$$\dot{M} \simeq \frac{1}{2} \left(\dot{M}_{\|} + \dot{M}_\perp \right) \ . \tag{4.33}$$

In order to test our proposed interpolation formulas, we study the problem of Bondi-Hoyle accretion in a magnetized plasma using the `RAMSES` MHD code (Teyssier 2002) over a range of field strengths and sonic Mach numbers relevant for star formation. These simulations employ the adaptive mesh refinement (AMR) capabilities of the code to retain high spatial resolution where it is needed—close to the accreting object—while allowing for a large computational domain to prevent the boundaries of the domain from influencing the steady-state flow. As noted above, we do not consider the effects of stellar winds or outflows on the accretion rate. The numerical methodology is described in the next section.

4.3 Numerical Methods

Our methods are similar to those described in Cunningham et al. (2012). Here we summarize the methods, highlighting the significant differences in this work, present the results of our convergence study, and point to where the reader can find additional details if interested.

We solve the equations of ideal MHD for an isothermal gas with a fixed point mass at the origin:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{v} = -S_M(\mathbf{x}) \ , \tag{4.34}$$

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = -\nabla \left(P_{\rm th} + \frac{\mathbf{B}^2}{8\pi} \right) + \frac{\mathbf{B} \cdot \nabla \mathbf{B}}{4\pi} - \frac{G M_* \rho}{\mathbf{x}^2}\hat{\mathbf{x}} - S_{\rm M}(\mathbf{x}) \cdot \mathbf{v} \ , \tag{4.35}$$

$$\frac{\partial B}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) = 0 \ , \tag{4.36}$$

$$P_{\rm th} = \rho c_{\rm s}^2 \ . \tag{4.37}$$

Here $\mathbf{v}$ is the velocity of the gas, $\mathbf{x}$ is the position relative to the sink particle, and $\mathbf{B}$ is the magnetic field. Self-gravity of the gas is neglected. In the code, the point mass is represented by a fixed sink particle located at the center of the computational domain. The term S_M allows for mass accretion onto the central point mass if gas flows into a sphere of radius $4\Delta x$, where Δx is equal to the size of the grid cell on the finest AMR level. The accreted gas's momentum is also removed from the grid, though the particle is held stationary at the center of the domain.[4] The sink particle is allowed to accrete mass but not magnetic flux, and it accretes as much mass in a timestep Δt as it can without introducing a new local maximum in the Alfvén speed amongst the cells located within a shell with radius r between $4\Delta x$ and $6\Delta x$ from the accreting particle. That is,

$$S_{\rm M}(\mathbf{x}) = \begin{cases} \dfrac{1}{\Delta t} \max \left(0 \ , \rho - \dfrac{B}{4\pi v_{\rm A,max}^2} \right) & \text{if } |\mathbf{x}| < 4\Delta x \\ 0 & \text{otherwise} \end{cases} \ , \tag{4.38}$$

where $v_{\rm A,max} = \max(v_{\rm A}(\mathbf{x}); 4 \leq |\mathbf{x}|/\Delta x \leq 6)$. The reader can also see the paragraph containing Equation (7) of Cunningham et al. (2012) for more details on the sink particle algorithm.

Since the sink particle accretes mass but not flux, the cells interior to the sink particle radius decouples the gas from the field. In reality, non-ideal MHD effects remove the majority of the accreted gas from the field within the accretion disk < 100 AU from the star; see equation (48) of (Li & McKee 1996) or the review of Armitage (2011). Our sink particles will typically have a radius of ~ 500 AU, so our treatment of non-ideal MHD effects requires a sub-grid model; our prescription was given above.[5] Furthermore, this also means that gas just interior to the sink particle radius could be artificially affected by non-ideal effects. Nonetheless, both the exact prescription for how gas is removed from the field lines and the size of the sink particle are unimportant as long as the gas entering the sink region

---

[4] We note that the absence of the $-S_{\rm M}(\mathbf{x}) \cdot \mathbf{v}$ term in the equations of Cunningham et al. (2012) is a typographical mistake.

[5] Non-ideal effects can also play a role at larger distance ($\lesssim 1000$ AU) within shocks that originate from the collision of in-falling gas and the magnetic field that has been freed from accreted material (Li & McKee 1996).

has accelerated to free-fall. If this has occurred, the in-falling gas has causally disconnected from the surrounding medium and any artificial prescriptions cannot alter the far-field gas. We discuss how are results are independent of the sink particle conditions in more detail in Section 4.2.

In addition to the very small magnetic flux the star gains by accretion, the star might also generate its own field through dynamo action. The fields of newborn stars are observed with strengths of order kGauss, but the dipole component of the field falls off as $(R_*/R)^3$, making the stellar field strength a few μGauss at $\sim$ 10 AU, which is already smaller than the field in the ISM. Therefore, we neglect the field generated by the star itself.

For all our integrations, the gas is initially uniform with density ρ_0 and sound speed $c_{\rm s}$. The magnetic field is initially set to be uniform in the $\hat{z}-$direction with a magnitude set by β. The speed of the gas is initially set to v_0, which is oriented either parallel or perpendicular to the $\mathbf{B}$-field, except in one case where we orient $\mathbf{v}$ at an angle of 45 degrees. We explore a parameter space of β and $\mathcal{M}$. We consider β values of $10^{30},\ 10^2,\ 10,\ 1,\ 0.1,$ and 0.01 and sonic Mach numbers that range from 0.014 to 44.7. For a given β, we select our velocities to be either equal to the Alfvén velocity or -1, -1/2, 1/2, or 1 decade from this value. This gives us a combination of runs that are both sub and super-sonic as well as sub and super-Alfvénic. Table 4.1 tabulates the parameter space explored, and Figure 4.2 shows this parameter space graphically. The plot identifies four regions of parameter space, depending on whether $\mathcal{M}$ and $\mathcal{M}_{\rm A}$ are greater or less than one. Including the stationary runs of Cunningham et al. (2012), our runs explore two of these regions quite well ($\mathcal{M}, \mathcal{M}_{\rm A} > 1$ and $\mathcal{M}, \mathcal{M}_{\rm A} < 1$). The empty region ($\mathcal{M} < 1$, $\mathcal{M}_{\rm A} > 1$) is explored by hydrodynamic models of Bondi-Hoyle flow (Ruffert 1996). The final region ($\mathcal{M} > 1$, $\mathcal{M}_{\rm A} < 1$) is only explored by one simulation. Typical star forming regions have $\mathcal{M}_{\rm A} \approx 1$ (Crutcher 2012), so we also explore two cases with an Alfvén Mach number of unity.

We carried out our computations using the `RAMSES` code (Teyssier 2002), an adaptive-mesh-refinement (AMR) code with an oct-tree data structure. The computational domain is a three-dimensional Cartesian domain with a length of $50\,r_{\rm B}$ in each direction. We discretize the domain onto a Cartesian base-level grid of 64^3. Denote this level as level $L = 0$. We allow for seven additional levels of refinement ($L = 1, 2, ..., 7$), with each level incrementing the grid-cell density by 2^3 above the previous level. Each grid cell in the domain is initially refined up to level L if its distance x from the center of the domain satisfies

$$\left(\frac{25}{2^{L+1}}\right) r_{\rm B} < x < \left(\frac{25}{2^{L}}\right) r_{\rm B} \ .$$

That is, initially the grid is a set of concentric spheres of increasing refinement as the radius decreases. We also allow for further adaptive refinement if a particular pair of zones has a steep density gradient: if any component of $(\Delta x/\rho)\nabla\rho$ exceeds 1/2, those cells are refined. In this evaluation, the ρ in the denominator is the average of the two cell densities. This second criterion is met only at later times when transient features develop near the sink particle.

Seven AMR levels sufficiently resolve the relevant lengths scales for the majority of our runs. Since our runs include thermal pressure, gas motion, and magnetic fields, we want to ensure not just that the length scale $r_{\rm AB}$ scale is resolved—as was done in Cunningham et al. (2012)—but that the Alfvén-Bondi-Hoyle length scale (Equation 4.23) is adequately resolved. The maximum level of refinement provides an effective resolution of $\Delta x = 50\,r_{\rm B}/(64 \cdot 2^7 \text{ cells}) = r_{\rm B}/(164 \text{ cells})$. Table 4.1 tabulates the value $r_{\rm ABH}/\Delta x$. All length scales are resolved by at least 7 cells on the finest level. We note that we are allowing one less level of refinement compared to Cunningham et al. (2012), who allowed up to $L_{\rm max} = 8$. Even though some runs have the smallest length scale resolved by ≤ 8 zones, we have confirmed through numerical convergence studies that reducing the number of AMR levels from $L_{\rm max} = 8$ to $L_{\rm max} = 7$ does not affect the steady-state accretion rate for several of our runs; see Figure 4.3, where the mass accretion rate for several examples is compared as a function of $L_{\rm max}$. In cases where increasing $L_{\rm max}$ changes the steady state accretion rate by more than 30%, we include additional levels of refinement until the disparity between simulations diminishes.

The cases $(\beta, \mathcal{M}) = (0.1, 44.7)$, (0.1,4.47) and (0.01,4.47) require special treatment. For the first, $r_{\rm ABH} \approx r_{\rm BH} \approx r_{\rm B}/2000$ is not adequately resolved by even one cell at the finest level of refinement with our standard procedure. Furthermore this implies the sink particle—having a radius of four times the finest grid cell—exceeds the smallest length scale and thus no longer approximates a point particle. Since $r_{\rm BH}$ is orders of magnitude smaller than the other two length scales—and consequently the ratio of pressures $P_{\rm ram}/P_{\rm B} \approx 200$—we expect the inertia of the gas to dominate the dynamics, so the time for the flow to reach a steady state should be of order $t_{\rm ABH} \approx t_{\rm BH} = r_{\rm BH}/v_0 = t_{\rm B}/\mathcal{M}^3 \ll t_{\rm B} = r_{\rm B}/c_{\rm s}$. Unperturbed fluid traverses only a small part of the computational domain in the time required for the flow to achieve steady-state. In order to adequately resolve the Bondi-Hoyle length scale in this case, we reduce the size of the box by a factor of 2^7, making the length of the domain $\approx 780\,r_{\rm BH}$. We also allow one additional level of refinement, making the finest level of refinement smaller by an additional factor of 2 so that $r_{\rm BH}/\Delta x \approx 21$. A steady state is reached in a few $t_{\rm BH}$, so we need not worry about boundary conditions affecting the state of the flow near the accretor.

For the cases of $(\beta, \mathcal{M}) = (0.1, 4.47)$, and $(0.01, 4.47)$, $r_{\rm ABH}/\Delta x \approx 4$ and 1, respectively, when $L_{\rm max} = 7$. A convergence study showed that, for the parallel orientation, two additional levels were required for the former case and one for the latter. Convergence is achieved in the latter case even though the ABH length, $r_{\rm ABH}$, is resolved by only about 2 zones. The former required more levels because $\mathcal{M}_{\rm A} = 1$, and we find that the accretion rates for these cases are most sensitive to the flow morphology $\sim r_{\rm ABH}$ from the accretor, and thus require the most resolution at these scales (see Figure 4.3). For the perpendicular orientations, only one additional level was required to show convergence.

All quantities are computed in the cell centers, except for the magnetic field, which is computed on the cell faces. The magnitude of a particular cell's magnetic field is then the average of the magnitude of the cell faces.

We set $\rho_0 = 10^{-8}(M_*/r_{\rm B}^3)$ in all our simulations. The total mass of the gas is then $(50\,r_{\rm B})^3\rho_0 \approx 10^{-3}M_* \ll M_*$, justifying our neglect of the gas's self gravity. Integrations are run to a final time t_{end} sufficiently long to attain a statistically steady accretion rate onto the central particle. Using a stellar mass of $M_{\odot}$ with $T = 10$ Kelvin gas, $r_{\mathrm{B}} \sim 22,000$ AU. For our default resolution with seven levels of refinement, the finest level has a resolution of ~ 135 AU, and the radius of the sink particle is 540 AU. For the Mach 44.7 run where the box size is reduced, the radius of the sink particle is ~ 2 AU. For this run only, the stellar field could influence the gas surrounding the sink particle. However, given the high momentum of the gas ($\mathcal{M}_{\mathrm{A}} = 10$), this run will mimic non-magnetic hydrodynamic flow where additional non-ideal effects play little-to-no role in setting the final accretion rate.

4.4 Results

4.4.1 Morphology

All of our subsonic runs are also sub-Alfvénic, making the gas morphologies and the final accretion rates well approximated by the stationary models of Cunningham et al. (2012). In this section we describe the supersonic cases, particularly the $\mathcal{M} = 1.41$ and $\mathcal{M} = 4.47$ runs.

Figures 4.4 and 4.5 show snapshots late in the simulations after a steady state accretion rate has been established. These two-dimensional slices through the center of the computational domain show the gas density (color bar), velocity of the gas (arrows), and magnetic flux direction (lines) for $\beta \geq 0.1$ and $\mathcal{M} = 1.41$ and 4.47. Figure 4.6 takes the $\mathcal{M} = 1.41$ and $\beta = 1$ runs and plots the local values of $\mathcal{M}$, β, $\mathcal{M}_{\mathrm{A}}$, and B^2 at the same late time.

The general evolution of the runs goes as follows. Initially, gravity pulls nearby gas towards the sink particle, pinching the magnetic field perpendicular to the far-field flow direction for the parallel orientation, and parallel to the flow for the perpendicular orientation. Gas flows relatively undisturbed until it hits the developing shock at the Mach cone or, in some cases, a bowshock propagating upstream. These shocks retard the gas to sub-magnetosonic velocities, and the gas continues to flow along field lines downstream of the shocks. Near the source, field lines are drawn towards the sink particle, creating a network of pathways for gas to flow onto the accretor. The extent which field lines can be dragged toward the source depends on the values of $\mathcal{M}$ and $\mathcal{M}_{\mathrm{A}}$; stronger fields are more resistant to bending (compare, for example, field lines downstream of the shocked region for the $\mathcal{M} = 1.41$ parallel runs in the left panel of Figure 4.4). Mass-loaded field lines that reach the sink are relieved of their gas, eliminating the gravitational force pulling holding them at the sink. Like a released bowstring, the field snaps back into the surrounding gas (this is prominently shown in the perpendicular orientation for $\mathcal{M} = 1.41$ and $\beta = 1$ in Figure 4.5; here the downstream field lines were recently released). While the mass accretion rate reaches an approximate steady state, the morphology of the flow shows larger fluctuations. The details of this morphology depend on the initial orientation of the field to the flow, which we now consider in turn.

Parallel Orientations

In the parallel case, there are two types of shock: hydrodynamic, in which $\mathbf{B}$ is unaffected, and switch-on shocks, in which a perpendicular component of the field appears behind the shock front (Draine & McKee 1993). The conditions for the occurrence of a switch-on shock are (1) $\mathcal{M}_{\rm A} > 1$: (2) $v_{\rm A} > c_s$, corresponding to $\beta < 2$; and (3) the post-shock flow must be less than $v_{\rm A} \cos\theta_2$, where the subscript 2 denotes post-shock quantities and θ is the angle between the magnetic field and the flow velocity. The first two requirements ensure that the shock velocity exceeds the fast-wave velocity v_F, which is $\max(v_{\rm A},\ c_{\rm s}) = v_{\rm A}$ in this case. The third, post-shock requirement translates to an upper bound on the pre-shock velocity. For isothermal gas ($\gamma = 1$), as in our simulations, this upper bound is infinite (Draine & McKee 1993).

Gravity amplifies ρ and B relative to the background values ρ_0 and B_0, but primarily inside $r_{\rm ABH}$. Elsewhere, shocks can also produce density and/or field enhancements. For the parallel orientations, a Mach cone develops immediately for all the supersonic runs and typically extends far beyond $r_{\rm ABH}$. This is the enhanced density region surrounding and downstream from the accretor in Figure 4.4. At the later times shown in this Figure, the Mach cone may have joined onto shocks propagating upstream of the accretor, which also may end up disturbing the Mach cone's shape (e.g., the $\mathcal{M} = 1.41$ runs in Figure 4.4). In the case of $\mathcal{M} = 4.47$ and $\beta = 0.1$, the Mach cone shock front is located only close downstream from the accretor. Here, the unshocked, low-β, fast-moving gas drags the shocked gas downstream along field lines rather than allowing the Mach cone to extend at the same opening angle. Upstream shocks develops for all $\mathcal{M} = 1.41$ runs, either immediately when $\beta = 1$ ($\mathcal{M}_{\rm A} = 1$) or at later times for $\beta > 1$. Figure 4.7 plots the shock front location as a function of time for $\beta = 1$ and $\mathcal{M} = 1.41$. A least-squares fit to an exponential function suggests the shock will vanish at $\sim 3r_{\rm ABH}$, where $\mathcal{M}_{\rm A}$ drops to unity.

As noted above, switch-on shocks occur only for $\beta < 2$. Our simulations show that perpendicular field components can develop in the flow at a finite distance downstream of the shock; of course, if the shock is not exactly parallel, then the upstream perpendicular component of the field can be amplified by the compression in the shock. In either case, the perpendicular field component causes material to pile up on one side of the sink particle, and the inertia of the gas drives kinks in the field farther downstream from the shock (see the left side of Figure 4.4). As the field piles up, magnetic pressure eventually dominates and the field straightens itself out, but overshoots, collecting on the opposite side of the accretor. The resulting morphology upstream is an oscillating motion of the field and the flow with a period of $\sim t_{\rm B}$.

Material not immediately upstream of the accretor flows into the Mach cone and then primarily travels through the region immediately downstream of the shock towards the sink particle. However, we note that for these particular simulations, the region of reduced B along the Mach cone is most likely due to numerical reconnection because the field flips orientation over a few grid cells; we have not carried out higher resolution simulations to test if this structure is converged. However, we have ensured that the mass accretion rate is converged, as discussed in the next subsection. The Mach cone results in the downstream being less dense than the background, rather than material forming a downstream wake as in the hydrodynamic limit.

For $\mathcal{M} = 4.47$, the bowshock only forms for $\beta = 0.01$. Even though $\beta = 0.1$ gives $\mathcal{M}_{\rm A} = 1$ initially, the region interior to $r_{\rm ABH}$ is so close to the accretor that any decrease with β also occurs with an increase in $\mathcal{M}$, resulting in $\mathcal{M}_{\rm A} < 1$ never being satisfied. For larger β, even less field enhancement occurs upstream. As a result, the $\mathcal{M} = 4.47$ runs resemble non-magnetic flows, with a downstream “wake" forming as the region of gas shocked from the Mach cone. The majority of the mass accretion occurs through this wake.

Perpendicular Orientations

For the perpendicular orientation, shocks occur if the flow velocity exceeds v_F, which is $(v_{\rm A}^2 + c_{\rm s}^2)^{1/2} = c_{\rm s}(1 + 2/\beta)^{1/2}$ in this case. Even if this condition is not initially satisfied, a magnetosonic wave launched from the sink particle boundary can steepen into a weak shock as it moves upstream from the sink particle. We see this, for example, in the $\mathcal{M} = 1.41$, $\beta = 1$ case. Initially $v_F = (3/2)^{1/2} v_0$. Immediately upstream of the shock, both ρ and B are increased from compression, but ρ has increased even further from material falling down field lines. This results in an increased β and reduced v_F. Figure 4.7 plots this front location. Since β has increased to ~ 1.2, the wave (which has steepened into a weak shock) moves at a speed $\sim (1 + 2/1.2)^{1/2} c_s - v_0 \sim 0.22 c_s$ upstream, relative to the accretor. Perpendicular to the flow, a weak shock moves outward at $\sim v_F$.

The $\mathcal{M} = 1.41$ runs all show the development of a dense irrotational disk around the accretor interior to $\sim 0.25 r_{\rm B}$, with the disk normal perpendicular to the incoming flow. For $\beta = 1$ and 10, this disk attaches to a downstream wake. In the $\beta = 100$ case, colliding flows have made the inner $\sim 0.25 r_{\rm B}$ flow unstable, similar to the oscillating flow we saw in the parallel cases. The weak field ends up draped around the shocks that form around the sink particle. For $\mathcal{M} = 4.47$, again the flow resembles the non-magnetic case. No bowshock is launched except in the $\beta = 0.1$ case, but, as discussed above, this is a transient of the flow.

We also perform one run at $\beta = 1$, $\mathcal{M} = 1.41$ with a 45-degree angle between the flow and magnetic field. Within $\sim r_{\rm ABH}$, the flow tends to align itself with the local magnetic field, and the general flow resembles that of the parallel orientation. The remaining runs, with $\mathcal{M} = 44.7$ and $\beta = 0.1$ or with $\mathcal{M} = 1.41,\ 4.47$ and $\beta = 10^{30}$, are dynamically dominated by the ram pressure of the gas and therefore closely resemble non-magnetic hydrodynamic flow.

4.4.2 Mass Accretion Rate

For each simulation, the mass accretion rate, $\dot{M}$, rises with time and then levels off after a few $t_{\rm B} = r_{\rm B}/c_{\rm s}$. Each simulation is run until the rate plateaus for a least a few $t_{\rm ABH}$. The rate quoted in Table 1 is the average over the last 1/6 of the integration, following Cunningham et al. (2012).

Figure 4.8 gives an example of the time evolution of $\dot{M}$ for the $\beta = 1$ runs. The rate is averaged over $0.2t_{\rm B}$ bins to reduce noise. In general, when $\mathcal{M} < 1$ or $\mathcal{M}_{\rm A} \gg 1$, the orientation of the field makes little difference in the final accretion rate despite the very different morphologies. When $\mathcal{M}_{\rm A} \approx 1$, the perpendicular rate always exceeds the parallel rate. When $\mathcal{M} \gg 1$, $\dot{M}_{\rm HL}$ well approximates the perpendicular accretion rate, even, surprisingly, when $\mathcal{M}_{\rm A} \approx 1$.

In the high β (≥ 100) subsonic runs of both our work and that of Cunningham et al. (2012), the accretor undergoes a period of rapid accretion before the accretion rate suddenly drops to $\sim 1/2$ of the original value (see Figure 6, right panel of Cunningham et al. 2012). The reason for this effect is that the magnetic field eventually becomes dynamically dominant after enough flux has built up near the accretor. Cunningham et al. (2012) showed that this occurs after $t \gtrsim (\beta/100)^{1/2} t_{\rm B}$ for a dynamically weak field and could take an arbitrarily long time as $\beta \to \infty$. We note that we do not see this effect for our $\mathcal{M} = 1.41$ run with $\beta = 100$. In this case, the magnetic flux is unable to appreciably build up within $r_{\rm ABH} \sim r_{\rm B}/4$ before the gas pulls the flux downstream. Below, when we determine the best fit parameters for our interpolation formulas, we use the initial (larger) steady-state value for the mass accretion rate since in astrophysical applications the flow is often not steady for long time periods.

Two requirements are needed to ensure that the accretion rate has converged. First, the value of $\dot{M}$ should not depend on the resolution of the grid. As explained in §3, we have verified that this is the case. Increasing the resolution also decreases the size of our sink particle, which has a radius of $4\Delta x$. The second requirement is to ensure that the sink particle boundary conditions cannot influence the value of $\dot{M}$. To do this, we require that the accreting gas pass through the fast magnetosonic point at $r > 4\Delta x$, so that it becomes causally disconnected from the ambient medium before encountering the sink region. For each run, we calculate the mass-weighted volume average of $\mathcal{M}_F = v/v_F$ for the cells either 5 or 6 Δx from the sink particle. Recall that gas is removed from flux tubes inside the sink region. The resulting low-density flux tubes are interchange unstable and will rise away from the accretor. Since we are interested in verifying that the accreting gas is causally disconnected from the ambient medium, we include only accreting gas (i.e., gas from cells where $\mathbf{v} \cdot \mathbf{x} < 0$, where $\mathbf{x}$ is the position vector from the sink particle's center) in calculating the average of $\mathcal{M}_F$. This averaged value is given in Table 1. For all our runs, we confirm that we have captured the transition.

With the simulation data, we can now determine $\beta_{\rm ch}$ and n in our proposed interpolation formulas (Equations 4.27 and 4.30) from Section 4.2.2 . Recall that these formulas generalized and built upon previous known analytic and numerical results. In particular, we have proposed a simple interpolation formula for the Bondi-Hoyle limit (Equation 4.9), which matches the simulation results of Ruffert (1996). We performed two $\beta = 10^{30}$ runs to verify this equation, finding it underestimates the true accretion rate by only 19% and 3% for $\mathcal{M} = 1.41,\ 4.47$, respectively. The mass accretion rate for the diagonal case lies between the the parallel and perpendicular rates; the average of the predicted parallel and perpendicular rates (Equation 4.33) reproduces this diagonal case to 18%.

We perform a least-squares fit to Equations (4.27) and (4.30) using the union of data from this work and from Cunningham et al. (2012). Since the values of the accretion rate can vary over orders of magnitude, we define the residuals in the least-squares function to be the difference of the logarithms rather than of the absolute values:

$$S = \sum (\log_{10} \dot{M}_{\rm data} - \log_{10} \dot{M}_{\rm fit})^2 \ . \tag{4.39}$$

Each data point is given an equal statistical weighting. Minimizing S, we find $\beta_{\rm ch} = 18.3 \pm 0.004$ and $n = 0.94 \pm 0.15$ with $S = 0.956$. We do not include the diagonal run or the two hydro runs in our fit.

The standard errors show that matching the data to these interpolation formulas is not terribly sensitive to the exact value of n. Since the data are consistent with using $n = 1$, we adopt this value for simplicity. Fixing $n = 1$, a least-squares fit to the data yields $\beta_{\rm ch} = 19.8 \pm 0.006$ with $S = 0.96$. We fix $\beta_{\rm ch}$ as 19.8.

In Section 4.2.2 we wrote all accretion rates in terms of $\dot{M}_{\rm B}$, which is constant across our entire parameter space. Since we are studying gas initially in uniform motion, normalizing to $\dot{M}_{\rm BH}$ (Equation 4.10) is also useful. Indeed with this normalization, our parallel accretion rate can be written in terms of one parameter

$$\mathcal{M}_{\rm BH}\beta^{1/2} = 2^{1/2}\mathcal{M}_{\rm BH}\left(\frac{c_s}{v_{\rm A}}\right) \ , \tag{4.40}$$

which varies as B^{-1} if the other parameters are held constant ($\mathcal{M}_{\rm BH}$ is defined in Equation 4.12; recall that $\mathcal{M}_{\rm BH} \to 1$ as $\mathcal{M} \to 0$). Equations (4.27) and (4.30) can be rewritten as

$$\frac{\dot{M}_{\parallel}}{\dot{M}_{\rm BH}} = \left[1 + \frac{4.4}{(\mathcal{M}_{\rm BH}\beta^{1/2})}\right]^{-1} \ , \tag{4.41}$$

$$\frac{\dot{M}_{\perp}}{\dot{M}_{\rm BH}} = \min\left\{1 \ , \ \mathcal{M}_{\rm BH}\left[1 + \frac{4.4}{(\mathcal{M}_{\rm BH}\beta^{1/2})}\right]^{-1}\right\} \ . \tag{4.42}$$

The perpendicular rate, however, potentially requires knowledge of both $\mathcal{M}_{\rm BH}$ and β individually. Figures 4.9 and 4.10 plot these fits for the parallel and perpendicular orientations. These fits are able to reproduce the simulation data to within a factor of three.

For typical molecular cloud values of $\mathcal{M} \sim 5$ and $\beta \sim 0.04$ (Crutcher 1999), which corresponds to $\mathcal{M}_{\rm A} = 0.71$, $\mathcal{M}_{\rm BH} = 5.15$, and $\mathcal{M}_{\rm BH}\beta^{1/2} = 1.03$, Equation (4.33) gives an accretion rate of $4.2 \times 10^{-3}\ \dot{M}_{\rm B}$. For the fiducial parameters given in Equation (4.4), this corresponds to $4.3 \times 10^{-9}\ M_\odot\ {\rm yr}^{-1}$, which is not that different from the hydrodynamical prediction (Equation 4.10) of $\dot{M}_{\rm BH} = 6.9 \times 10^{-9}\ M_\odot\ {\rm yr}^{-1}$. However, for smaller Mach numbers of ~ 2 and 0, the ratio of the predicted average accretion rate to the hydrodynamic value decreases to 0.19 and 0.048, respectively. Figure 4.11 plots our parallel and perpendicular fits, normalizing to $\dot{M}_{\rm B}$ (Equation 4.3). Here the disparity between the perpendicular and parallel fits can be seen, especially when $\mathcal{M}_{\rm A} < 1$. We remind the reader that our runs only have one instance where $\mathcal{M}_{\rm A} < 1$ and $\mathcal{M} > 1$: $\beta = 0.01$ and $\mathcal{M} = 4.47$, and we found little difference between the parallel and perpendicular rates, whereas the fits predict the perpendicular rate should be 4.6 times the parallel rate. However, this region of parameter space is only a small region of the overall parameter space (Figure 4.2) and our fits do predict the disparity between the parallel and perpendicular rates in the other three regions, as well as when $\mathcal{M}_{\rm A} = 1$. Since we chose our non-magnetized limit to well reproduce the results of Ruffert (1996), our fits also succeed in predicting the accretion rates for $\mathcal{M} < 1$ and $\mathcal{M}_{\rm A} > 1$, even though we performed no runs ourselves in this region of parameter space.

4.5 Validity of the Steady-State Approximation

In both this work and that of Cunningham et al. (2012), we have made several approximations in our analysis of Bondi- and Bondi-Hoyle-type accretion: (1) the accretion must be in a steady-state (at least when averaged over times $\sim t_{\rm B} = r_{\rm B}/c_{\rm s}$); (2) the accreting gas must not be self-gravitating; and (3) the accretion rate must be determined by the mass of the particle, not by the gravitational collapse of the ambient medium. As we shall see, these approximations are all connected. We have also assumed that the ambient medium is uniform and that the particle is small compared to $r_{\rm ABH}$, but we shall not discuss these approximations here. To keep our discussion simple, we restrict ourselves to Bondi accretion.

We define the Bondi mass as

$$M_{\rm B} \equiv 4\pi\rho_0 r_{\rm B}^3 \ , \tag{4.43}$$

so that the Bondi accretion rate is

$$\dot{M}_{\rm B} \simeq \frac{M_{\rm B}}{t_{\rm B}} \ , \tag{4.44}$$

where the approximation consists of setting $\lambda \simeq 1$. The Bondi mass is the mass of gas located within the Bondi radius of the particle and is approximately the mass accreted within one Bondi time. For steady-state accretion, the mass of the particle must change slowly, i.e., the mass accreted in one Bondi time must be small compared to the particle mass: $\dot{M}_{\rm B} t_{\rm B} \ll M_*$. Therefore, the steady-state approximation reads

$$\frac{M_{\rm B}}{M_*} \simeq \frac{\dot{M}_{\rm B} t_{\rm B}}{M_*} \ll 1 \ . \tag{4.45}$$

The self-gravity of the ambient gas is characterized by the gravitational mass,

$$M_{\rm G} \equiv \frac{c_{\rm s}^3}{\sqrt{G^3\rho_0}} \ ; \tag{4.46}$$

the Bonnor-Ebert mass, the maximum mass of an isothermal sphere in hydrostatic equilibrium, is $1.182 M_{\rm G}$ (Ebert 1955; Bonnor 1956). We then have the identity

$$M_{\rm B} M_{\rm G}^2 = 4\pi M_*^3 \ , \tag{4.47}$$

which implies

$$\frac{M_{\rm B}}{M_{\rm G}} = \frac{1}{\sqrt{4\pi}} \left(\frac{M_{\rm B}}{M_*}\right)^{3/2} \ll 1 \tag{4.48}$$

(Equation 4.45). The condition for the accreting gas to be non-self-gravitating is that the mass inside the Bondi radius be small compared to the gravitational mass, $M_{\rm B} \ll M_{\rm G}$. Equation (4.48) then implies that *Bondi accretion is in a steady state if and only if the accreting gas is not self gravitating.*[6] In other words, the first two approximations listed at the start of this section are really only one approximation. The steady-state approximation, together with the identity (4.47), place an upper bound on the particle mass,

$$\sqrt{4\pi}\,\frac{M_*}{M_{\rm G}} = \left(\frac{M_{\rm B}}{M_*}\right)^{1/2} \ll 1\ , \tag{4.49}$$

so that

$$M_* \ll 0.36 \left(\frac{T}{10\,{\rm K}}\right)^{3/2} \left(\frac{10^4\ {\rm cm}^{-3}}{n_0}\right)^{1/2} \quad M_\odot\ , \tag{4.50}$$

where n_0 is the density of hydrogen nuclei in the ambient gas. The right-hand side of Equation (4.50) is just $M_{\rm G}/\sqrt{(4\pi)}$. If $M_* \gtrsim M_{\rm G}/\sqrt{(4\pi)}$, the gas mass within $r_{\rm B}$ is massive enough to be self-gravitating, and therefore the mass of the particle will not change slowly. For example, if $M_* = 1.0\,M_\odot$, then for the fiducial parameters above, $M_{\rm B} \approx 2.5\ M_\odot$, which exceeds $M_{\rm G} \approx 1.3\ M_\odot$: The fact that $M_{\rm B}$ exceeds $M_{\rm G}$ means that the gas is self-gravitating, and the fact that $M_{\rm B}$ exceeds M_* means that that accretion is not in a steady state. Note that Equation (4.50) is based on the assumption that turbulence inside $r_{\rm B}$ is negligible; if turbulence is important on that scale, as it may be in regions of high-mass star formation, then the theory presented here would have to be generalized, as it was for the non-magnetic case by Krumholz et al. (2006).

Finally, we compare the accretion rate due to gravitational collapse,

$$\dot{M}_{\rm G} \sim \frac{c_{\rm s}^3}{G}\ , \tag{4.51}$$

(Shu 1977) with that due to Bondi accretion. Observe that the mass accreted due to gravitational collapse in one Bondi time is very large, $\dot{M}_{\rm G} t_{\rm B} \sim M_*$, so that

$$\frac{\dot{M}_{\rm G}}{\dot{M}_{\rm B}} \sim \frac{M_*}{M_{\rm B}} \gg 1\ , \tag{4.52}$$

for steady Bondi accretion. We therefore have the apparently paradoxical result that the rate of accretion via gravitational collapse greatly exceeds that due to Bondi accretion when the latter is in steady state ($M_{\rm B} \ll M_*$). There are then two possibilities for normal Bondi

---

[6] Lee & Stahler (2011) also showed that a steady state is realizable for Bondi-Hoyle accretion when the gas is not self-gravitating.

accretion: First, the cloud in which the accreting particle is embedded could be gravitationally stable, making $\dot{M}_{\rm G}$ not meaningful (for the simple isothermal, unmagnetized case we are considering, that requires that the cloud mass, $M_{\rm c}$, be less than $M_{\rm G}$). Second, gravitational collapse could occur on a large scale, but not be focused on the accreting particle. A real molecular cloud is turbulent and inhomogeneous, and it can undergo gravitational contraction without having mass accumulate at a central point. Our analysis is valid provided the cloud is approximately uniform within a Bondi radius of the accreting particle.

In summary, Bondi-type accretion is in steady state if and only if the gas inside the Bondi radius is not self-gravitating,

$$\text{Steady state} \;\leftrightarrow\; M_{\rm B} \ll M_* \ll M_{\rm G}\ . \tag{4.53}$$

Figure 4.12 shows this schematically. The steady-state condition places an upper limit on the particle mass, $M_* \ll M_{\rm G}/\sqrt{(4\pi)}$ (see Equation 4.50).

This discussion is directly relevant to the issue of whether stars form by gravitational collapse or competitive accretion (Bonnell et al. 2001; Krumholz & McKee 2008). The isothermal sound speed, $c_{\rm s}$, must be replaced by the one-dimensional velocity dispersion, σ, in both $M_{\rm G}$ (McKee & Ostriker 2007) and $M_{\rm B}$ (Krumholz et al. 2006). In order for competitive accretion to dominate, one requires $\dot{M}_{\rm B} \gg \dot{M}_{\rm G}$, and according to Equation (4.52) this implies $M_{\rm B} \gg M_*$: The accreting gas must be self-gravitating, the accretion is not in a steady state, and the upper limit on the protostellar mass in Equation (4.50) does not apply. However, Equation (4.47) then implies that $M_{\rm G} \ll M_*$. Since M_* is less than the mass of the cloud from which it is accreting, $M_{\rm c}$, it follows that $M_{\rm G} \ll M_{\rm c}$: the cloud is very sub-virial. That is, the ratio of the kinetic energy in the cloud to the gravitational energy, which is of order the virial parameter $\alpha_{\rm vir} \equiv 5\sigma^2 R/GM_{\rm c}$, is much less than unity. Krumholz & McKee (2008) argued that since molecular clouds appear to have virial parameters of order unity, this implies that stars form by gravitational collapse, not competitive accretion. If the observed virial parameter reflects collapse instead of turbulence, then competitive accretion may be viable, but the accretion rates would be reduced by magnetic fields.

4.6 Summary and Discussion

The accretion of gas onto an object due to its gravity is generally referred to as Bondi accretion when the object is stationary and Bondi-Hoyle accretion when the object is moving. Such accretion has been employed in many observational, theoretical, and numerical studies to explain the growth of planets, brown dwarfs, stars, compact objects, and supermassive black holes, to name a few (e.g., Hopkins et al. 2006; Kokubo & Ida 2012; Andre et al. 2012; Toropina et al. 2012). Here we have determined the effects of a uniform magnetic field on Bondi-Hoyle accretion under the assumption that the gas is isothermal and that the accreting mass is a point particle, thereby generalizing the results of Cunningham et al. (2012) to include the effects of the motion of the accretor. In keeping with most previous treatments of Bondi and Bondi-Hoyle accretion, we did not consider the effects of stellar winds and outflows on the steady-state accretion rate.

Our primary application is to protostellar accretion, but our results should apply to stellar accretion in any medium in which the gas is approximately isothermal. We have not considered the effects of stellar winds, which in some cases are strong enough to suppress accretion. Our results might also be applicable to accretion onto supermassive black holes in active galactic nuclei. There the Bondi radius is $r_{\rm B} \simeq 3(M/10^8 M_\odot)(10^7\,{\rm K}/T)$ pc. Compton heating and cooling can maintain isothermality near the Bondi radius if the luminosity is sufficiently high (Woods et al. 1996). In some cases, Compton-heated gas is thermally unstable, and Gaspari et al. (2013) have shown that then Bondi accretion rates based on the temperature of the hot gas can underestimate the true accretion rate by up to two orders of magnitude.

The time-averaged mass-accretion rate for isothermal accretion flow onto a static point mass of mass M_* was determined by Bondi (1952), $\dot{M}_{\rm B} = 4\pi\lambda\rho_0 {r_{\rm B}}^2 c_s$, where λ is a numerical constant, ρ_0 is the ambient density, $r_{\rm B} = GM_*/c_s^2$ is the Bondi radius, and c_s is the isothermal sound speed. If the object is moving, then the morphology of the accretion flow and the accretion rate also depend on the sonic Mach number, $\mathcal{M} = v_0/c_{\rm s}$, where v_0 is the velocity of the mass through the ambient medium. If the medium is magnetized, two additional parameters enter, the plasma $\beta = 8\pi\rho_0 c_{\rm s}^2/B_0^2$ and θ, the angle between the the field and the velocity of the object relative to the medium. (For moving objects, β can be replaced by the Alfvén Mach number, $\mathcal{M}_{\rm A} = v_0/v_{\rm A,0} = (\beta/2)^{1/2}\mathcal{M}$, where $v_{\rm A,0}$ is the Alfvén velocity in the ambient medium). When both $\mathcal{M}$ and $\mathcal{M}_{\rm A}$ are large, the accretion resembles non-magnetized Bondi-Hoyle flow. When either $\mathcal{M}$ or $\mathcal{M}_{\rm A}$ is small, the ambient medium is approximately static and the flow resembles the stationary magnetized models of Cunningham et al. (2012). Here we have explored the case in which both magnetic fields and motion of the mass through the medium are important by performing three-dimensional simulations of a gravitating point particle accreting from an initially uniform, isothermal gas pervaded by a uniform magnetic field that is either parallel or perpendicular to the direction of motion. Since the magnetic flux in stars is small compared to that in the gas from which the stars formed, we assume that only gas, not magnetic flux, accretes onto the point mass (discussed in §4.3). Our main results are approximate expressions (4.27) and (4.30) for the accretion rates, which reduce to known numerical and analytic limits and agree with our simulation data and that of Cunningham et al. (2012) to within a factor of three (see Figures 4.9 and 4.10).

The key assumption underlying the theory of Bondi-Hoyle accretion is that the gas is not self-gravitating on the scale of the Bondi radius or, equivalently (as shown in Section 4.5) that the accretion rate is steady after averaging over the fluctuations that occur on time scales $\lesssim t_{\rm B}$. This assumption must be validated for each astrophysical situation that employs it. The conditions for the validity of the steady-state assumption are that the stellar mass be larger than the Bondi mass, $M_{\rm B} = 4\pi\rho_0 {r_{\rm B}}^3$, but smaller than $M_{\rm G}$, the mass at which self-gravity becomes important, so that $M_{\rm B} \ll M_* \ll M_{\rm G}$ (Equation 4.53). For the simple case we considered in Section 4.5, in which magnetic fields are negligible (and, as is true throughout this chapter, turbulence is also negligible), the steady-state assumption is valid for stars less than $\sim 0.4 M_\odot$ for fiducial molecular cloud parameters (Equation 4.50).

Sub-grid particle accretion methods have been employed to model protostellar accretion in large-scale numerical simulations of molecular clouds. Our results should be of particular utility for extending the sub-grid accretion models in such codes. Previous work has used unmagnetized accretion rates even though the sink particles were moving through a magnetized medium, thereby overestimating the true accretion rate onto the particle. We outline our implementation of Equations (4.27) and (4.30) in `ORION2` sink particles in Appendix B and demonstrate that this implementation succeeds in reproducing the correct accretion rate even when the accretion length scale $\sim r_{\rm ABH}$ is not well resolved (Figure B.3). However, it should be noted that our accretion rates apply to gas that is not turbulent, and so they do not include the reduction associated with vorticity (Krumholz et al. 2005).

Finally, our results have implications for the theory of star formation. At present, there are two main paradigms for the formation of massive stars: gravitational collapse, in which stars form via the gravitational collapse of a pre-existing protostellar core (McKee & Tan 2003), and competitive accretion, where protostars compete for gas from a common reservoir initially unbound to the stars (Zinnecker 1982; Bonnell et al. 1997, 2001). Our study shows that magnetic fields make competitive accretion scenarios for the growth of pre-main sequence stars less efficient than predicted from Bondi-Hoyle accretion rates. For example, the amount of suppression for a cloud with average values of $\beta \sim 0.04$ (Crutcher 1999) and $\mathcal{M} \sim 1/2$ (Bonnell et al. 2001) is a factor of ~ 20 (Figures 4.9 and 4.10). This reduction increases for lower β and $\mathcal{M}$. Models that employ Bondi accretion to transform molecular clouds into stars (e.g., Murray & Chang 2012) may be underestimating the timescale for the buildup of massive stars, and therefore, assuming these massive stars are what eventually destroy the cloud, are underestimating the lifetime of the molecular clouds in these models. Delayed buildup from magnetic fields would predict that these models have molecular clouds that persist beyond their typically observed lifetimes. In case of direct simulations, as long as the Alfvén-Bondi-Hoyle radius is resolved and the flow transitions to super-Alfvénic speeds, the accretion rates obtained will still be correct regardless of the subgrid model (e.g, Price et al. 2012, also see the Appendix).

Acknowledgements

The authors thank the peer reviewer for an insightful report that helped improve the general clarity of the chapter. The authors gratefully acknowledge support from (1) the National Science Foundation: A.T.L. through an NSF Graduate Fellowship, and C.F.M. and R.I.K. through grant AST-0908553 and AST-1211729, (2) the U.S. Department of Energy at the Lawrence Livermore National Laboratory: A.T.L. through grant LLNL-B569409, and R.I.K. and A.J.C. under contract DE-AC52-07NA27344, and (3) NASA: C.F.M. and R.I.K. through ATFP grant NNX13AB84G. Supercomputing support was provided through the NSF XSEDE at the University of Texas at Austin.

Table 4.1: Simulation Parameters

β	$\mathcal{M}$	$\mathcal{M}_{\rm A}$	$r_{\rm ABH}/\Delta x$	$t_{\rm end}/t_{\rm B}$	$\langle\mathcal{M}_{\rm fast}\rangle^a_{\vert\vert}$	$\langle\mathcal{M}_{\rm fast}\rangle^a_{\perp}$	$(\dot{M}/\dot{M}_{\rm B})_{\vert\vert}$	$(\dot{M}/\dot{M}_{\rm B})_{\perp}$	$(\dot{M}/\dot{M}_{\rm B})_{45^\circ}$
100	0.014	0.1	161	8	1.4	1.7	0.323	0.379	
100	1.41	10	54	9	2.2	1.6	0.363	0.332	
10	1.41	3.2	51	7	1.4	1.1	0.273	0.294	
10	4.47	10	8	4	3.8	3.5	0.012	0.011	
1	1.41	1	33	5	1.6	1.0	0.106	0.182	0.116
1	4.47	3.2	7	3	2.1	2.0	0.013	0.012	
0.1	0.447	0.1	8	3	1.7	1.5	0.064	0.061	
0.1^b	4.47	1	16	0.5	1.8	1.1	0.00163	0.0112	
0.1^c	44.7	10	21	3×10^{-4}	8.2	8.6	10^{-5}	8.12×10^{-6}	
0.01^d	4.47	0.32	2	0.5	1.3	1.00	0.0024	0.0026	
∞^e	1.41	n/a	55	5	6.7		0.4		
∞^e	4.47	n/a	8	5	5.5		0.01		

[a] Computed as the volume average over the cells 5 and 6 Δx from the sink particle. Cells are included if the gas is in-falling (i.e., if $\mathbf{v}\cdot\mathbf{x}<0$.)

[b] For this simulation, two additional levels of refinement are allowed for the parallel run, reducing the value of Δx by a factor of 2^2. For the perpendicular run, only one additional level is allowed, reducing Δx by a factor of 2.

[c] For this simulation, one additional level of refinement is allowed and each dimension of the computational domain is reduced from $50\,r_{\rm B}$ to $50/64\,r_{\rm B}$ by a factor of 2^7, reducing the value of Δx by a factor of 2^{7+1}.

[d] For this simulation, one additional level of refinement is allowed, reducing the value of Δx by a factor of 2.

[e] For these simulations, β is set to 10^{30} to approximate non-magnetic flow.

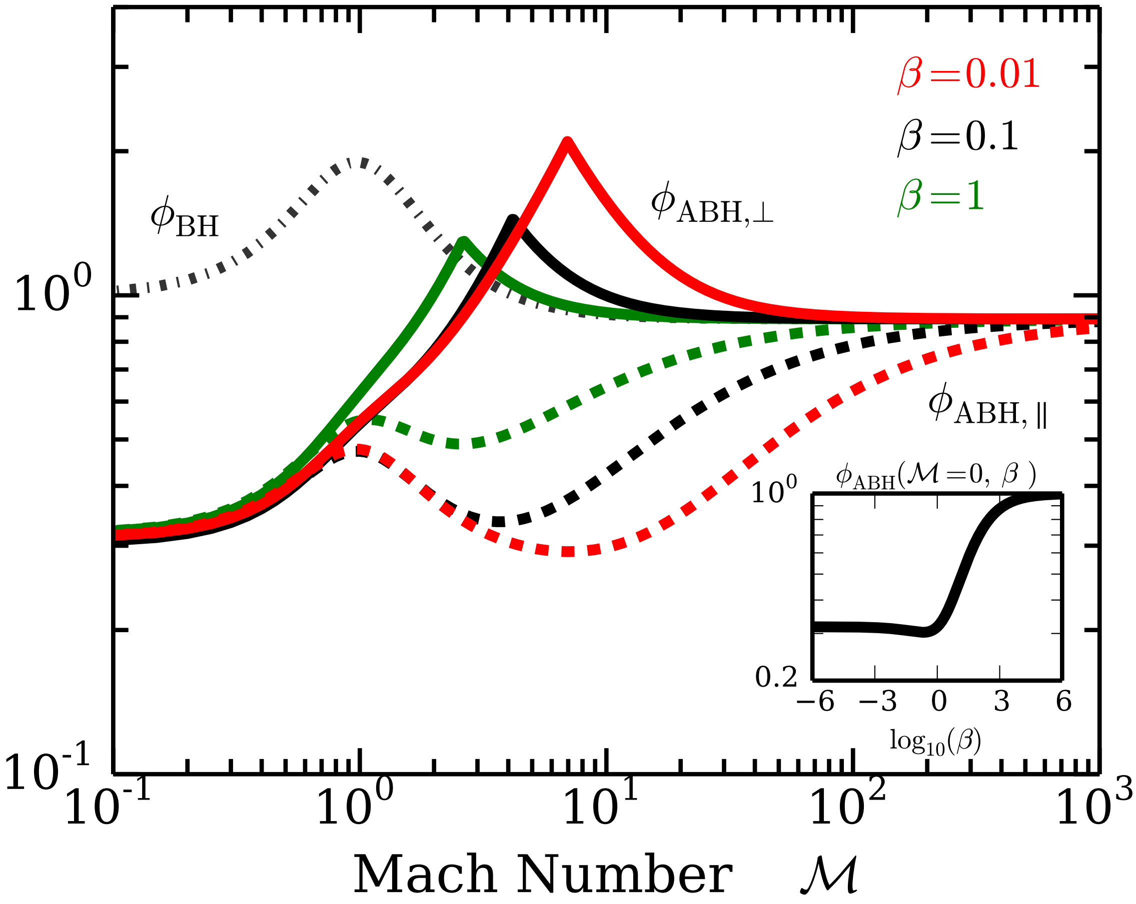

Figure 4.1: Various numerical parameters ϕ as a function of Mach number $\mathcal{M}$ and plasma β. Colors refer to different magnitudes of β, whereas the linestyles differ for different ϕ. The high $\mathcal{M}$ limit is $1/\lambda$. The subplot shows the stationary limit of $\phi_{\rm ABH}$ as a function of β. This has the same functional form as $\phi_{\rm AB}$, defined in the text, but with different fitting parameters $\beta_{\rm ch}$ and n, which are determined in §4.4.2. The asymptotes are $\sqrt{2/\beta_{\rm ch}} \approx 0.32$ and unity.

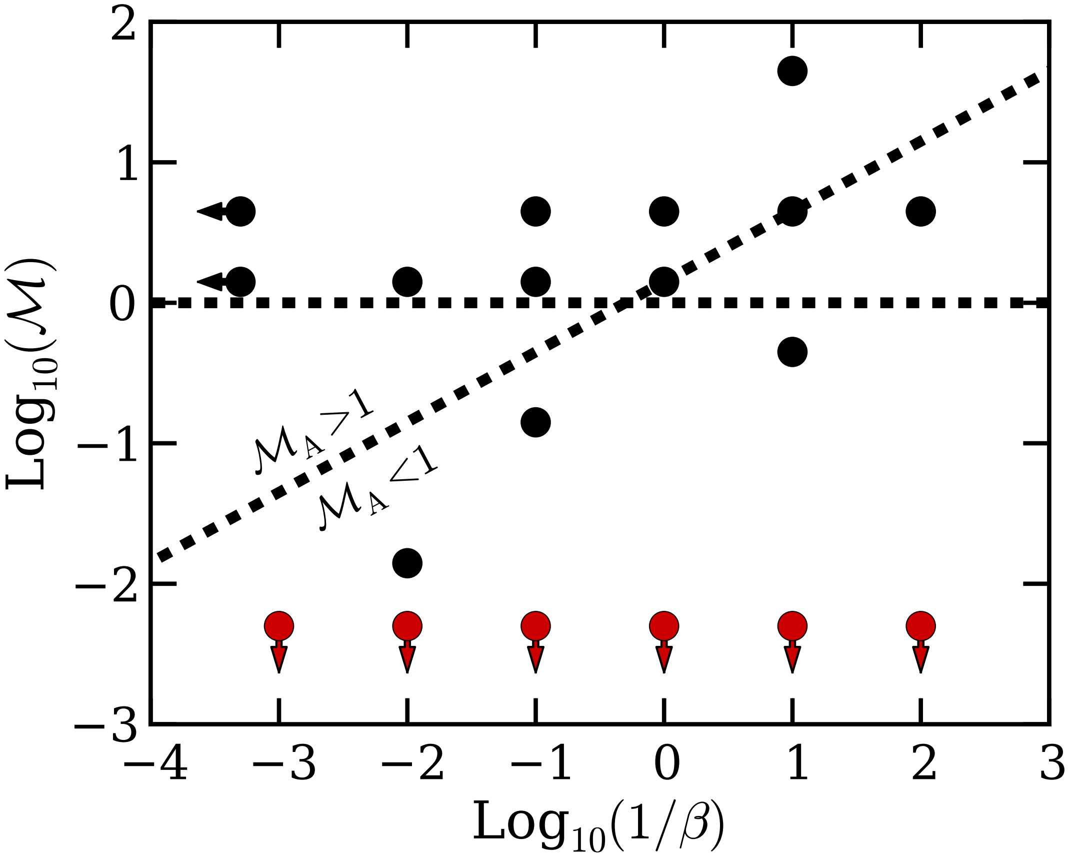

Figure 4.2: Parameter space to be studied. Black dots represent models explored in this work, with the two runs with arrows corresponding to $\beta = 10^{30}$. Red dots are the stationary models of Cunningham et al. (2012). With this choice of axes, the left vertical axis approximates non-magnetic flow, where the bottom horizontal axis approximates stationary flow. The diagonal line plots $\mathcal{M}_{\rm A} = 1$, while the horizontal line plots $\mathcal{M} = 1$. Our runs explore two regions of this parameter space quite well. The region of $\mathcal{M} < 1$ and $\mathcal{M}_{\rm A} > 1$, not explored by us, was studied by Ruffert (1994, 1996) in his investigations of non-magnetized isothermal Bondi-Hoyle accretion. Since typical star forming regions have $\mathcal{M}_{\rm A} \approx 1$ (Crutcher 2012), we also explore two cases with an Alfvén Mach number of unity.

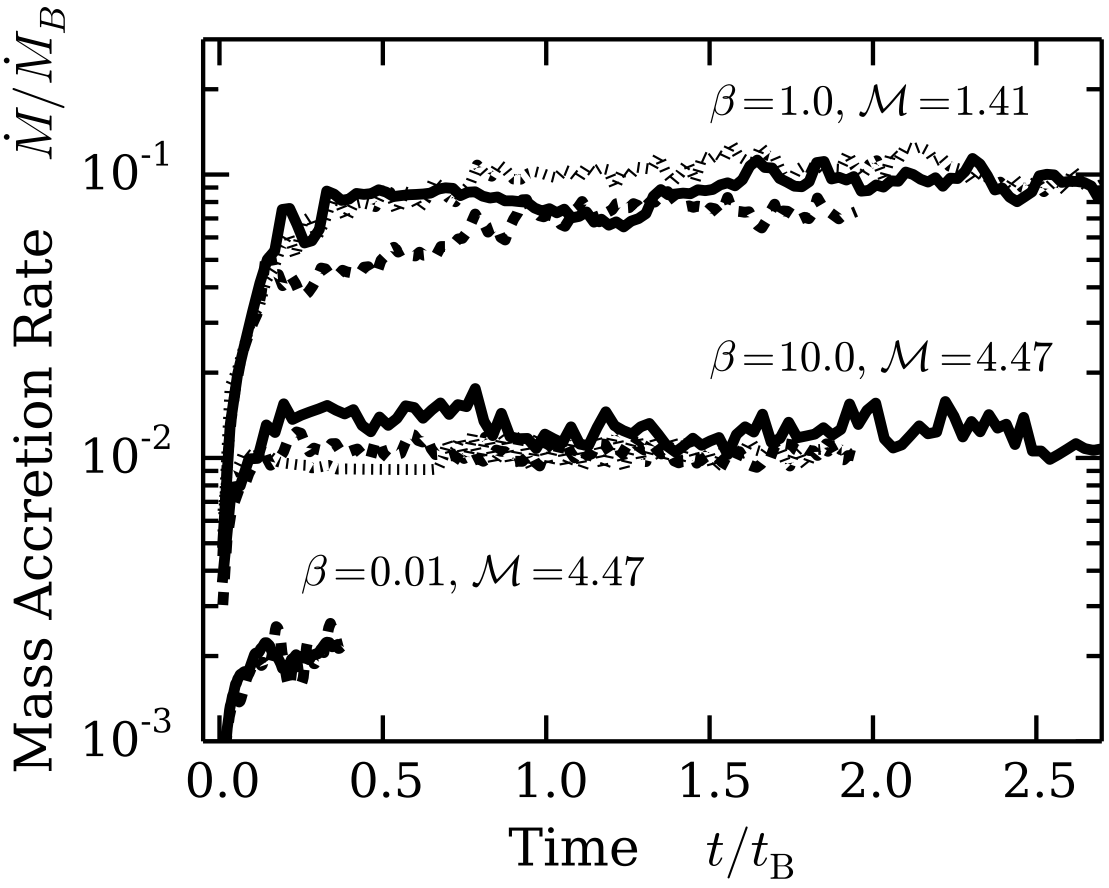

Figure 4.3: Convergence study for three of our marginally resolved models. For the $\beta = 1$ and 10 models, the dotted, solid, and dashed lines represent $L_{\rm max} = 6$, 7, and 8. The solid and dashed represent $L_{\rm max} = 8$, and 9 for the $\beta = 0.01$ model. Increasing the number of levels increases the ratio $r_{\rm ABH}/\Delta x$. For all of these runs the velocity and magnetic field are parallel. The sudden jump for $\beta = 10$, $L_{\rm max} = 6$ at $t/t_{\rm B} \approx 0.7$ occurs because an instability develops in the flow that allows magnetic flux to escape from the region surrounding the sink particle, allowing more mass to accrete. The $\beta = 0.01$ case appears converged even though for eight levels of refinement, $r_{\rm ABH}/\Delta x \sim 2$.

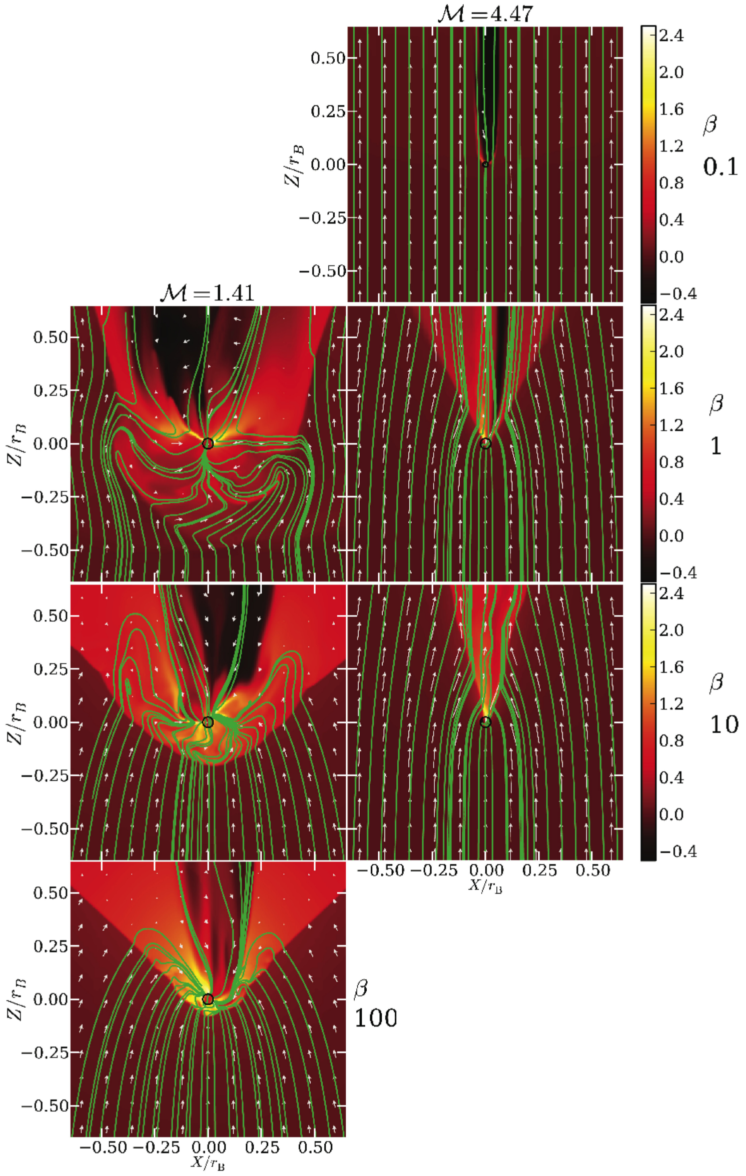

Figure 4.4: Slices in the $x - z$ plane showing the region near the sink particle for the parallel orientations. The left and right columns have $\mathcal{M} = 1.41$ and 4.47, respectively. From top to bottom, the rows show $\beta = 0.1, 1, 10$ and 100. All plots are shown at $t = 3t_{\rm B}$ except the $\beta = 0.1$ plot, which is at $0.5t_{\rm B}$. The color map indicates $\log_{10}(\rho/\rho_0)$, green lines represent magnetic flux tubes drawn from equidistant foot-points $0.5r_{\rm B}$ upstream of the sink particle, and white arrows represent the flow pattern in the plane of the slice. The black circle indicates the size of the sink particle, equal to $4\Delta x$.

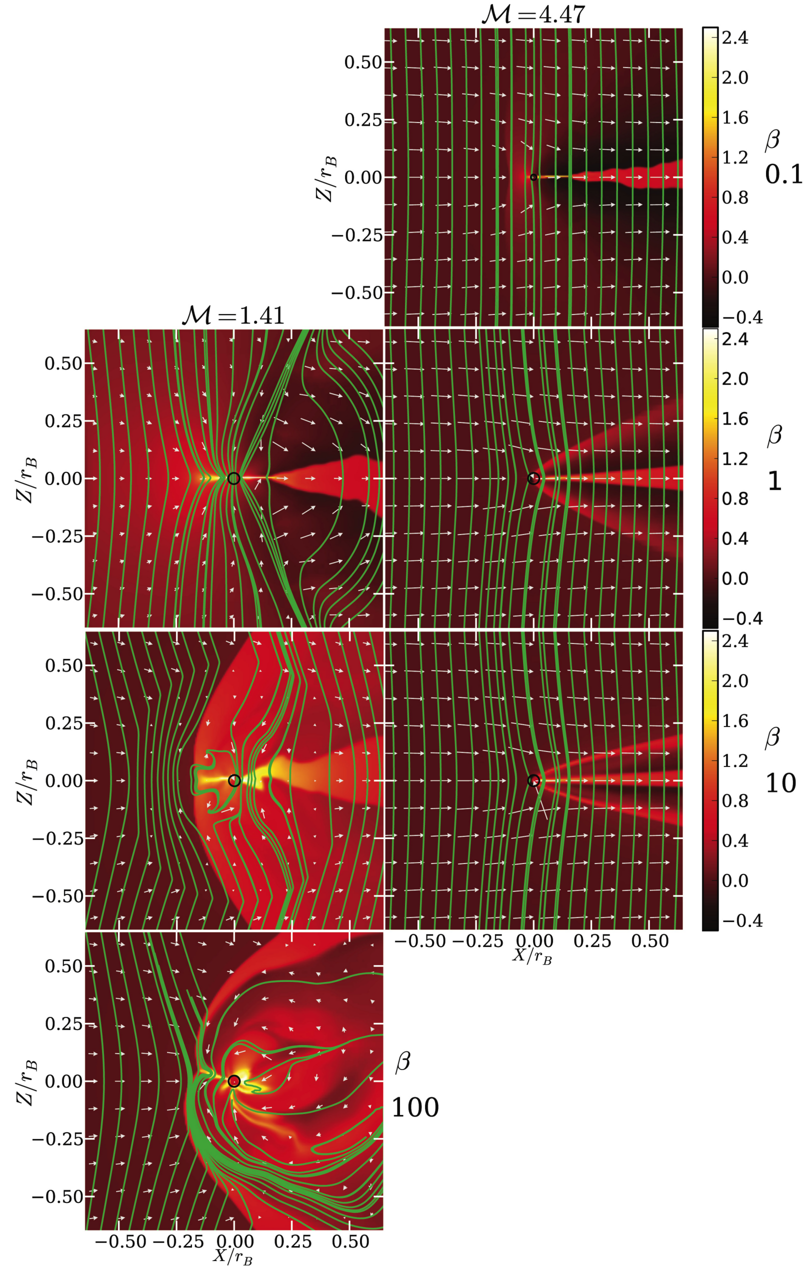

Figure 4.5: Same as Figure 4.4 but for perpendicular orientations.

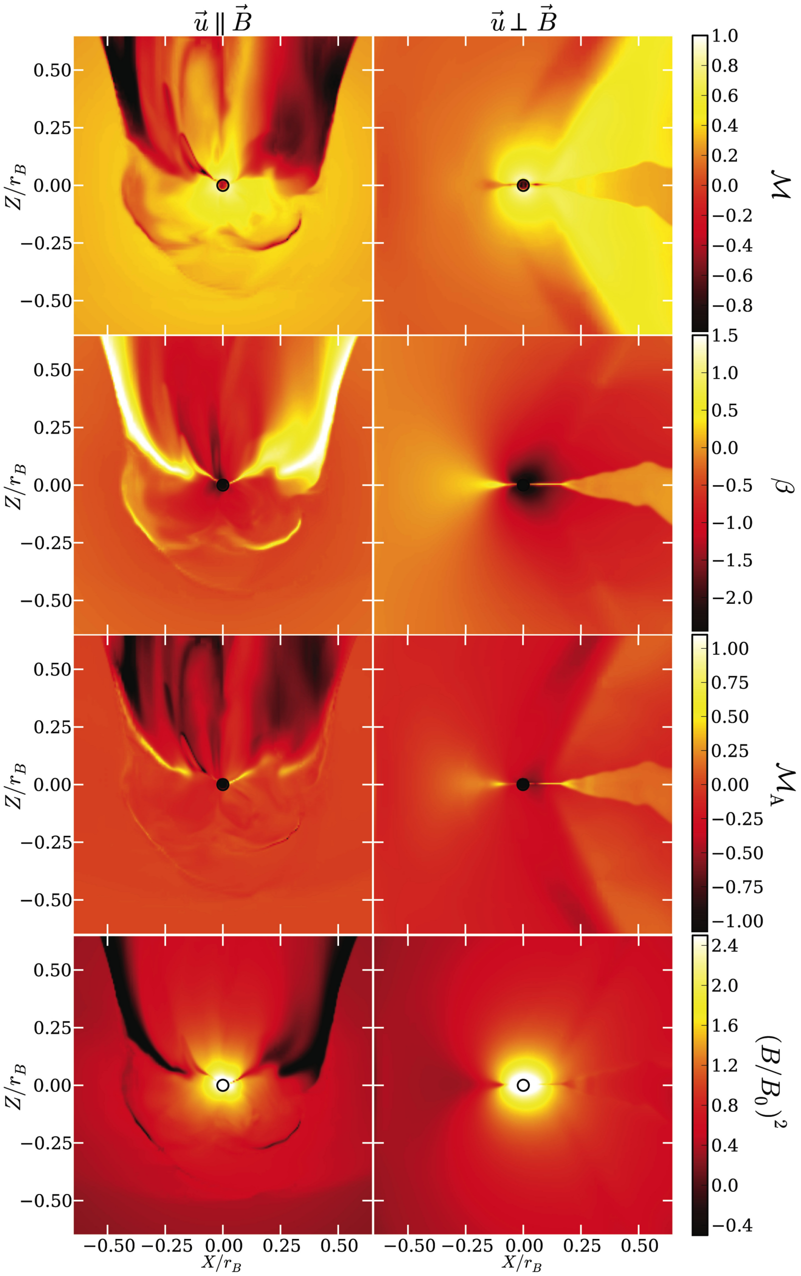

Figure 4.6: Characteristic flow quantities for $\mathcal{M} = 1.41, \beta = 1$ at $t = 3t_{\rm B}$. The left and right columns shown the parallel and perpendicular orientations. The top row shows $\mathcal{M}$, the second row shows β, the third row shows $\mathcal{M}_{\rm A}$, and the bottom row shows $(B/B_0)^2$. The colormaps are in $\log_{10}$ space, where the axes are linear (with units of $r_{\rm B}$). The black circle indicates the size of the sink particle, equal to $4\Delta x$.

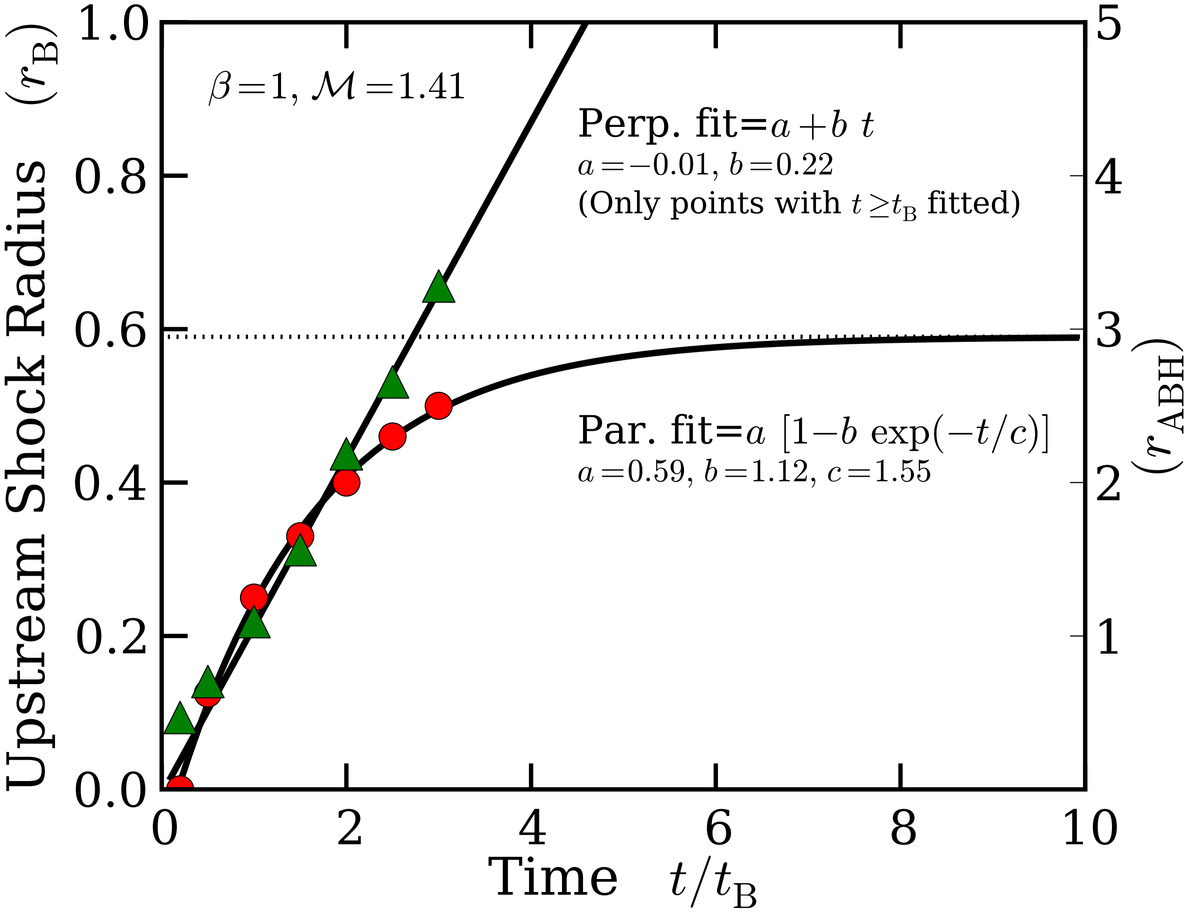

Figure 4.7: Location of the shock front along the upstream axis. To provide data for a fit, the location is plotted at $t/t_B = 0.2, 0.5$, and then every $0.5\, t_B$ until $3\, t_B$, when the simulations end. The parallel orientation is plotted with circles, perpendicular triangles. A least-squares fit to the function shown is performed. The parallel shock velocity tends to zero at $\sim 3\, r_{ABH}$, where the flow is unchanged from the background flow. For the perpendicular case, the shock maintains a nearly constant speed equal to the local magnetofast velocity.

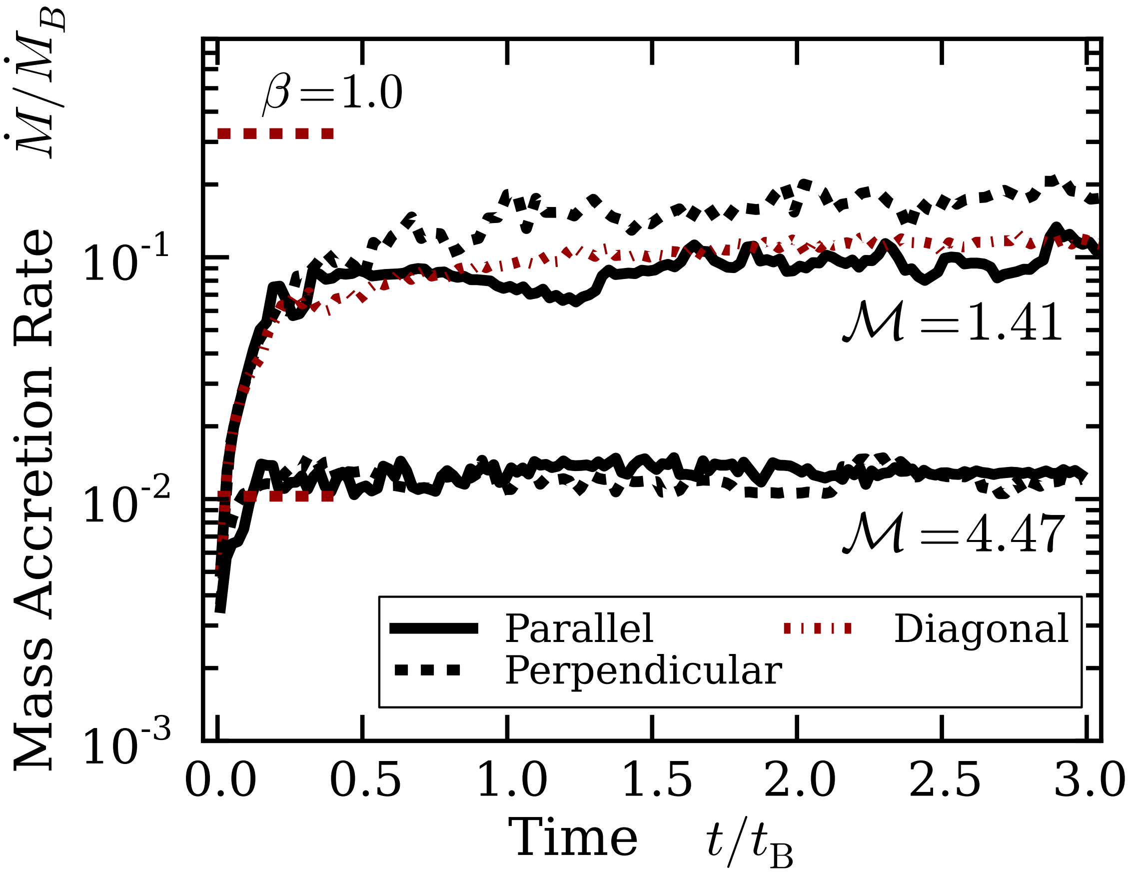

Figure 4.8: Mass accretion rates as a function of time and field orientation for $\beta = 1.0$. All rates are normalized to the Bondi accretion rate (Equation 4.3). The short (red) lines identify the steady-state Bondi-Hoyle accretion rates (Equation 4.9).

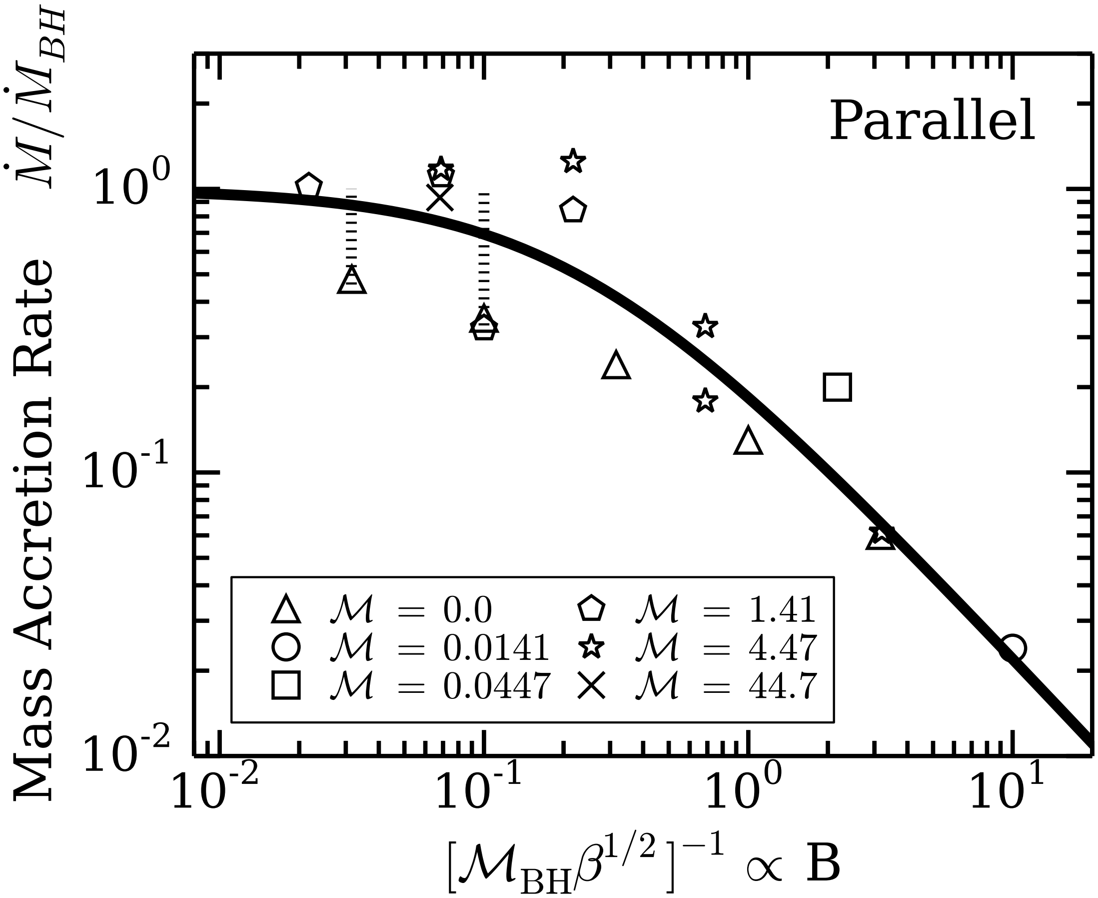

Figure 4.9: Steady-state mass accretion rates for the parallel orientations as a function of the plasma beta β (horizontal axis) and sonic Mach number $\mathcal{M}$ (symbols). Here we have defined $\mathcal{M}_{\rm BH} = v_{\rm BH,eff}/c_{\rm s}$. All accretion rates are normalized to our Bondi-Hoyle accretion rate (Equation 4.9). The solid line is our best fit $\dot{M}_{\|}$ with $\beta_{\rm ch} = 19.8$ and $n = 1.0$ (Equation 4.27 or 4.41). Subsonic runs with $\beta \geq 100$ are plotted with their second steady-state value. The dashed lines connect these points to their initial steady state values (no data point shown).

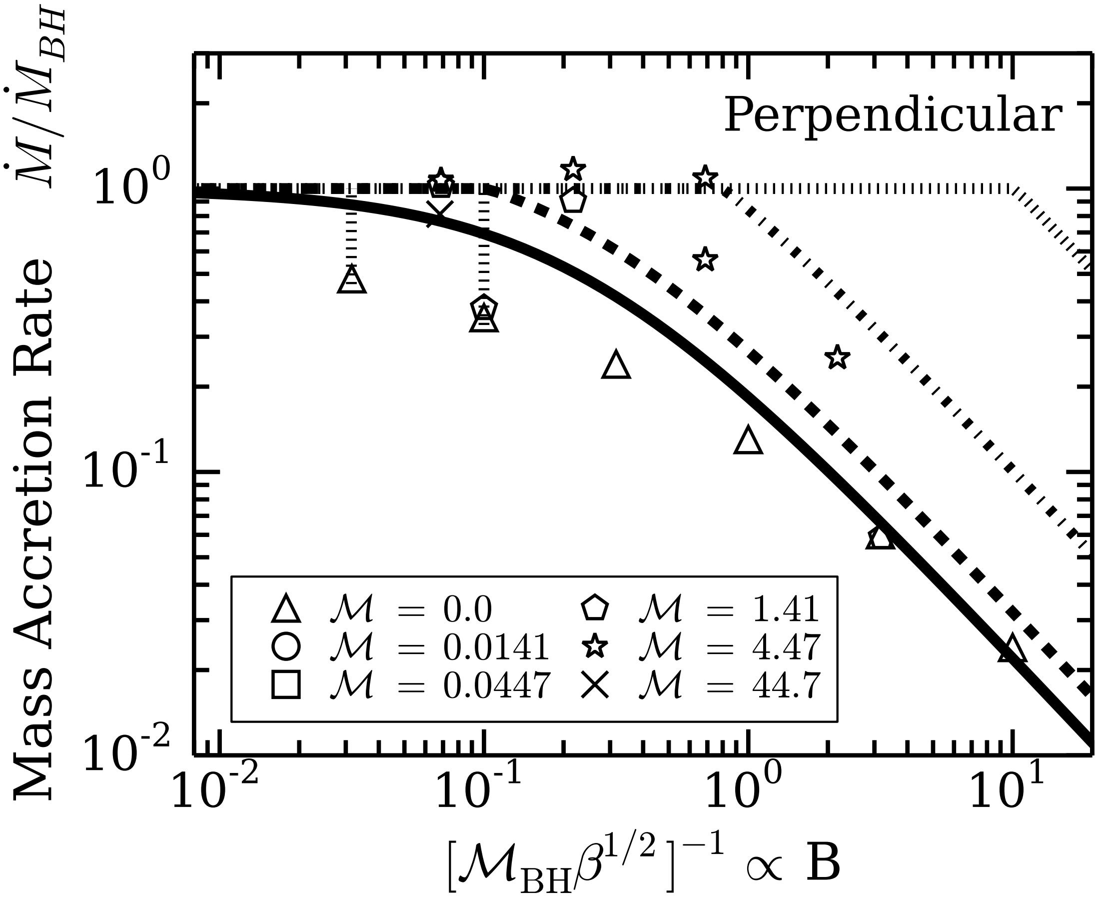

Figure 4.10: Steady-state mass accretion rates for the perpendicular orientations as a function of the plasma beta β (horizontal axis) and sonic Mach number $\mathcal{M}$ (symbols). All accretion rates are normalized to our Bondi-Hoyle accretion rate (Equation 4.9). The lines are our best fit $\dot{M}_\perp$ with $\beta_{\rm ch} = 19.8$ and $n = 1.0$ (Equation 4.30 or 4.42) for four values of $\mathcal{M}$: solid $\mathcal{M} = 0$; dashed $\mathcal{M} = 1.41$; dot-dashed $\mathcal{M} = 4.47$; and dotted $\mathcal{M} = 44.7$. Subsonic runs with $\beta \geq 100$ are plotted with their second steady state value. The dashed lines connect these points to their initial steady state values (no data point shown).

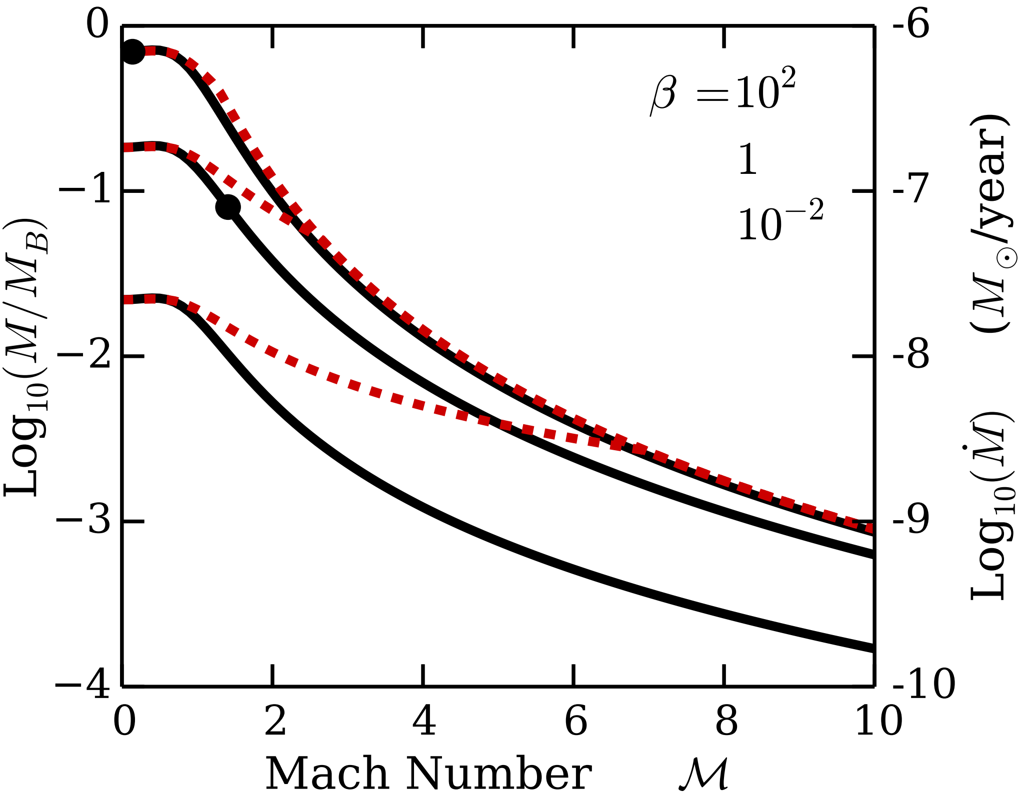

Figure 4.11: Mass accretion rate as a function of sonic Mach number (x-axis) and plasma β. From top to bottom, the curves represent decreasing β values. The solid curves show the parallel fit while the dashed curves show the perpendicular fit (Equations 4.27 and 4.30). The right y-axis uses the fiducial parameters given in Equation (4.3) $M_* = 0.4\,M_\odot$ yr^{-1}, $n_0 = 10^4$ cm^{-3} and $T = 10$ K. The points identify where $\mathcal{M}_\mathrm{A} = 1$. The fits are identical at low Mach numbers, and the perpendicular rate always equals or exceeds the parallel rate. Once $\mathcal{M}_\mathrm{BH,eff} > \mathcal{M}_\mathrm{ABH,eff}/\mathcal{M}_\mathrm{BH,eff}$ (Equation 4.30), the perpendicular fit becomes identical to the hydrodynamic fit, which is well-approximated by the $\beta = 100$ curves.

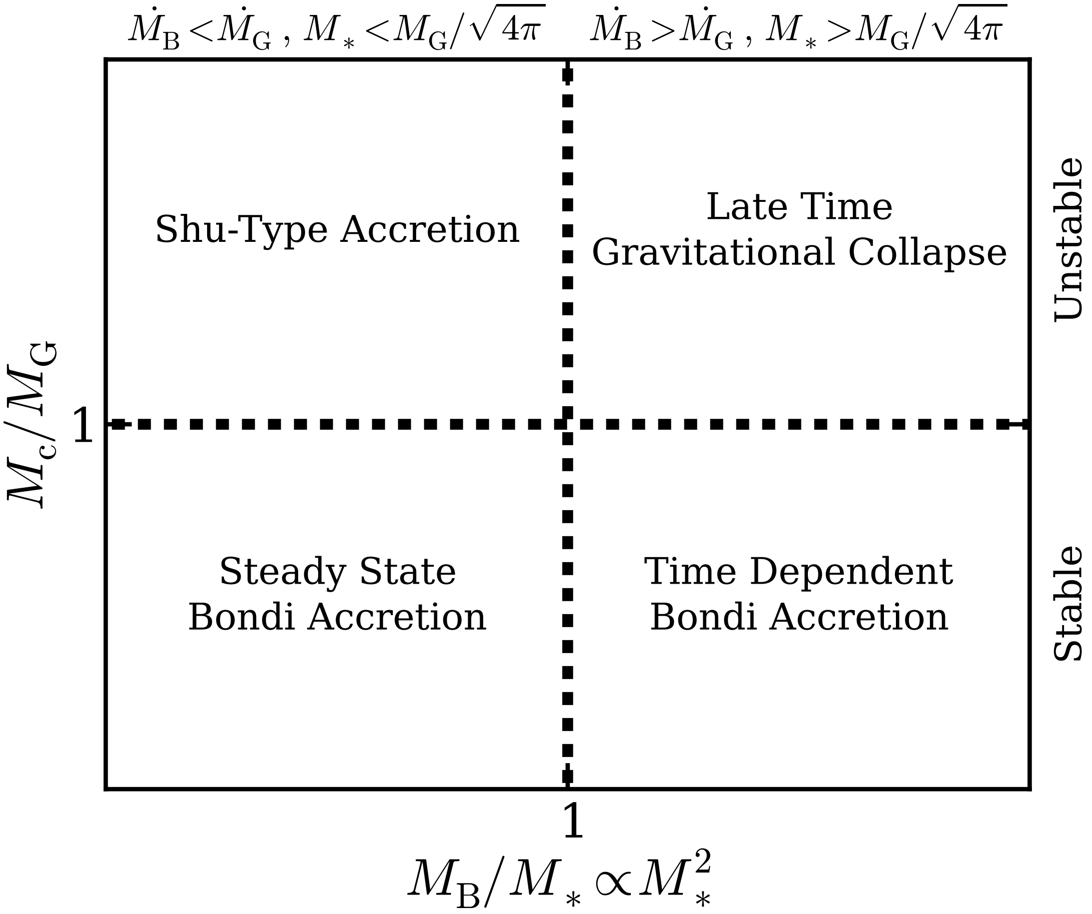

Figure 4.12: Method of accretion as a function of the stellar mass M_* and the host core's mass $M_{\rm c}$. The core is stable against its own self gravity if its mass is less than $M_{\rm G}$ (Equation 4.46). The gas interior to radius $r_{\rm B}$ is not self gravitating if this gas mass $M_{\rm B}$ is less than M_* (or equivalently, $M_* < M_{\rm G}/\sqrt{4\pi}$). If both of these relations are satisfied, steady-state Bondi accretion occurs. Once the stellar mass grows so that $M_{\rm B} \gtrsim M_*$, the accreted mass is comparable to M_* and this accretion rate becomes time dependent. Regardless, if the core is collapsing via its own self-gravity, accretion occurs by gravitational collapse and a steady state is never realized.

Chapter 5

Radiative Feedback in the Formation of Population III Stars

[1] The formation of the Universe's first stars initiated the transition from a pristine dark Universe to a complex, metal-rich Universe filled with stars. The exact role these first stars played in reionization, metal-enrichment, and the formation of subsequent generations of stars depends on their initial mass function, which is likely regulated by the stars' own UV feedback. We have carried out a numerical study on the effect of stellar feedback and the development of HII regions around the first stars. Our simulations draws initial conditions from the spherical models of Tan & McKee (2004) and McKee & Tan (2008) utilizes the adaptive mesh refinement code ORION2. We track the ionizing radiation from a forming protostar using a newly implemented ray tracing routine to study the development of an HII region around that protostar. Once the star exceeds $\sim 300 M_{\odot}$, the radius of the HII region exceeds the star's gravitational radius $GM/c_s^2 \sim 1.5 \times 10^3$ AU, where the sound speed of the ionized gas is comparable to the free-fall velocity. This allows the heated gas to escape from the star's gravitational pull, and the accretion onto the star dramatically reduces.

5.1 Introduction

Our Universe today bears little resemblance to the Universe at early times. At the time of recombination and the release of the cosmic microwave background (at redshift $z \sim 1100$), the structure of the Universe was uniform and devoid of stars, galaxies, and any elements beyond Lithium. Density perturbations grew and dark matter began to amass in some of the first structures in the Universe. Around $z \sim 20 - 30$, the formation of $\sim 10^6\, M_{\odot}$ dark matter minihalos along the intersection of filaments and sheets created reservoirs where baryonic matter collected (e.g., Abel et al. 2002; Yoshida et al. 2003). These $\sim 10^3\, M_{\odot}$ gas

---

[1] Large portions of this chapter will eventually be published as Lee, A. T., McKee, C. F., Stacy, A. R., Rosen, A.R., Klein, R.I. Radiative Feedback in the Formation of Population III Stars I. HII Regions. in the Monthly Notices of the Royal Astronomical Society. Appropriate permissions to reproduce it here have been obtained from the co-authors.

clouds (e.g., Bromm et al. 2002) are where it is believed that enough hydrogen amassed to develop a sufficient abundance of H_2 to cool the gas through rovibrational transitions and enable collapse (e.g., Haiman et al. 1997; Tegmark et al. 1997; Yoshida et al. 2003). This collapse commenced the formation of the first metal-free primordial stars (e.g., Omukai & Nishi 1998a), the so-called Population III (Pop III) stars.

Since Pop III stars' initial rays of light, cosmic evolution has largely been driven by stars. The stellar winds and supernova explosions of the first generations of stars enriched the Universe with elements beyond lithium (e.g., Mori et al. 2002; Karlsson et al. 2013). Furthermore, the first stars and their potentially violent finales commenced the reionization of the developing intergalactic medium (e.g., Kitayama et al. 2004; Sokasian et al. 2004; Whalen et al. 2004; Johnson et al. 2007a). How much these Pop III stars affected their surroundings, either through their radiative or mechanical output, depends on their initial mass function (IMF). The Pop III IMF is believed to be top-heavy, dominated by stars that are tens to hundreds of solar masses (Abel et al. 2002; Dopcke et al. 2013), compared to the present-day IMF that is dominated by low-mass stars (Kroupa 2002; Chabrier 2003). Simulations have shown that these first stars can have a variety of masses, ranging from Sun-like to hundreds of solar masses (Clark et al. 2011; Stacy et al. 2012), though simulations have yet to converge on the exact form for the IMF.

The IMF of Pop III stars informs the role these stars played in reionization and the evolution of the early Universe. Just as in the metal-enriched case, the luminosity of a Pop III star scales super-linearly with mass, and super-luminous stars can wreak havoc on their surroundings. For example, massive Pop III stars emit copious Lyman-Werner radiation between 11.2 eV and 13.6 eV, which can dissociate molecular hydrogen (Bromm & Yoshida 2011). This presents an extra challenge to subsequent generations of star formation, since H_2 is an important coolant in these high-redshift environments (Haiman et al. 2000; Glover & Brand 2001; Johnson et al. 2007b). Indeed, simulations have shown that the Lyman-Werner radiation from a single massive Population III star is sufficient to prevent further cooling and star formation in its natal halo (Omukai & Nishi 1998a). Though these first stars unlikely played a dominant role in directly reionizing the developing IGM, they likely impacted subsequent star formation. Fully understanding the Pop III IMF will inform whether subsequent star formation contributed to reionization, or whether the developing first galaxies (and their central black holes) played a larger role (Greif & Bromm 2006).

These stars can also undergo a variety of stellar deaths. For non-rotating stars, those in the ranges $40 - 140\ M_\odot$ or $> 260 M_\odot$ likely collapse directly into black holes, dispersing, at most, a tiny amount of metal-enriched mass into its surroundings, but will continue to irradiate in X-rays and above as they accrete mass (Jeon et al. 2012; Hummel et al. 2015). With rotation, these black holes may also become the first collapsar gamma-ray bursts (Woosley 1993; Stacy et al. 2011, 2013) observable with current and planned observatories (Salvaterra 2015; Ghirlanda et al. 2015). Stars falling in the intermediate range $140 - 260\ M_\odot$ end their lives as pair-instability supernovae, events that disrupt the entire star and leave behind no compact remnant (Heger et al. 2003). Stars between $8 M_\odot$ and $140 M_\odot$ end their lives as traditional Type II supernovae. While individual Pop III stars are too faint to be detectable by next-generation telescopes like the *James Webb Space Telescope* (Gardner et al. 2006), Pop III stars of lower mass ($< 0.8 M_\odot$) could survive to the present day, allowing for direct detection in the Milky Way through, for example, *Gaia* (Perryman et al. 2001) or *SkyMapper* (Keller et al. 2007). These direct detections, or the lack thereof, could place strong constraints on the lower mass range of Population III stars (Hartwig et al. 2015).

Previous theoretical work has found that typical stellar masses of Pop III stars are smaller than that of the $\sim 10^3\ M_\odot$ host cloud primarily due to the stellar radiative feedback emitted by these first stars (e.g., Omukai & Inutsuka 2002; McKee & Tan 2008; Stacy et al. 2016). As stellar mass increases, bipolar HII regions develop above and below the accretion disk of these first stars, halting accretion in these regions. The disk eventually photo-evaporates from the increasing stellar radiation, setting the final mass of the star (McKee & Tan 2008). Simulations have begun to bear out this picture. The two-dimensional radiative-hydrodynamic simulations of Hosokawa et al. (2011), Hosokawa et al. (2012b), and Hirano et al. (2014) show UV radiation playing a primary role in setting a star's final mass, though fails to capture the true dynamical nature of the three-dimensional accretion process. This work has since been extended by the three-dimemsional non-axisymmetric simulations of Stacy et al. (2012), Susa et al. (2014), Stacy et al. (2016), and Hosokawa et al. (2016), for example, which have continued to support this picture that stellar feedback plays an important role in setting the star's final mass.

Initial disagreements between simulations resulted primarily from different combinations of physical processes used. As our numerical capabilities have advanced, these simulations are converging on the physics involved in the Pop III star formation process, differing primarily in their implementations and the initial conditions used. A substantial numerical challenge in all of these simulations is the implementation of the direct UV component from growing stellar sources. The stellar feedback component dramatically affects the hydrodynamics and chemistry of the surrounding gas, so this radiation must be tracked both in space and in time. This is typically done through ray tracing approaches. Early simulations that incorporated the direct UV component from stellar sources were predominately two-dimensional (e.g., Hosokawa et al. 2011) to reduce the computational demand, whereas three-dimensional simulations have used coarse ray tracing schemes (Greif et al. 2009; Stacy et al. 2016).

While simulations support the basic picture of Pop III formation above, simulations are not yet in agreement. This picture may also be too simplistic: stars do not form in isolation. Fragmentation of the host cloud and gravitationally unstable accretion disks may form multiple objects, which can form higher-order systems even before stellar feedback becomes important, as is found in Stacy & Bromm (2013), or these higher-order systems can ultimately disrupt or never form in the first place, as is found in Hosokawa et al. (2016). These simulations typically considered only one or a few sets of initial conditions. Hirano et al. (2014) considered instead 110 star forming regions, and found that the true final mass of the star likely depends on a combination of feedback effects and the dynamics of the gas within the halo (Hirano et al. 2014). The final masses of Pop III stars from these simulations span from $\sim M_\odot$ to $\sim 10^3\ M_\odot$. Another challenge in these simulations is achieving the resolution necessary to capture the small-scale dynamics of the accretion disks. The best simulations resolve down to $\sim$ 10 AU, which makes resolving the vertical structure of the accretion disk ($\sim$ AU) as well as small-scale fragmentation within the accretion disk a challenge. Treatment of the hydrodynamics on these small scales become important when modeling how mass gets transported within the accretion disk onto the surface of the radiating star. Episodic accretion from the disk puffs up the radius of the star, which additionally lowers its surface temperature. The ultimate reduction of UV feedback may allow the central star to grow beyond the UV-regulated limits suggested by earlier studies (Hosokawa et al. 2016).

Recently, we implemented a new ray tracing in our ORION2 gravity-hydrodynamics code. This scheme achieves parallel scaling out to thousands of processors, a major advance over all the existing ray tracing schemes to date (Rosen et al. 2017), and gives us, for the first time, the capabilities of running simulations that incorporate a high resolution direct radiation component. In this final chapter, we present the preliminary work for such three-dimensional simulations. In Appendix C, we discuss our implementation of hydrogen-ionizing physics in the ray trace routine and show several tests that demonstrate excellent coupling with ORION2's hydrodynamics physics module. The simulations reported here follow the growth of a Pop III star until radiative feedback from the massive star develops an expanding HII region. We employ ORION2's sink particle routines to represent the growing protostar. The simulations of this chapter draw initial conditions from the spherical models of Tan & McKee (2004) and McKee & Tan (2008). Particularly, these simulations allow us to test the coupling of the ray trace with the gravity-hydrodynamic physics already present in ORION2. From here, future work will incorporate more realistic initial conditions that are drawn from the cosmological simulations of Stacy & Bromm (2013).

This chapter is organized as follows. Section 5.2 describes the numerical methodology of our star formation code ORION2. Section 5.3 discusses the results of these simulations. Section 5.4 summarizes and discusses the results.

5.2 Numerical Methods

5.2.1 Overall Methodology

In this study of the growth and evolution of the first stars, we use the ORION2 adaptive mesh refinement (AMR) code (Li et al. 2012). The physics modules used in these simulations include self-gravity (Martin et al. 2008), hydrodynamics (Li et al. 2012), gravitating and accreting sink particles (Krumholz et al. 2004a; Myers et al. 2013a; Lee et al. 2014), and protostellar feedback due to protostellar luminosity (Offner et al. 2009). In addition, ORION2 now has a highly-parallelized ray tracing routine to model the ionizing radiation emitted from forming protostars (Appendix C, Rosen et al. 2017). ORION2 uses the Chombo library for AMR (Applied Numerical Algorithms Group & LBNL Computational Research Division 2012) and an extended version of the Constrained Transport scheme from PLUTO (Mignone et al. 2012) to evolve the system. Each AMR level is evolved through discrete time steps Δt_l, which differ for each level l with the property $\Delta t_{l+1} = 0.5\Delta t_l$. We define the one-dimensional cell size of a cell on level l as Δx_l, which relates to Δx_{l+1} by the same 0.5 factor. Details regarding the implementation of these modules in ORION2 can be found in the references. Below we highlight any parameter choices and assumptions we made and address aspects of the modules that will be specific to this work.

ORION2 uses a Cartesian adaptive grid where every cell has a set of conserved quantities: the mass density ρ, the momentum density $\rho\mathbf{v}$ of gas moving at velocity $\mathbf{v}$, the specific internal energy e, and a set of tracer fields. For this work, we trace the mass density of neutral atomic hydrogen ρ_H. Future work will include our recently completed primordial chemistry package based off the packages of Gray & Scannapieco (2010) and Glover & Abel (2008), which will trace and evolve these species along with twelve other relevant atoms, molecules, and ions of hydrogen and helium to accurately calculate the proper heating and cooling rates in primordial environments.

Independent of the grid, ORION2 includes Lagrangian sink particles to model regions of gas that have collapsed beyond the resolution of the grid. These sink particles represent the growing protostars and can interact with the gas through gravity and radiation. The characteristics of the protostar (such as its luminosity) are determined by a protostellar evolution model, described below, which is based off the model of Stacy et al. (2016). Sink particles accrete mass, momentum, and energy from a sphere centered on the sink particle that extends four computational zones in radius (Krumholz et al. 2004a).

The equations governing the evolution of the gas in ORION2 are

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(\rho\mathbf{v}) - \sum_p \dot{M}_p W(\mathbf{r}_p)\ , \tag{5.1}$$

$$\frac{\partial(\rho\mathbf{v})}{\partial t} = -\nabla\cdot(\rho\mathbf{v}\mathbf{v}) - \nabla P - \rho\nabla(\psi+\psi_p) - \sum_p[\dot{\mathbf{p}}_p W(\mathbf{r}_p)]\ , \tag{5.2}$$

$$\frac{\partial(\rho e)}{\partial t} = -\nabla\cdot[(\rho e + P)\mathbf{v}] - \rho\mathbf{v}\cdot\nabla(\psi+\psi_p) + \sum_p[\dot{E}_{rad,p} - \dot{\epsilon}_p W(\mathbf{r}_p)]\ , \tag{5.3}$$

$$\frac{\partial\rho_t}{\partial t} = -\nabla\cdot(\rho_t\mathbf{v}) + \sum_p \dot{C}_{rad,p}(\rho_t, e) - \dot{M}_{p,t}W(\mathbf{r}_p)\ , \tag{5.4}$$

$$\nabla^2\psi = 4\pi G\rho\ , \tag{5.5}$$

$$P = \frac{\rho k_{\mathrm{B}} T}{\mu m_{\mathrm{H}}} = (\gamma-1)\rho e\ . \tag{5.6}$$

These equations are the standard equations of mass, momentum, and energy conservation, a continuity equation for the mass density of each tracer ρ_t, and Poisson's equation for the gravitational potential ψ of the gas. The equations are closed with the ideal gas law for the pressure P. The summations over the particles p incorporate how the sink particles interact with the gas. The gravitational potential of the particles evaluated at a cell center is the sum

$$\psi_p = -\sum_p \frac{GM_p}{|\mathbf{r}_p|} \ , \tag{5.7}$$

where $\mathbf{r}_p$ is the vector connecting the sink particle and a cell center. The distribution of how much mass, momentum, and energy that is removed from the accretion zone surrounding the sink particle is described by a weighting kernel

$$W(\mathbf{r}_p) \propto \exp\left(-(|\mathbf{r}_p|/\Delta x_{l_{\max}})^2\right) \ , \tag{5.8}$$

which has units of 1/volume and is zero for $|\mathbf{r}_p| > 4\Delta x_{l_{\max}}$. Accelerations from the particles are softened by increasing the distance one finest cell length. Mass accretion also removes mass from the tracer fields but preserves the relative mass fractions of the gas. The mass accreted also imparts momentum and energy to the particle at rates $\dot{\mathbf{p}}_p$ and $\dot{E}_p$. For this work, we will also track the ionizing radiation from stellar sources; this feedback will impart energy into the gas, as well as change its composition, at volumetric rates $\dot{E}_{rad,p}$ and $\dot{C}_{rad,p}(\rho_t, e)$, respectively. How these terms are evaluated is described in Appendix C. The ideal gas law relates the pressure to the temperature T and the mean molecular weight μ of the gas. The remaining symbols are standard constants: $k_{\rm B}$ is Boltzmann's constant and $m_{\rm H}$ is the mass of a neutral hydrogen atom. The second equality in Equation (5.6) is the thermodynamic relation between pressure and the internal energy. This expression incorporates the adiabatic index γ, defined as the ratio of the heat capacities at constant pressure and volume.

Furthermore, the sink particles are updated via the N-body equations

$$\frac{dM_p}{dt} = \sum_{\rm cells} \dot{M}_p \ , \tag{5.9}$$

$$\frac{d\mathbf{x}_p}{dt} = \mathbf{v}_p \ , \tag{5.10}$$

$$\frac{d\mathbf{v}_p}{dt} = -\nabla\psi - \sum_{q\neq p} \frac{GM_q}{|\mathbf{r}_{pq}|^2 + \Delta x_{l_{\max}}^2}\hat{\mathbf{r}}_{pq} + \sum_{\text{cells } i} \frac{\dot{\mathbf{p}}_i}{M_p} \ , \tag{5.11}$$

where the 'cell' sum is done over the accretion zone of particle p. The distance vector between two sink particles is $\mathbf{r}_{pq}$, which defines the unit vector $\hat{\mathbf{r}}_{pq} = \mathbf{r}_{pq}/|\mathbf{r}_{pq}|$.

5.2.2 Chemistry

Above $\sim 10^2$ Kelvin, rotational and vibrational modes of H_2 can be excited, which can lower the cell's value of γ below 5/3. Figure 5.1 shows γ for H_2, assuming an ortho to para ratio of 3:1, which is appropriate for primordial gas (Glover 2013). A proper calculation of a cell's adiabatic index is the sum over all species

$$\frac{1}{\gamma - 1} = \sum_{\rm species} \frac{n_i/n_t}{\gamma_i(T) - 1} \ , \tag{5.12}$$

where n_i is the number density of the species, n_t is the total number density, and $\gamma_i(T)$ is the adiabatic index of species i. If we consider only atomic and molecular hydrogen, and helium, only the molecular hydrogen γ depends on temperature; the remaining $\gamma_i = 5/3$. Since the γ value for molecular hydrogen depends on the temperature and temperature is not a quantity tracked in the above hydrodynamic equations, Equation (5.12) is simultaneously solved with

$$e = \frac{1}{\gamma - 1} \frac{\rho k_{\rm B} T}{\mu m_{\rm H}} \tag{5.13}$$

to obtain γ (and T, if needed) for each cell.

Beyond γ, chemistry plays a vital role in triggering primordial star formation. Primordial gas proceeds to collapse because of the eventual formation and cooling of molecular hydrogen. H_2 is unable to form via dust grains as it is believed to form in the contemporary universe (Gould & Salpeter 1963). Furthermore, the heating and cooling rates vary considerably from local star forming environments. The proper calculation of μ, γ, and these heating and cooling rates would involve solving a chemical network for the pristine gas found in Pop III environments. The relevant species include three states of atomic hydrogen (H, H^+, H^-), three states of atomic deuterium (D, D^+, D^-), three states of atomic helium (He, He^+, He^{++}), two states of molecular hydrogen (H_2, H_2^+), two states of molecular deuteride (HD, HD^+), and free electrons (e^-) (Glover & Abel 2008). Here superscripts give the charge of each species, with $+$ representing ions, and $-$ representing anions. This network would include relevant formation scenarios for molecular hydrogen, including the three-body reaction that occurs above number densities of $n_{\rm H} \sim 10^8$ cm^{-3}:

$$\mathrm{H + H + H \rightarrow \ H_2 + H} \ , \tag{5.14}$$

with a smaller contribution of H_2 coming from reactions involving already present H_2 molecules and helium atoms (Palla et al. 1983).[2] The increasing fraction of H_2 is accompanied with increased H_2 line cooling, resulting in a slight drop in temperature as the H_2 fraction approaches unity.

In this preliminary study, we will approximate the details above. Omukai & Nishi (1998b) and Ripamonti et al. (2002) considered one-dimensional calculations of accretion onto a forming Pop III protostar and found that the accreting gas can be accurately treated as isentropic with an adiabatic index $\gamma \approx 1.1$ due to H_2 cooling; i.e., each mass element satisfies the relation $P = K\rho^{1.1}$ with a constant entropy parameter K. We will adapt this value for γ globally; future work will incorporate a primordial chemistry package that will allow γ to vary both spatially and temporally.

---

[2] The rate of this crucial three-body reaction has been debated in the literature, with rate coefficients varying by orders of magnitude (Turk et al. 2011a) prior to the recent quantum mechanical calculations of Forrey (2013).

5.2.3 Primordial Stellar Evolution Model

Once sink particles form, an internal model is used for gas that has been accreted. This model incorporates the protostellar mass M_*, the history of the accretion rate $\dot{M}_*$ and an approximate set of equations for the stellar structure to determine quantities like the protostellar radius R_*, the effective temperature T_*, the luminosity L_*, and the number flux of photons with energies above 13.6 eV ($\dot{N}_{\rm ion}$). Our protostellar evolution model follows that of Stacy et al. (2014, 2016), which is derived by fitting to the stellar evolution models of Stahler et al. (1986b), Stahler et al. (1986a), Omukai & Palla (2003), Hosokawa et al. (2010b), and Hosokawa et al. (2012a). These evolution models particularly determine the evolution of the Pop III star as it contracts toward the main sequence. In this chapter we will be primarily focused on stars that exceed $\sim 100 M_\odot$, so the exact details of what follows will not impact the results of this chapter. Nonetheless, we recap the model here, since it will play an important role in future studies.

The stellar luminosity is determined by two components, the internal luminosity $L_{\rm int}$ and the accretion luminosity $L_{\rm acc}$:

$$L_* = L_{\rm acc} + L_{\rm int} = \alpha \frac{G M_* \dot{M}_*}{R_*} + L_{\rm int} \ . \tag{5.15}$$

To smooth the episodic accretion rate expected from accreting from a discretized grid, we smooth $\dot{M}_*$ over the last 32 finest time steps. Here α parameterizes the fraction of the accretion energy that is radiated from the star. The gravitational energy of the infalling gas can either be radiated away, advected into the interior of the star, or, in the case where magnetic fields are present, be extracted through magnetic forces. In an extreme limit, the gas can gently settle into the atmosphere of the star, losing all its excess energy so that the infalling gas's thermal state matches that of the stellar atmosphere. In reality, the infalling gas advects additional energy into the interior of the star and $\alpha < 1$.

For this particular simulation, the exact value of α does not impact the results: after $M_* \gtrsim 100 M_\odot$, the internal luminosity $L_{\rm int}$ will dominate the total luminosity output. The value of α can matter at earlier times where the advected energy can result in inflation of the stellar radius (Hosokawa et al. 2013). Nonetheless, we adopt $\alpha = 0.99$, following estimates from Hosokawa et al. (2016).

The Pop III stellar model begins on the Hayashi track of the Hertzsprung-Russell diagram. These tracks are described by a nearly constant effective temperature, which is ∼3000 to 5000 Kelvin. The lower opacities of Pop III stars result in higher effective temperatures compared to their contemporary counterparts. We choose an initial effective temperature of 4500 Kelvin; the resulting luminosity is then

$$L_{\rm Hay} = 4\pi R_*^2 \sigma_{\rm SB} (4500 \ {\rm K})^4. \tag{5.16}$$

As the stellar mass increases, nuclear reactions eventually commence in the core and the star transitions to the Henyey track, characterized by a nearly constant luminosity. Fitting to the $10 M_\odot$ zero-metallicity track yields

$$L_{\rm Hen} = 10^{3.5} L_\odot \left(\frac{M_*}{9 M_\odot}\right)^{22/5} \left(\frac{T_*}{10^4\ {\rm K}}\right)^{4/5} \tag{5.17}$$

(Henyey et al. 1955). We use the Hayashi luminosity for the sink particle until its magnitude is exceeded by the Henyey luminosity. At that point, the Henyey luminosity is used. Once the radius of the protostar in the stellar model approaches the zero-age main sequence (ZAMS) luminosity, we switch over to the ZAMS luminosity:

$$L_{\rm ZAMS} = 140 L_\odot \left(\frac{M_*}{M_\odot}\right)^2 \tag{5.18}$$

(Hosokawa et al. 2010b). We note that since the radius of the protostar can fluctuate from rapid and episodic accretion, this model allows the star to enter and leave the ZAMS.

The radial evolution of the star also proceeds in two phases, the adiabatic accretion phase and the Kelvin-Helmholtz (KH) contraction phase (Stahler et al. 1986a; Hosokawa et al. 2010b). In which phase the protostar resides depends on how well the protostar is able to radiate away energy. When the accretion time scale $t_{\rm acc} \sim M_*/\dot{M}_*$ is shorter than the KH timescale $t_{\rm KH} \sim GM_*^2/R_* L_*$, energy is inefficiently radiated away from the surface. In this phase, the radius R_* gradually increases with mass. In the opposite regime, the protostar contracts until it reaches the zero-age main sequence (ZAMS).

We can derive an approximate stellar evolution model for the stellar radius using the models of Hosokawa et al. (2010b) and Hosokawa et al. (2012a). The simplest models assume a constant $\dot{M}_*$. These papers solved for the radial structure of the star, and included the effects of deuterium burning even before hydrogen burning commences. They considered two accretion geometries, namely, accretion via an accretion disk and spherical accretion. The former can be viewed as the gentler of the two, where the accreted gas impacts a small fraction of the stellar surface. The latter, on the other hand, can create an accretion shock over a substantial fraction of the stellar atmosphere. In addition, Hosokawa et al. (2010b, 2012a) found that their protostellar radii undergo a swelling phase near the end of the adiabatic accretion phase (see, e.g., their Figure 2 for the spherical accretion geometry or their Figure 4 for disk-like accretion in Hosokawa et al. 2010b). In this stage, the rapidly increasing luminosity output from the center of the star creates what they call a "luminosity wave" that transports entropy outward toward the surface. As it moves toward the surface, the star continues its gradual expansion, which lowers the opacity of the star as it grows. The surface layers of the star eventually absorb part of this luminosity wave and bloat to larger radii. Our model below will not capture this swelling phase that divides the adiabatic accretion and KH contraction phases. However, we note that this swelling phase occurs before the ionizing flux of the star becomes substantial ($M \leq 10 M_\odot$), so this phase will not impact our initial results here.

The details of the protostar's radial evolution depend on the accretion geometry. For purely spherical accretion, the radius during the adiabatic phase can be approximated as

$$R_{\rm ad,sph} = 49 R_\odot \left(\frac{M_*}{M_\odot}\right)^{1/3} \left(\frac{\dot{M}_*}{\dot{M}_{\rm fid}}\right)^{1/3} , \tag{5.19}$$

where $\dot{M}_{\rm fid} = 4.4 \times 10^{-3}\ M_\odot/{\rm yr}$ (Omukai & Palla 2003; Hosokawa et al. 2010b; Stacy et al. 2012). The Kelvin-Helmholtz phase is given by

$$R_{\rm KH,sph} = 141 R_\odot \left(\frac{M_*}{10 M_\odot}\right)^{-2} \left(\frac{\dot{M}_*}{\dot{M}_{\rm fid}}\right) . \tag{5.20}$$

Around $M_* \sim 10 M_\odot$, Kelvin-Helmholtz contraction begins. For purely spherical accretion with a constant mass accretion rate–ignoring the swelling phase for now–setting the radius equal to the minimum of these two quantities at any given time generally reproduces this result (Stacy et al. 2016). These approximations are also consistent with the 2D semi-analytic model of Tan & McKee (2004).

If the accretion geometry is disk-like, the protostar evolves differently than the spherical case just described. With a smaller energy flux compared to spherical accretion, the protostar contracts sooner when accreting through a disk. Hosokawa et al. (2010b) found that the radius for spherical accretion models are at most three times that of disk models. For the disk values of $R_{\rm ad,dk}$ and $R_{\rm KH,dk}$, we set them as one-third the values of the spherical case (Stacy et al. 2016).

Finally, the ZAMS radius is given as

$$R_{\rm ZAMS} = 0.28 R_\odot \left(\frac{M_*}{M_\odot}\right)^{0.61} . \tag{5.21}$$

We note from Hosokawa et al. (2010b) that this radius is smaller for Pop III stars compared to stars of the same mass with metallicity. The lack of C, N, and O elements delay the onset of the CNO-cycle until these elements are generated through triple-α burning, and the proton-proton nuclear reaction chain is unable to halt contraction until these stars have contracted to lower radii.

Following Stacy et al. (2016), we use our choice of α to delineate between these spherical and disk-like accretion tracks. During every update of the star's radius, we set the radius as

$$R_* = (1-\alpha)\min(R_{\rm ad,sph}, R_{\rm KH,sph}) + \alpha \min(R_{\rm ad,dk}, R_{\rm KH,dk}) . \tag{5.22}$$

If this radius is less than $R_{\rm ZAMS}$, we reset the radius as the ZAMS radius. Equation (5.22) is a very rough approximation that captures the fact that more energy is advected into the stellar interior in spherical accretion than in disk accretion. As noted above, the exact value of α is not important for us since we are focusing on the stage where the star is massive enough to be on the main sequence.

When the accretion rates are low, the above expressions can lead to rapidly decreasing radii, which can become unphysically low, especially at early times. Indeed, this would require the protostars to shrink at rates faster than the Kelvin-Helmholtz rate. If we evaluate $\dot{R}_*$ and its magnitude exceeds $R_*/t_{\rm KH}$, where the previous value of R_* is used in this criterion, we restrict $\dot{R}_*$ to be exactly $-R_*/t_{\rm KH}$. This allows our protostars to smoothly decline in radius, but still contract rapidly enough to high effective temperatures once the star has grown beyond $\sim 10 M_\odot$.

Once the new radius is determined, it is used to calculate a new stellar luminosity. If the radius is within 5% of the ZAMS radius, we set the luminosity to $L_{\rm ZAMS}$. Else we set the luminosity equal to the maximum between $L_{\rm Hen}$ and $L_{\rm Hay}$. The previous value of L_* is used to estimate T_* for the Henyey luminosity. The exception to this luminosity calculation is if the luminosity exceeds the Eddington rate

$$L_{\rm edd} = 3.8 \times 10^6 L_\odot \left(\frac{M_*}{100 M_\odot} \right) \ . \tag{5.23}$$

at which point we cap the luminosity to $L_{\rm edd}$. Again, this limit can occur at early times when the accretion rate is large (Hosokawa et al. 2012a), and at later times when the star grows beyond $\sim 100 M_\odot$ in mass.

Figure 5.2 gives an example radius evolution from our stellar evolution model when the mass accretion rate is fixed at $\dot{M} = 10^{-3}\ M_\odot/{\rm yr}$. The radius initially grows via the adiabatic phase, eventually transitioning to KH-contraction toward the main sequence.

5.2.4 Ray Tracing

To model the radiative feedback from the massive protostar in our simulation, we employ ORION2's ray tracing module (Rosen et al. 2017). Particularly, we use this module to trace the ionizing radiation emitted from stars; both this component of the ray tracing routine, as well as how the rays interact with the cells, are described in Appendix C. For this simulation, we do not include the radiation pressure imparted from the ionizing radiation and therefore the ray tracing module only changes the composition and internal energy of the gas.

The primary role of the protostellar evolution model is to calculate the number of ionizing photons emitted from the star. Given L_* and R_*, they give T_*, which can be used to compute the integral

$$\dot{N}_{\rm ion} = 4\pi R_*^2 \int_{\nu_0}^{\infty} \frac{\pi B_\nu(T_*)}{h\nu} d\nu$$

where $\nu_0 = 13.6\ {\rm eV}/h$ is the minimum frequency to ionize hydrogen. The logarithm of this integral is computed as a table and linear interpolation is done to quickly calculate the number of ionizing photons coming from a given star. Figure 5.2 also shows $\dot{N}_{\rm ion}$ for the constant $\dot{M}_*$ example.

5.2.5 Initial Conditions

We draw our initial conditions from the spherical models of Tan & McKee (2004) and McKee & Tan (2008), which studied the gravitational collapse of a virialized cloud. For simplicity, the cloud is initialized without rotation, therefore we do not expect an accretion disk to form in this simulation. To incorporate the effects of H_2 cooling without directly incorporating chemistry, we follow the approximation of these papers, Omukai & Nishi (1998a), and Ripamonti et al. (2002), who found these effects can be modeled by assuming a global adiabatic index of $\gamma \approx 1.1$. This is incorporated by assuming a polytropic relationship between the pressure and density: $P = K\rho^{1.1}$ for a given entropy parameter K. With these assumptions, Tan & McKee (2004) derived the hydrostatic equilibrium solution for the total mass interior to radius r (their Equation 14):

$$M_r = 980 \left(\frac{r}{\mathrm{pc}}\right)^{7/9} K'^{10/9}\, M_\odot \ , \tag{5.24}$$

where K' is the normalized entropy parameter $K' = K/(1.88 \times 10^{12}\ \mathrm{cgs})$. Equivalently, the density at a given radius is given by $\rho = (7/36\pi)M_r/r^3$. Without radiative feedback from the central protostar, they derive a mass accretion rate of

$$\dot{M}_* = 0.026 K'^{15/7} \left(\frac{M_*}{M_\odot}\right)^{-3/7} M_\odot/\mathrm{yr} \ , \tag{5.25}$$

(Equation 8 of Tan & McKee 2004). The mass accretion rate scales as $t^{-3/10}$, and the time required to build a star to a given stellar mass is

$$t_* = 27 K'^{-15/7} \left(\frac{M_*}{M_\odot}\right)^{10/7} \mathrm{yr}. \tag{5.26}$$

These equations will serve as our basis for comparison when radiative feedback is included. For this simulation, we set $K' = 1$. Figure 5.3 shows the initial radial profile of ρ. For the ray tracing routine, we assume a chemical composition that is initially neutral hydrogen and helium with mass fractions $X = 0.76$ and $Y = 0.24$, respectively. This gives a mean molecular weight of $\mu = 1.22$.

We employ inflow/outflow boundary conditions in our simulation. The finite size of the computational domain and this choice of boundary conditions creates two artifacts that must be addressed. First, in our case where gas at the boundaries is generally moving toward the center of the box, inflow/outflow boundary conditions simply copy the state quantities of the gas in the boundary cells to generate the gas that flows into the domain. This inflow may not be representative of the actual inflowing gas where we to initialize a larger volume of the halo. Second, the Poisson solver treats the edge of the domain as an equipotential surface, which would not be the case if the box were extended to include more of the halo.

Our simulations will run until the star reaches $\sim 300 M_\odot$ in size, which occurs in $\sim 10^5$ years (McKee & Tan 2008). Considering the inflowing speed of the gas, we use a box that is $L_{\rm box} = 0.78$ parsecs along each dimension. The gas that forms the star comes primarily from the central 50% of the box, and this also ensures that boundary conditions for the hydrodynamics and self-gravity will not significantly influence our results.

5.2.6 Grid Parameters and Refinement Criteria

The power of AMR codes comes from their ability to refine only the areas of interest. This is particularly useful in simulations of star formation, which have large dynamic ranges but which only a small volume requires high resolution to accurately track the collapse of gas into stars. During the simulation, grids can be added and removed based on criteria set by the user.

Our simulations uses a base grid (level 0) of 128^3 and allows up to $l_{\rm max} = 5$ levels of refinement, giving an effective resolution of 40 AU. Atop the base grid, we statically refine the region around the sink particle that forms in the center of the domain. All level 0 cells within a cube centered on the protostar that spans $\pm(1/2^4)L_{\rm box}$ in each direction are refined to level 1. Similarly, we refine all level 1 cells to level 2 if they reside in a similar cube that spans $\pm(1/2^5)L_{\rm box}$ in each direction. This process is repeated all the way to the maximum level. These choices allow the central ~ 1000 AU to be located on the finest two levels. The refined grid is redrawn every two coarse time steps by recursively calling the above criterion for each level.

Finally, sink particles are created on the finest level if the density of a cell exceeds the Truelove-Jeans density:

$$\rho_{\rm TJ}(J = 1/4, l = l_{\rm max}) = \frac{\pi J^2 c_s^2}{G\Delta x_l^2} = \frac{\pi c_s^2}{16 G \Delta x_{l_{\rm max}}^2} \tag{5.27}$$

(Krumholz et al. 2004a). This criterion creates sink particles when the Jeans length ($\sim \sqrt{c_s/G\rho}$) is no longer resolved by four grid cells, which has been shown to prevent artificial fragmentation (Truelove et al. 1997). The sink particle's initial mass is equal to the excess mass in that cell $M_{*,0} = [\rho - \rho_{\rm TJ}(1/4, l_{\rm max})]\Delta x_{l_{\rm max}}^3$, and our protostellar model is initialized with starting values for R_* and L_*. For these initial conditions, a $\sim M_\odot$ star is formed at $t = 0$ and no additional protostars form over the course of the simulation.

5.3 Results

To test ORION2 against the analytical predictions of Tan & McKee (2004), we first run the simulation without radiative feedback effects. Figure 5.4 shows the mass of the protostar as a function of time. Comparing with Equation (5.26), we note excellent agreement between our numerical results and the analytical prediction of Tan & McKee (2004). The offset occurs because the sink particle that forms at $t = 0$ cannot immediately accrete all of the material within the sink particle's accretion zone (see Section 5.2.6). Instead, this material gets accreted over the first few timesteps, which offsets the $M_*(t)$ curve compared to the one predicted by Tan & McKee (2004).

From here we considered the effects of protostellar feedback on the infalling gas. Without any initial angular momentum, infall proceeds radially and no accretion disk forms during the simulation. Therefore, radiation can effectively shut down accretion onto the star if the heated ionized gas can escape the gravitational pull of the star, which occurs when the radius of the HII region exceeds the star's gravitational radius

$$r_g = \frac{GM_*}{c_s^2} = 525 \left(\frac{M_*}{100 M_\odot}\right) \left(\frac{T}{2.5 \times 10^4 \text{ K}}\right)^{-1} \text{ AU} \tag{5.28}$$

in size (McKee & Tan 2008). The sink particle's accretion zone spans four grid cells in radius, equivalent to $r_a = 157$ AU for our highest resolution simulations. This radius considerably exceeds the radius of our central Pop III star, which is less than $100 R_\odot \sim 0.5$ AU in size. Because the gas within the sink particle's accretion zone has become modified through accretion, it has a reduced optical depth that allows radiation to escape more readily. For the resolution of this simulation, the radius of the accretion zone matches r_g when $M_* = 30 M_\odot$. Since the ionizing radiative output is considerable at these mass (Figure 5.5), artificial breakout of the HII region could occur if we employed the ray tracing routine at this mass. Indeed, we confirmed this to be the case by employing the ray tracing routine once $M_* \approx 45 M_\odot$ and $r_g/r_a = 1.5$. Within $\sim 2 \times 10^3$ yrs, the accretion rate decreased from the expected $\dot{M}_* = 5 \times 10^{-3}$ $M_\odot$/yr to 10^{-5} $M_\odot$/yr.

The primary goal of this chapter is to study the breakout of the HII region around a massive primordial star. In this case where the total angular momentum of the system is low, this breakout effectively sets the star's final mass. Therefore, we seek to demonstrate two cases: one where despite radiative feedback the HII region remains confined within the star's gravitational radius, and another where the star has exceeded $M_* \sim 275 M_\odot$ and the HII region grows and exceeds this radius (Tan & McKee 2004). We note that when rotation is included, the density distribution becomes modified and breakout above and below the denser accretion disk can occur at earlier masses (McKee & Tan 2008).

Therefore, we turn the ray tracing routine on at two particular instances, once when $M_* \approx 127 M_\odot$ and again when $M_* \approx 300 M_\odot$. The former's HII region should remained confined near the star, where the latter's HII region should eventually break free. The evolution of the HII region expands within $\sim 10^4$ years (McKee & Tan 2008), so we run these two simulations for $\geq 2 \times 10^4$ years to assess whether the HII region will remain confined or expand. For the high mass case, however, we note that since the ray tracing is immediately turning on, the expansion of the HII region will initially proceed faster than what would be predicted from theory. In a more self-consistent simulation, an HII region would have already been present at this stellar mass.

In Appendix C, we discussed our ionization sub-cycling routine and showed that for the simplified D-Type ionization front test, using a ray trace timestep that is 10% the smallest value of the recombination timescale $1/\alpha_B n_e$ or the temperature evolution timescale $\dot{e}/e$, where e is the internal energy of the gas. Even when the timestep criterion was increased to 50%, our results still remained within 10% of the analytical prediction. For both of these simulations, we use sub-cycling limitation of 100% instead, since the ray tracing routine becomes computationally demanding once several layers of AMR are introduced.[3] This will inevitably introduce some error into our results, but the introduced error should not significantly change the intended results.

Figure 5.6 shows example slices of the gas temperature and neutral hydrogen mass fraction for both of these cases after radiative feedback has been turned on, and a circle with radius r_g is drawn as a dotted circle. In the lower mass instance, $r_g/r_a = 3.3$, where the higher mass case has $r_g/r_a = 10.1$. For the lower mass snapshots, where $\dot{N}_{\rm ion} \approx 10^{50.2}$, the HII region remains confined within the gravitational radius of the star. The resulting mass accretion rate fluctuates and slightly reduces from the predicted rate of $\dot{M}_* \approx 3\times10^{-3}\ M_\odot/\mathrm{yr}$ (Tan & McKee 2004). The infalling gas replenishes the region just outside the sink particle's accretion radius with neutral hydrogen. As a result, only the gas just outside the accretion radius (and within the gravitational radius) becomes heated and partially ionized. Overall, the radiative feedback is unable to significantly alter the dynamics of the gas.

Alternatively, the $M_* \approx 300M_\odot$ ($\dot{N}_{\rm ion} \approx 10^{51}$) star creates an HII region that ultimately breaks free from the star. Within the first $\sim 10^3$ years, the radius of the HII region exceeds $\sim r_g/2$, where the infalling gas compares to $\sim 2c_s$. At this stage, the HII region launches a shockwave ahead of the ionized gas, which eventually accelerates beyond r_g. During this phase, we notice a sharp decrease in the total accretion rate onto the star, with the majority of the infalling gas piling up within the post-shock region. Had rotation been included, some of the shocked gas could have fallen onto the accretion disk and continued accreting this way. In our case, this breakout effectively sets the final mass of the central star at $M_* \approx 301M_\odot$, with accretion rates dropping several orders of magnitude.

We do identify one error that is introduced by using larger timesteps than we had used in Appendix C. Once the HII region forms, the time step is typically determined by the recombination timescale limit. The larger time steps mean that each individual ray carries more photons and rays travel farther before becoming depleted and terminated. Since we employ an operator-split method and apply recombinations only after the rays have all terminated, the rays travel farther than they would had recombinations been included self-consistently. Additionally, if multiple rays pass through a cell, rays moving in one direction may reach a cell first and ionize the cell before subsequent rays arrive. Coupled with the fact that the grid becomes coarser as the distance from the star increases, so that ray splitting occurs less frequently, asymmetrical ray artifacts become more pronounced. The result of these errors is asymmetrical spokes of heated and partially ionized gas that extend beyond a generally spherical HII region. These spokes generally develop and fade as the gas heats and cools, due to the random initial orientation of the rays at the start of each ray trace. In the future, using a more restrictive time step and even allowing the rays to divide more readily can help eliminate these artifacts.

---

[3] Particularly, recall that the ray trace is called at least once on every finest time step.

5.4 Summary and Future Work

We have examined the formation of a Pop III star resulting from the gravitational collapse of a virialized spherical cloud. Tan & McKee (2004) and McKee & Tan (2008) described the rate of collapse of one of these primordial gas cores, whose physical properties are set mostly by the cooling of H_2. Tan & McKee (2004) described the rate of collapse of these gas cores, and we find this simple analytical picture is reproduced with the ORION2 code when radiative feedback from the forming star is not included (Figure 5.4).

When radiative feedback is included, the heated photoionized gas within the HII region breaks free of the central star when this region grows beyond the gravitational radius $r_g = GM_*/c_s^2$ in size. When the gravitational radius is comparable to the sink particle's accretion radius, which spans four cells on the finest AMR level, the HII region can artificially break free and shut off the accretion onto the central star, a result of the gas density being artificially modified in a region that greatly exceeds the physical stellar radius. With future studies, where we will have ∼AU resolution on the finest cell level, the similarly reduced accretion zone around the star will allow us to obtain a more self-consistent solution for the evolution of the HII region at early times. In this case, the increased mass density of the infalling gas will confine the ionized gas, as opposed to our current work where the accretion zone exceeds the gravitational radius until the star has grown beyond $\sim 30 M_\odot$. To overcome this issue, we employed our ray tracing routine only when the mass of the star has exceeded $100 M_\odot$, and the gravitational radius is $\gtrsim 3$ times the accretion radius of the star.

For these simulations, as the spherical cloud collapses and the star grows, the total mass accretion rate gradually decreases . Eventually the hydrogen-ionizing luminosity $\dot{N}_{\rm ion}$ exceeds the incoming neutral hydrogen flux and the radiation can begin developing a long-lasting HII region. As the ionization luminosity continues to grow, this region can grow beyond r_g. Employing radiative feedback when the star is $M_* = 127 M_\odot$, the higher mass accretion rate and a lower ionizing luminosity flux prevents the HII region from growing beyond r_g. However, our $M_* = 300 M_\odot$ star creates an HII region that spreads beyond r_g and effectively shuts off the accretion rate. These simple simulations shown here corroborate prior studies (e.g., McKee & Tan 2008; Hosokawa et al. 2016; Stacy et al. 2016) that show that the ionizing radiative feedback of the first stars plays an important role in determining the star's final mass.

Section 5.2 highlighted several additional improvements we will employ in future studies of Pop III stars, including incorporating a primordial chemistry package and further improvements to our stellar evolution model. An additional improvement we will incorporate is the use initial conditions drawn from cosmological simulations. We plan to draw our initial conditions from Stacy & Bromm (2013), who ran N-body and SPH simulations using Gadget-2 (Springel 2005). They evolved a 1.4 comoving Mpc box from cosmological initial conditions at $z = 100$ using a ΛCDM cosmology with $\Omega_\Lambda = 0.7$, $\Omega_{\rm M} = 0.3$, $\Omega_{\rm B} = 0.04$, and $H_0 = 70$ km/s/Mpc. As the simulation progressed, they continued to refine ten regions that formed dark matter mini-halos, each tens of parsecs in size. Our next set of simulations will simulate the central $(\rm pc)^3$ [physical] region of their Halo 9, starting at redshift $z = 24.71$. To prepare the data for use in ORION2, we developed a projection routine that transforms SPH simulation to AMR grid data. This routine, described in Appendix D, allows us to project the baryonic gas component, centered around the location of highest density, onto a 512^3 grid. This region is well inside the dark matter halo such that the gravitational potential from the dark matter is uniform, and therefore we do not have to include dark matter in our ORION2 simulations. Utilizing AMR to achieve resolutions down to a few AU, we will be able to study the collapse and fragmentation of the first star forming clouds, allowing us to assess the role radiative feedback plays in the determining the initial mass function, multiplicity, and ultimate fate of the first stars.

Acknowledgements

ATL thanks the members of his dissertation committee for comments that improved this chapter. ATL acknowledges partial support from a UC Berkeley Dissertation Fellowship at UC Berkeley, the LLNL-B569409 grant from the Lawrence Livermore National Lab, TCAN grant NNX-14AB52G from the National Aeronautics and Space Administration, and Stella Offner at the University of Massachusetts Amherst. We used the Python package Yours Truly (Turk et al. 2011b) to produce all the figures that present simulation data.

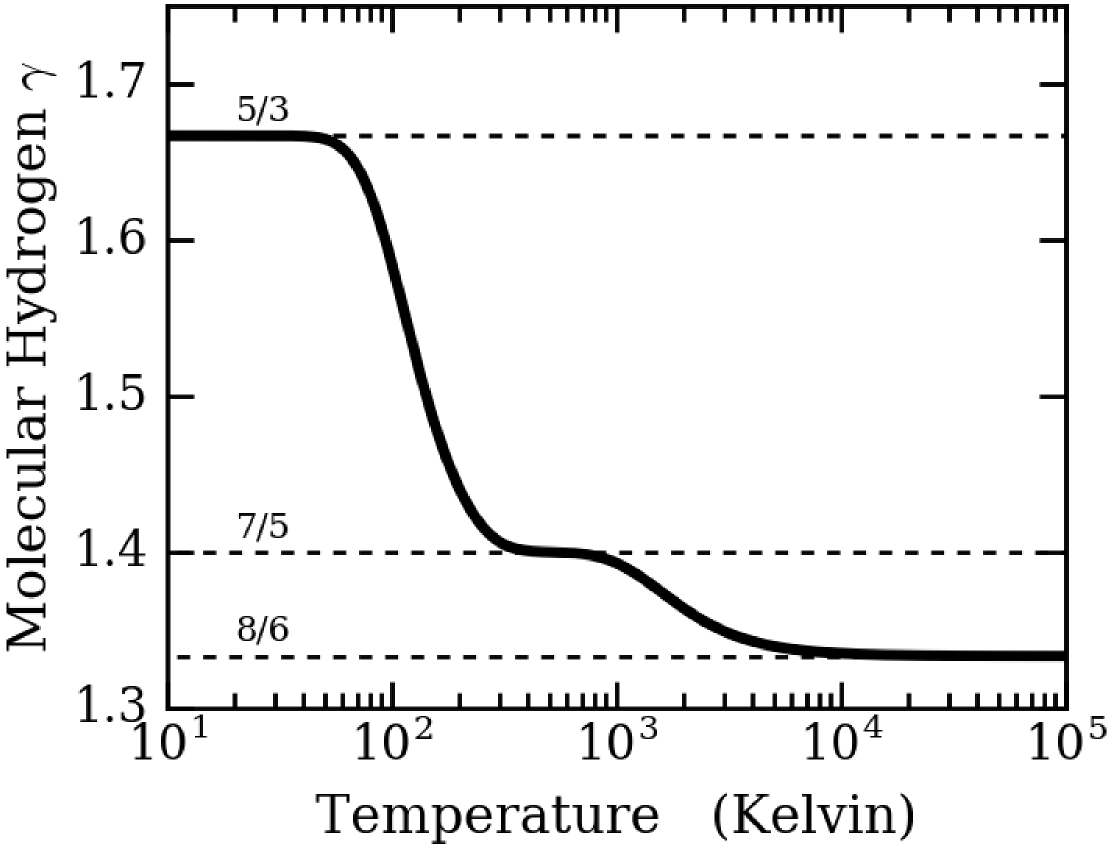

Figure 5.1: Adiabatic index of molecular hydrogen, found by summing 500 terms in its partition function, assuming an ortho to para ratio of 3:1. Above $\sim 10^2$ and $\sim 10^3$ Kelvin, rotational and vibrational modes become excited, respectively.

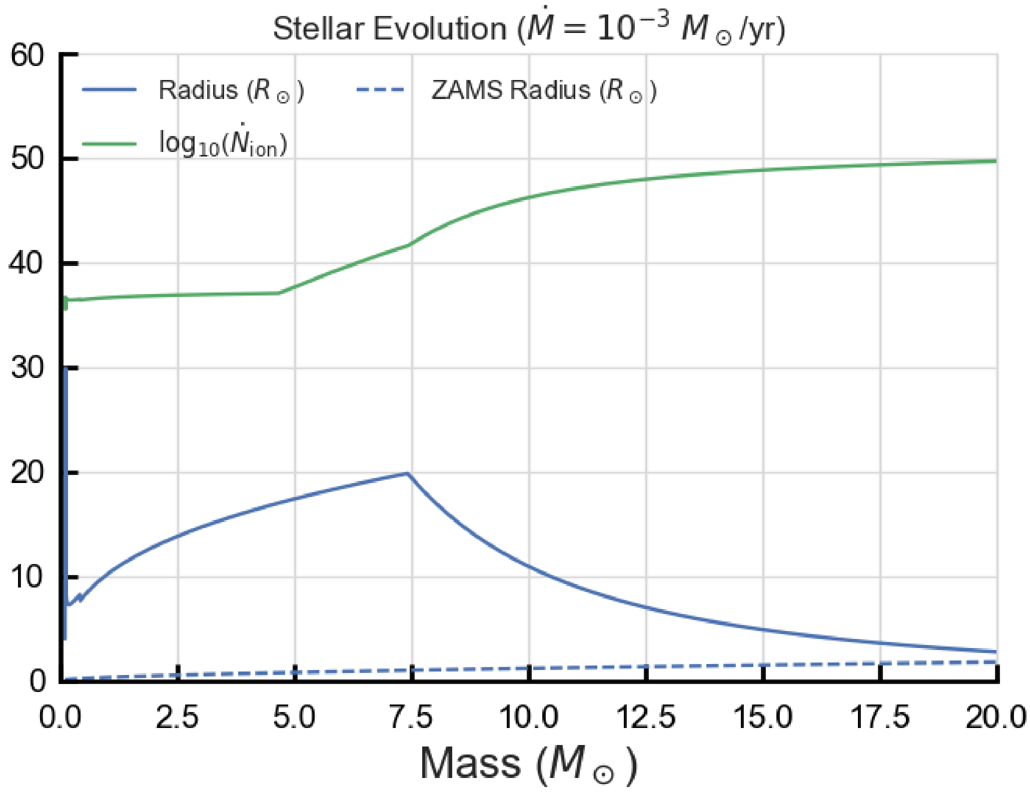

Figure 5.2: Example of our stellar evolution model in the case of a constant mass accretion rate of $\dot{M} = 10^{-3} M_\odot$/yr, showing the logarithm of $\dot{N}_{\rm ion}$ (top, units of number of ionizing photons/sec) and the stellar radius R_* (bottom, units of $R_\odot$). The dashed line shows the main sequence radius as a function of mass.

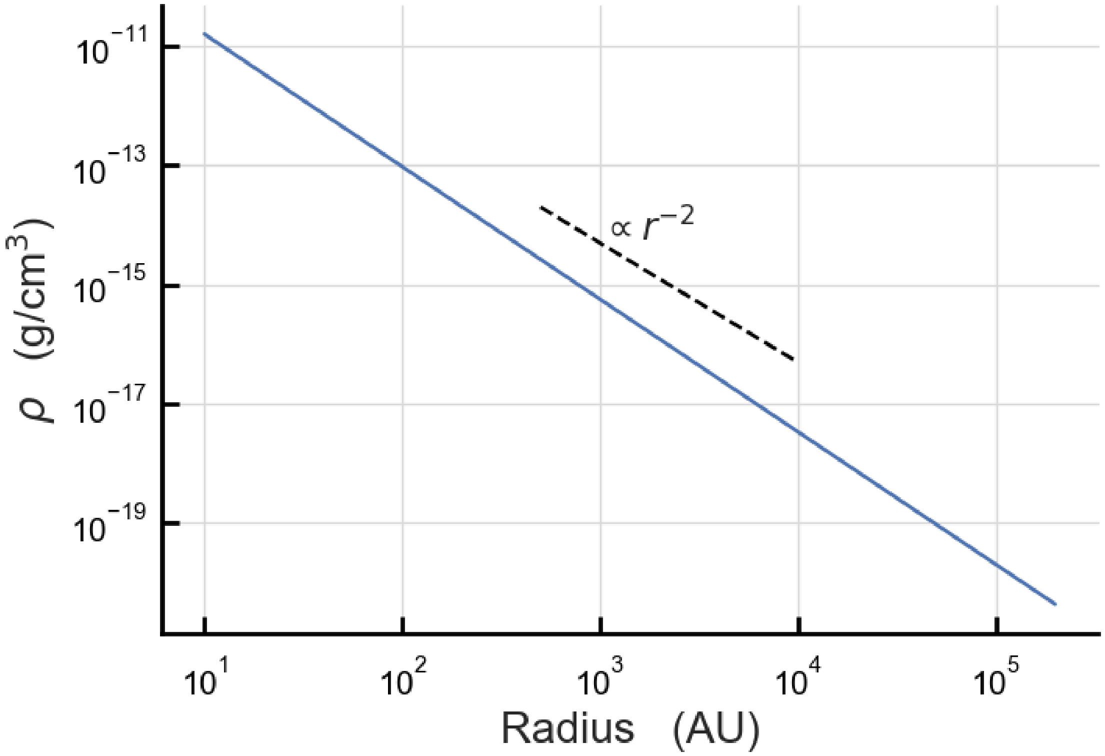

Figure 5.3: Initial equilibrium solution for the density profile from Tan & McKee (2004). Rotation is not included and a polytropic relationship $P = K\rho^{1.1}$ is assumed. The dashed line shows a r^{-2} profile, for comparison.

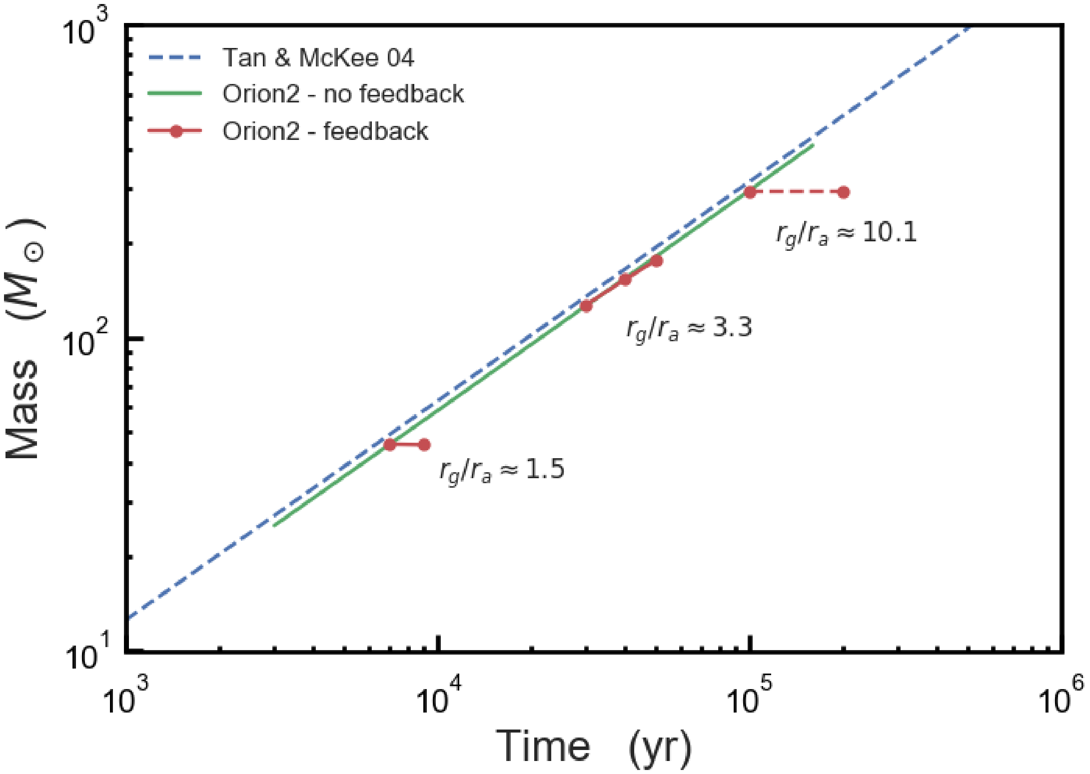

Figure 5.4: Mass of the central star as a function of time. Without radiative feedback, we get excellent agreement with the analytical prediction of Tan & McKee (2004, also our Equation 5.26), shown as the dashed line. When feedback is included but the accretion zone of the star compares to the gravitational radius (Equation 5.28), the HII region can break out and artificially shut off accretion onto the star. When $M_* \approx 120 M_\odot$, the feedback is unable to sufficiently ionized the infalling gas, and the star continues to grow. At $M_* \approx 300 M_\odot$, the ionizing luminosity ionizes the within and beyond r_g, and accretion onto the star precipitously drops within $\sim 10^3$ years. Though the simulation was only run for $\sim 10^4$ years, we extend the line to $t = 2 \times 10^5$ years for clarity.

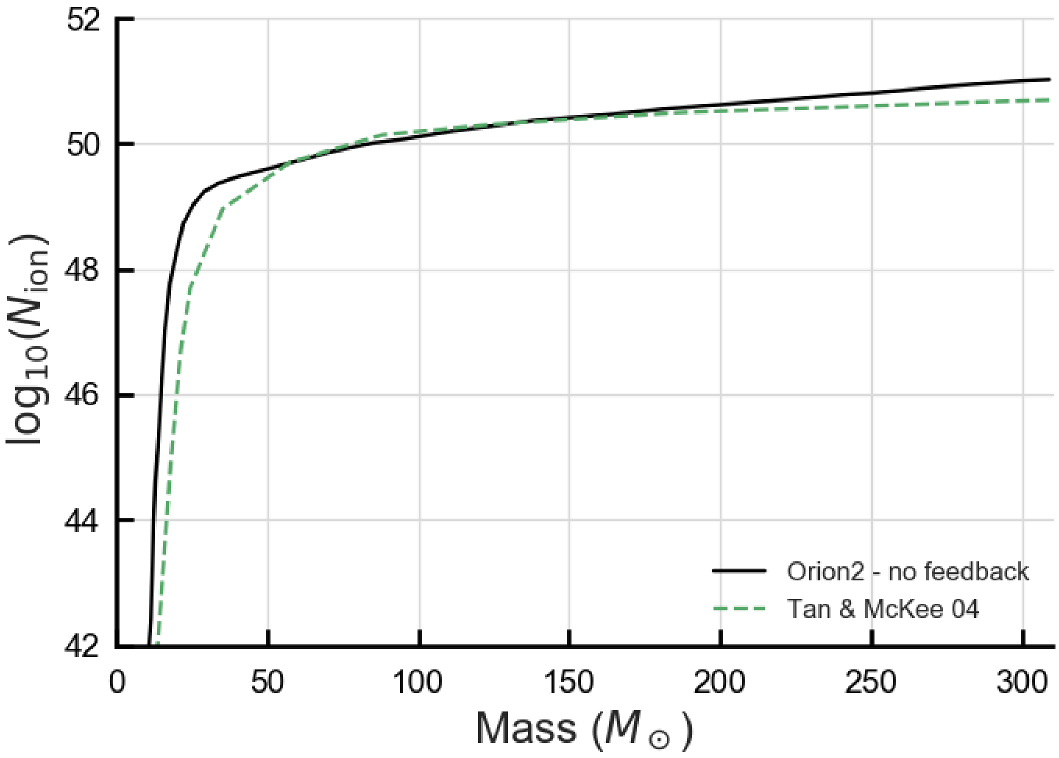

Figure 5.5: Number of hydrogen-ionizing photons emitted from the central star as a function of stellar mass. The dashed line shows the results of Tan & McKee (2004). The ionizing luminosity of the central star when the accretion rate is not modified by feedback effects, is shown as the solid line.

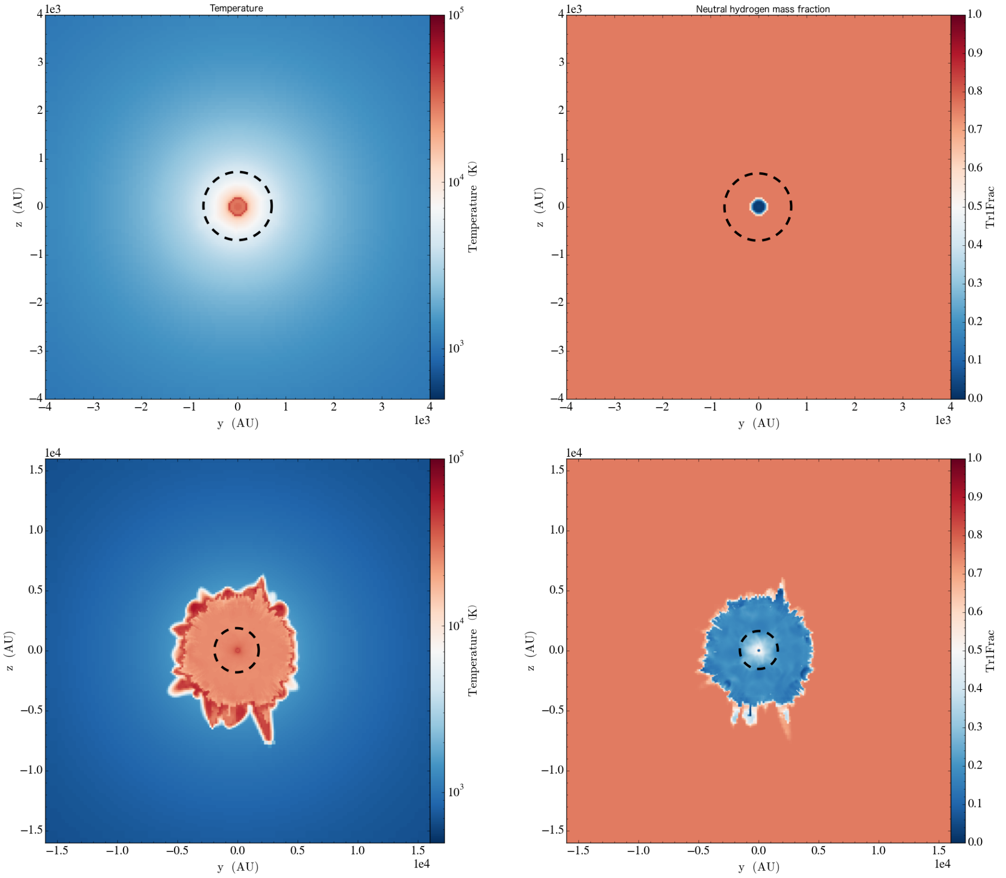

Figure 5.6: Confined and expanding HII regions, as seen using temperature slices (left column) through the center of the domain and the neutral hydrogen mass fraction (right column). Both plots zoom in toward the sink particle region. The gravitational radius of the central star is shown as a dotted circle. (Top) Example slices when the initially $M_* \approx 120 M_\odot$ star has reached $\approx 140 M_\odot$. The ionized gas barely extends beyond the accretion radius of the sink particle, and the gas just beyond the sink particle's radius becomes partially ionized and heated. The overall accretion rate fluctuates but does not substantially decrease. (Bottom) Example slice when $M_* \approx 300 M_\odot$. After 10^3 years, the HII region has moved beyond the gravitational radius and continues to move outward as a D-Type ionization front. The value of $\dot{M}_*$ has dropped over three orders of magnitude.

Appendix A

Background Protoplanetary Disk Model

[1] For numerical estimates in this paper, we adopt the standard disk model derived in the review by Chiang & Youdin (2010). The disk has surface densities

$$\Sigma_{\rm g} = 2200\,F\left(\frac{r}{\rm AU}\right)^{-3/2}\ {\rm g\,cm^{-2}} \tag{A.1}$$

$$\Sigma_{\rm d} = 33\,F\,Z_{\rm rel}\left(\frac{r}{\rm AU}\right)^{-3/2}\ {\rm g\,cm^{-2}} \tag{A.2}$$

in gas (g) and dust (d). The dimensionless parameters F and $Z_{\rm rel} \equiv (\Sigma/\Sigma_{\rm g})/0.015$, typically of order unity, describe how much total mass the disk has relative to the minimum-mass solar nebula, and how metal-rich the disk is compared with a gas of solar abundances, respectively. The minimum-mass solar nebula ($F = 1$, $Z_{\rm rel} = 1$) uses a condensate mass fraction for solar abundances of $\Sigma_{\rm d}/\Sigma_{\rm g} = 0.015$ (Lodders 2003). Values of $Z_{\rm rel} > 1$ correspond to supersolar metallicities $\Sigma_{\rm d}/\Sigma_{\rm g} > 0.015$. Integrated to $r = 100\,{\rm AU}$, equation (A.1) yields a total disk mass of $0.03 F M_\odot$.

At the disk midplane, the gas temperature, scale height, and density are given by

$$T = 120\left(\frac{r}{\rm AU}\right)^{-3/7}\ {\rm K} \tag{A.3}$$

$$H_{\rm g} = 0.022 r\left(\frac{r}{\rm AU}\right)^{2/7} \tag{A.4}$$

$$\rho_{\rm g0} = 2.7\times10^{-9} F\left(\frac{r}{\rm AU}\right)^{-39/14}\ {\rm g\,cm^{-3}}\,. \tag{A.5}$$

These are adapted from Chiang & Goldreich (1997), adjusted for a disk obeying (A.1)–(A.2), orbiting a pre-main-sequence star of mass $M_* = 1 M_\odot$, radius $R_* = 1.7 R_\odot$, and temperature $T_* = 4350\,{\rm K}$.

---

[1] Large portions of this appendix have been previously published as Lee, A. T., Chiang, E., Asay-Davis, X., Barranco, J. 2010. Forming Planetesimals by Gravitational Instability. II. How Dust Settles to its Marginally Stable State. ApJ, 725, 1938.

A.1 The Super-Linear Relation Between μ_0 and Σ_d/Σ_g

We derive μ_0 as a function of Σ_d/Σ_g under the assumption of a constant Ri. Some evidence supporting a constant Ri was found in our simulations (§3.5.1). The density profile for constant Ri is used in a number of papers (Sekiya 1998; Youdin & Shu 2002; Paper I) and we begin by repeating the result, neglecting self-gravity as we have throughout our paper. The dust-to-gas ratio is given by

$$\mu(z) = \left[\frac{1}{1/(1+\mu_0)^2 + (z/z_d)^2}\right]^{1/2} - 1 \tag{A.6}$$

where

$$z_d \equiv \frac{Ri^{1/2}\, v_{\max}}{\Omega_K} \tag{A.7}$$

is a characteristic dust height and $v_{\max} = \eta\Omega_K r$ (see equations 3.5 and 3.6) is a constant equal to the difference in azimuthal velocity between a strictly Keplerian flow and dust-free gas. The dust density drops to zero at

$$z = \pm z_{\max} = \pm\frac{\sqrt{\mu_0(2+\mu_0)}}{1+\mu_0} z_d\,. \tag{A.8}$$

A comment on equation (A.6), in the limit that $\mu_0 \gg 1$: except where μ is nearly constant near $z \ll z_{\max}/\mu_0$ and where it falls to zero near $z = z_{\max}$, the shape of $\mu(z)$ is that of $1/z$. This form follows simply from the constancy of Ri. Because the numerator of Ri is approximately constant with μ (§3.5.2), the denominator must be as well: $\partial v_\phi/\partial z \sim (v_{\max}/\mu)/z \sim$ constant, which implies $\mu \propto 1/z$. From this we can deduce the super-linear trend between μ_0 and Σ_d/Σ_g as follows. The integral of μ with respect to z is proportional to the total surface density of dust Σ_d. Because $\mu \propto 1/z$, flattening off as z decreases below $z_{\max}/\mu_0$, this integral varies as $\log\mu_0$. Then $\mu_0 \propto \exp\Sigma_d$, crudely.

More formally, we have

$$\Sigma_d = 2\int_0^{z_{\max}} \rho_d\, dz = 2\rho_{g0}\int_0^{z_{\max}} \mu\, dz \tag{A.9}$$

where ρ_{g0} is the midplane gas density, assumed constant because $z_{\max} \ll H_g$. The gas density profile always well approximates a Gaussian (see footnote 4), from which it follows that $\Sigma_g \approx \sqrt{2\pi}\rho_{g0}H_g$. Then

$$\frac{\Sigma_d}{\Sigma_g} = \sqrt{\frac{2}{\pi}}\frac{1}{H_g}\int_0^{z_{\max}} \mu\, dz\,. \tag{A.10}$$

Inserting (A.6) into (A.10) we have

$$\sqrt{\frac{\pi}{2}}\frac{H_{\mathrm{g}}}{z_{\mathrm{d}}}\frac{\Sigma_{\mathrm{d}}}{\Sigma_{\mathrm{g}}} = \log[1+\mu_0+\mu_0^{1/2}(2+\mu_0)^{1/2}] - \frac{\mu_0^{1/2}(2+\mu_0)^{1/2}}{(1+\mu_0)}\,. \tag{A.11}$$

In the limit $\mu_0 \gg 1$, the exponential dependence of μ_0 on $\Sigma_{\mathrm{d}}/\Sigma_{\mathrm{g}}$ is evident. Equation (A.11) is plotted in Figure A.1, with $Ri = 0.25$ and $v_{\mathrm{max}}/c_{\mathrm{s}} = 0.025$. Overlaid is the same equation but with varying $Ri = Ri_{\mathrm{crit}} \approx 0.25(\mu_0/9)$, the relation we found in Paper I (see Figure 5 of that paper). The two data points representing the maximum μ_0 achieved in this paper are also plotted. The data track the variable $Ri_{\mathrm{crit}}(\mu_0)$ curve much better than the constant Ri curve.

Finally note that $H_{\mathrm{g}}/z_{\mathrm{d}} \propto c_{\mathrm{s}}/v_{\mathrm{max}}$ enters into equation (A.11) the same way that $\Sigma_{\mathrm{d}}/\Sigma_{\mathrm{g}}$ does. Thus μ_0 increases super-linearly with $c_{\mathrm{s}}/v_{\mathrm{max}}$ as well. This result leads us to suspect that our numerical results for μ_0 (2.9, 26.4) depend sensitively on our choice for $v_{\mathrm{max}}/c_{\mathrm{s}} = 0.025$. In this paper we did not run simulations with different $v_{\mathrm{max}}/c_{\mathrm{s}}$ and so did not test this suspicion.

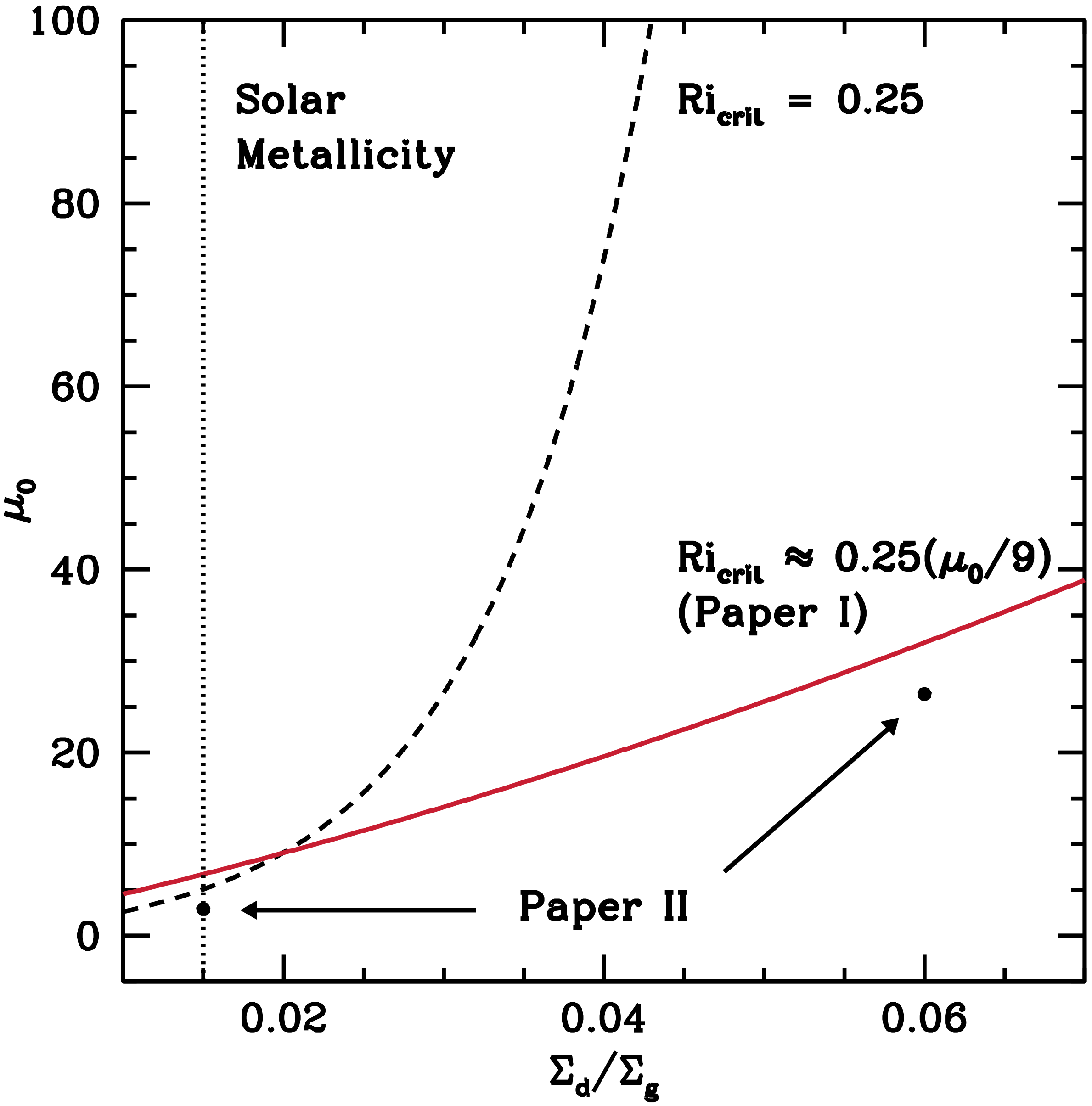

Figure A.1: Super-linear trend between the midplane dust-to-gas ratio μ_0 and height-integrated metallicity Σ_d/Σ_g for dust profiles characterized by a spatially constant Richardson number $Ri_{\rm crit}$. Equation A.11 is plotted twice: the dashed curve uses $Ri_{\rm crit} = 0.25$, whereas the solid curve varies $Ri_{\rm crit}$ according to the relation found in Paper I: $Ri_{\rm crit} \approx 0.25(\mu_0/9)^{1.0}$ (see Figure 5 of Paper I). Both curves fix $v_{\rm max}/c_{\rm s} = 0.025$. The maximum values of μ_0 achieved in this paper are plotted as points. These data follow the variable $Ri_{\rm crit}(\mu_0)$ curve more closely than the constant $Ri_{\rm crit}$ curve, corroborating the evidence we found in §3.5.1 that $Ri_{\rm crit}$ is spatially constant but varies with μ_0 (equivalently Σ_d/Σ_g).

Appendix B

Sink Particle Sub-grid Model

[1] A collapsing molecular cloud forms structures that can be several to many orders of magnitude smaller in size than the original cloud. Large scale astrophysical simulations attempt to resolve these structures in order to follow their evolution, but this requires overcoming additional computational burdens, be it the reduced time step needed to ensure numerical stability or resource demands due to the increased memory requirements of the computational domain. Sink particle methods have been developed for both Lagrangian (Bate et al. 1995; Hubber et al. 2013) and Eulerian mesh (Krumholz et al. 2004b; Federrath et al. 2010) codes to allow for collapsing flows that proceed beyond the finest resolved scale of the simulation. Material that enters these sink particles can be removed from the computational domain according to an analytical prescription that is intended to best estimate the physical processes that occur at those unresolved scales. In this section, we develop an implementation for embedding Lagrangian sink particles into an Eulerian mesh to model the accretion of an ideally magnetized gas, extending the approaches of Krumholz et al. (2004b, 2006) for non-magnetized flow. We review the criterion for the creation of a sink particle in a magnetized medium and we then determine the accretion rate of the sink particle, allowing for the finite resolution of the data. Only these prescriptions depend on the strength of the local magnetic field, whereas the others (sink particle mergers, coupling the sink particle's gravity to the hydrodynamics, etc.) do not and so are left unchanged. It should be borne in mind that, based on the observation that stars have far less magnetic flux than the gas from which they formed, we assume that the sink particles accrete mass but not magnetic flux.

B.1 Sink Particle Creation

In simulations of gravitational collapse, mass accumulates in a small fraction of the grid cells. Sink particle algorithms must be able to identify whether these regions would continue to collapse if they were afforded higher resolution. On physical grounds, Jeans (1902) showed

---

[1] Large portions of this appendix have been previously published as Lee, A. T., Cunningham, A., McKee, C., Klein, R. 2014. Bondi-Hoyle Accretion in an Isothermal Magnetized Plasma. ApJ, 783, 50.

that perturbations on scales larger than the Jeans length,

$$\lambda_{\rm J} = \left(\frac{\pi c_{\rm s}^2}{G\rho}\right)^{1/2} , \tag{B.1}$$

are unstable since thermal pressure cannot resist the self-gravity of the gas. Truelove et al. (1997) showed that Eulerian simulations are subject to purely numerical fragmentation if this Jeans scale is not resolved by at least four cells. A sink particle is introduced at the center of a cell when the cell mass density ρ exceeds a critical density, which we term the Truelove-Jeans density,

$$\rho_{\rm TJ} = \frac{\pi J^2 c_{\rm s}^2}{G\,\Delta x^2} , \tag{B.2}$$

where $J = \Delta x/\lambda_{\rm J}$ is the (user-provided) inverse of the number of cells required resolve the local Jeans length. Once this is satisfied, a sink particle is initialized with mass $(\rho - \rho_{\rm TJ})\Delta x^3$, and an equal mass is removed from the gas in the host cell. Myers et al. (2013b) extended this to incorporate ideal MHD, deriving a magnetic Truelove criterion: sink particles are initialized in cells whose density exceeds

$$\rho_{\rm TJ,mag} = \rho_{\rm TJ}\,(1 + 0.74/\beta) \tag{B.3}$$

(see their Appendix A). The additional term arises in Equation (B.3) because of the inclusion of magnetic pressure, which also acts to prevent gravitational collapse. Federrath et al. (2010) arrived at a similar condition. Following the work of Myers et al. (2013b), we adopt $J = 1/8$ in our Truelove criterion for sink particle creation.

B.2 Sink Particle Accretion

B.2.1 Estimated Accretion Rate, $\dot{M}_{\rm fit}$

After the sink particle has formed, it will continue to accrete nearby gas. The rate of accretion onto the sink particle may be determined by processes that occur on scales that cannot be resolved in many modeling applications of interest. Here we develop an expression for the accretion rate onto sink particles that incorporates our new interpolation formulas (Equations 4.27 and 4.30) and that works at all resolutions. Our results for the accretion rate are based on the assumption that the ambient medium is uniform, but in simulations of star-forming regions, the medium is far from uniform. We therefore need an expression for the accretion rate that depends on locally measured quantities. To obtain this, we ran the models in this work and in Cunningham et al. (2012) at a range of different resolutions and developed a prescription for the correct accretion rate based on quantities measured in the vicinity of the sink particle.

Two limits for the sink-particle accretion rate may be considered: The well-resolved case, in which $r_{\rm ABH}$ is much larger than the cell size, and the under-resolved case, in which it is much smaller. We need a prescription for the accretion rate that works in these limits, as well as in the intermediate case. Since $r_{\rm ABH} \propto M_*$, more massive stars are likely to have accretion flows that are well resolved, whereas low-mass stars are likely to have under-resolved flows (see Section 4.5).

First consider the well-resolved case. In this case, the exact prescription for how much mass gets removed from the host and neighboring cells of the sink particle is not important since the flow's transition to velocities exceeding the fast magnetosonic velocity, $v_{\rm F}$, is resolved. Once $v \geq v_{\rm F}$, the flow becomes causally disconnected from the background and will collect near the sink particle regardless of the conditions of the surrounding medium. If our prescription underestimates the mass accretion rate, gas will collect in the sink particle's host cell until the gas density exceeds $\rho_{\rm TJ,mag}$, at which point a sink particle will be formed and immediately merged with the existing sink particle (Krumholz et al. 2004b). If our prescription overestimates the accretion rate, the density in the superfast infall will drop below the correct value, and the accretion rate in the next time step will be reduced. Thus, in the well-resolved case, the mass accretion rate is effectively set by the supersonic infall. The work described earlier in this paper using `RAMSES` is an example of this regime, since we ensured that the length scale $r_{\rm ABH}$ was resolved. The insensitivity to the accretion rate algorithm is also true for simulations of global supersonic collapse onto a particle (Shu 1977).

Next, consider the under-resolved regime, where $r_{\rm ABH}$ is not well resolved. In this regime, the flow inside the sink cell is causally connected to the rest of the flow for $v_0 < v_{\rm F}$ (the subfast case) and to the flow in the downstream Mach cone for $v_0 > v_{\rm F}$ (the superfast case). The prescription for the amount of gas to be taken from the particle's host cell is important in this regime: not all the gas that flows through the cell should necessarily accrete onto the sink particle. Furthermore, the amount of gas in the particle's host cell determines the pressure support in the cell. The correct accretion rate in this regime is the Alfvén-Bondi-Hoyle rate that we have determined. The problem is that this accretion rate depends on the properties of the ambient medium, which we have assumed is homogeneous; in a simulation of a star-forming region, however, there is no homogeneous ambient medium. We therefore must estimate the accretion rate from the values of the parameters in the vicinity of the sink particle. Star-forming regions are supersonically turbulent, and Krumholz et al. (2006) showed that such turbulence has two countervailing effects on the accretion rate: the rate is increased by the density fluctuations, but decreased by the vorticity in the flow. Our simulations include the first effect so long as density variations are well-resolved (i.e., except in shocks). We have not included vorticity, however, so our results are necessarily approximate when applied to a turbulent medium.

The results we have obtained for the accretion rate depend on quantities—the ambient density, ρ, the ambient plasma β, the Mach number of the flow past the accretor, $\mathcal{M}$, and the angle between the flow velocity and the field, θ—that are assumed to be constant far from the accreting particle. In star-forming regions, however, these quantities are not constant far from the accretor. To deal with this problem in simulations, we have developed a two-step procedure: We first measure these quantities near the accretor; these values are denoted with a bar (e.g., $\bar{\rho}$). In the limit of low-resolution, these values can be used for the ambient values that appear in our results for the accretion rate in Section 4.2.2. However, when the flow is resolved, gravitational focusing amplifies the density and magnetic field near the accretor, and values of these quantities must be extrapolated in order to estimate the ambient values; these estimates for the ambient values are denoted by a dagger (e.g., $\rho^\dagger$). We do not distinguish the local and extrapolated values of the Mach number of the flow relative to the sink particle since they are the same in the limit of steady accretion (i.e., $\mathcal{M}^\dagger \simeq \bar{\mathcal{M}}$).

In general, the angle θ between $\mathbf{v}$ and $\mathbf{B}$ may be time-dependent, so it is not possible to infer the ambient value of θ by extrapolation; we therefore set $\theta^\dagger = \bar{\theta}$. With Equations (4.27) and (4.30) for $\dot{M}_\parallel$ and $\dot{M}_\perp$, our estimate of mass accretion rate onto a sink particle is then

$$\dot{M}_{\rm fit} = \dot{M}_\perp(\rho^\dagger, \beta^\dagger, \bar{\mathcal{M}}) \sin^2\bar{\theta} + \dot{M}_\parallel(\rho^\dagger, \beta^\dagger, \bar{\mathcal{M}}) \cos^2\bar{\theta} \ . \tag{B.4}$$

In order to compute local values of the quantities $\bar{\rho}$, $\bar{\beta}$, $\bar{\mathcal{M}}$, and $\bar{\theta}$, we take averages over all the cells in a spherical shell of radius $r_{\rm avg} \pm \Delta x$ around the sink particle, where Δx is the smallest grid size of the computational domain. The shell radius must be larger than $4\,\Delta x$, the region from which mass is removed from the cells and deposited onto the particle (Krumholz et al. 2004b), but not so large that the sink particle is sampling far from the local region that sets its current accretion rate. We have found the best results by sampling over a shell with radius $r_{\rm avg} = 11\Delta x$. In particular, the value of $\bar{\mathcal{M}}$ is computed as a mass-average over the volume of the shell, and $\bar{\theta}$ is the angle between the volume-averaged magnetic field and volume-averaged momentum directions within this shell. We have found the best results by imposing the refinement criterion that every AMR level cover a sphere with a radius of 16 zones centered on the sink particle.

As noted above, if the flow near the sink is well resolved—i.e., if $r_{\rm avg} < r_{\rm AB}$, $r_{\rm BH}$—then the density and magnetic field will be amplified by gravitational focusing, so that $\bar{\rho} > \rho^\dagger$ and $\bar{B} > B^\dagger$. First consider the density. Following Krumholz et al. (2004b), we determine $\rho^\dagger$ by assuming that the density near the sink particle is well approximated by the stationary Bondi (1952) solution, $\rho(r) = \rho^\dagger\alpha(r/r_{\rm B})$, where the function α is determined by a set of transcendental equations. We evaluate this function at $r_{\rm avg}$, so that $\rho(r_{\rm avg}) = \bar{\rho}$. To incorporate relative motion between the sink particle and the gas, we instead normalize the radius in α to $r_{\rm BH}$ (Krumholz et al. 2004b), giving

$$\bar{\rho} = \rho^\dagger\alpha(r_{\rm avg}/r_{\rm BH}) \ . \tag{B.5}$$

The function $\alpha(r)$ is a monotonically decreasing function with the limit $\alpha \to 1$ as $r \to \infty$. In Figure B.1 we plot $\alpha(r/r_{\rm BH})$. In the absence of a magnetic field, our estimate for the ambient density would be $\rho^\dagger = \bar{\rho}/\alpha(r_{\rm avg}/r_{\rm BH})$. Since the magnetic field limits the compression, we adopt an ansatz for the ambient density in which $\rho^\dagger = \bar{\rho}\alpha^{-\chi}$, where χ goes smoothly from 1 in the hydrodynamic limit to 0 (i.e., no compression) in the limit of a strong field; an explicit expression for χ as a function of $\beta^\dagger$ will be given below. At sufficiently high resolution, directly simulating the accretion onto a totally absorbing sphere is more precise than our accretion rate fits. We have found that this condition is met when $\Delta x \lesssim r_{\rm ABH}/8$. A resolution-dependent, piecewise prescription for the argument of $\alpha^{-\chi}$ is necessary to give a precise accretion rate in the limit of an asymptotically converged grid resolution, regardless of the error of our approximate fit.

Whatever the functional form for $\alpha^{-\chi}$, it has several requirements. In order to have the sink particle transition to a totally absorbing sphere in the high-resolution limit, the value of $\rho^\dagger$ should be similar to the near-sink value of $\bar{\rho}$— i.e., $\alpha^{-\chi}$ should be of order unity. For under-resolved flows, $r_{\rm avg}$ is sampled farther from the sink particle. In the weak field limit, $\alpha^{-\chi}$ should approach $1/\alpha(r_{\rm avg}/r_{\rm BH})$, which gives an accurate correction factor between $\rho^\dagger$ and $\bar{\rho}$ (Krumholz et al. 2004b). As β decreases, the value of $\alpha^{-\chi}$ should decrease until the field becomes so strong that any gravitational enhancement of ρ occurs well within $r_{\rm avg}$. For smaller values of β, $\alpha^{-\chi}$ should rise back up to be of order unity. We adopt

$$\rho^\dagger = \bar{\rho}\left(\alpha\left[\max\left(1, \frac{8\Delta x}{r_{\rm ABH}}\right)\frac{11 r_{\rm ABH}}{8 r_{\rm BH}}\right]\right)^{-\chi} \tag{B.6}$$

as our functional form for $\rho^\dagger$, which we later show achieves all the requirements above. Note that the factor 11/8 does not have any special significance; it is the result of our choice of $r_{\rm avg} = 11\Delta x$ and our result that the criterion for being well resolved is $\Delta x < r_{\rm ABH}/8$.

The magnetic field is also amplified in the accretion flow. For 1D compressions, $B \propto \rho$ so that $\beta \propto \rho/B^2 \propto 1/\rho$. The accretion flow is far more complicated than that, but we use this simple relation as the basis for our ansatz for β. Guided by this asymptotic consideration, we choose the following ansatz for $\beta^\dagger$:

$$\beta^\dagger = \bar{\beta}\left[\alpha\left(\frac{11\Delta x}{r_{\rm BH}}\right)\right]^{\chi}. \tag{B.7}$$

For the length scales $r_{\rm BH}$ and $r_{\rm ABH}$, the input quantity $\beta^\dagger$ is used in the expressions given by Equations (4.8) and (4.23), so Equation (B.7) an implicit function for $\beta^\dagger$. Note that, in contrast to the prescription for $\rho^\dagger$, the prescription for $\beta^\dagger$ does not require an explicit piecewise transition with resolution: the piecewise prescription for $\rho^\dagger$ transitions our estimate to a totally absorbing sphere in the limit of high resolution, and by definition this is insensitive to the field strength.

It remains to give an expression for $\chi(\beta^\dagger)$. We construct this function so that the functional form of the density profile (Equation B.6) best reproduces the azimuthally-averaged steady state density profiles in Cunningham et al. (2012). The functions are fit to the profile in the equatorial plane, defined as the plane perpendicular to the original magnetic field direction that also goes through the center of the sink particle. After some experimentation, we obtained a reasonably good fit with

$$\chi(\beta^\dagger) = \begin{cases} 0 & : \ \log_{10}\beta^\dagger < -3.1, \\ 1.27 - 0.5/(\beta^\dagger)^{0.13} & : \ -3.1 \le \log_{10}\beta^\dagger \le 2.0, \\ 1 & : \ \log_{10}\beta^\dagger > 2.0. \end{cases} \tag{B.8}$$

This function is a monotonically increasing function of $\beta^\dagger$. With this final parameter specified, Equations (B.7) and (B.8) are solved simultaneously by iteration until $\beta^\dagger$ converges to one part in 10^4 or until $\beta^\dagger > 10^9$. At this point, the four inputs $\rho^\dagger$, $\beta^\dagger$, $\bar{\mathcal{M}}$, and $\bar{\theta}$ are known and $\dot{M}_{\rm fit}$ can be determined.

We now show that this formulation satisfies the criteria given above. First, in the high-resolution limit, the sink should become totally absorbing, which requires that $\alpha^{-\chi}$ be of order unity. In this case, the argument of α is $11r_{\rm ABH}/(8r_{\rm BH})$, so that $\alpha^{-\chi} \to [\alpha(11/8)]^{-1} \simeq 1/2$ in the high-β limit. In the low-β limit, $\chi \to 0$, so that $\alpha^{-\chi} \to 1$. For intermediate values of β, α depends on both β and $\mathcal{M}$ as shown in Figure B.2. The smallest value of $\alpha^{-\chi}$ occurs for $\mathcal{M} = 0$; it is 0.064 at $\beta = 0.022$.

Next, for under-resolved flows ($\Delta x > r_{\rm ABH}/8$), α is evaluated at $r_{\rm avg}/r_{\rm BH} = 11\Delta x/r_{\rm BH}$ and is resolution dependent. We argued that in the weak-field limit, $\alpha^{-\chi}$ should approach $[\alpha(r_{\rm avg}/r_{\rm BH})]^{-1}$; this occurs naturally, since $\chi \to 1$ in this limit. In the strong-field limit, we required $\alpha^{-\chi} \simeq 1$; this is satisfied since $\chi \to 0$ in this limit. At intermediate values of β, we suggested that $\alpha^{-\chi}$ should have a minimum. Although we have not portrayed $\alpha^{-\chi}$ for different values of Δx in Figure B.2, this figure does show the expected minimum when Δx is at the boundary between low and high resolution ($\Delta x = r_{\rm ABH}/8$). As one moves into the low-resolution regime (increasing Δx), the argument of α increases and so does $\alpha^{-\chi}$. As a result, the values of $\alpha^{-\chi}$ for the high-resolution case in Figure B.2 provide a lower bound for the values in the low-resolution case.

B.2.2 The Adjusted Accretion Rate, $\dot{M}_{\rm sink}$: Capping the Alfvén Velocity

With $\dot{M}_{\rm fit}$ given by the prescription above, mass is extracted from a sink region within $4\,\Delta x$ of the particle as an operator-split source term that is applied every fine AMR level time step increment Δt. In well-resolved accretion flows care must be taken in extracting mass from the grid. In such cases we do not want to introduce a new local maximum in the speed of magnetosonic waves—similar to what was described in Section 4.3. In the opposite case of poorly-resolved accretion flow (e.g when $r_{\rm ABH}$ is not resolved) not introducing a new maximum in the value of $v_{\rm A}$, could arbitrarily diminish the accretion rate onto the sink particle. We therefore define a characteristic square-velocity V^2 as the maximum of two quantities, depending on whether $r_{\rm ABH}$ is resolved ($r_{\rm ABH} < \Delta x$) or not ($r_{\rm ABH} \geq \Delta x$):

$$V^2 = \max \begin{cases} \bar{v}^2_{\rm A,max} & : \quad r_{\rm ABH} < \Delta x \\ \dfrac{\bar{v}^2_{\rm A,max}\Delta x^2}{r^2_{\rm ABH}} & : \quad r_{\rm ABH} \geq \Delta x \end{cases} . \tag{B.9}$$

The value for $\bar{v}^2_{\rm A,max}$ is computed by taking the maximum Alfvén speed inside the same spherical shell described above. This defines the $\Delta\rho$ that can be extracted while only introducing a new local maximum in the Alfvén speed when $r_{\rm ABH}$ is not resolved and avoiding vanishing time-step pathologies when it is resolved. The value of $\Delta\rho$ extracted from a particular cell near the sink particle is set to the minimum of two quantities while holding the specific kinetic energy of the gas constant:

$$\Delta\rho = \min \left\{ \begin{array}{lcl} (\dot{M}_{\rm fit}\Delta t/\Delta x^3)W(r) & : & \text{Mass accretion estimated from fit.} \\ \rho - B^2/(4\pi V^2) & : & \text{New maximum Alfvén velocity avoided.} \end{array} \right| \;, \tag{B.10}$$

where r is the distance of that cell's center from the sink particle and the function $W(r)$ is a Gaussian kernel that extends out to $r = 4\,\Delta x$ and is normalized to unity (Krumholz et al. 2004b). Note that the second expression for $\Delta\rho$ is non-negative by definition. These two measures in the piecewise definition of $\Delta\rho$ are guided by physical considerations. However, if the sink particle accretes faster than the background flow can supply material, a void will open around the sink particle. If the density contrast between the void and the surroundings is allowed to become arbitrarily deep, the stability of the hydrodynamic scheme could be adversely impacted. This is particularly true in the limit of $\beta \to \infty$, where our Alfvén cap would not prevent this pathology. Therefore, we further impose the constraint that if the $\Delta\rho$ results in a particular cell having density less than $\rho^\dagger/10$, then $\Delta\rho$ is adjusted so that the cell's density is floored at $\rho^\dagger/10$. Finally, mass and momentum that is extracted from the grid is added to the sink particle mass in a manner than preserves global mass conservation,

$$\Delta M_{\rm sink} = \sum_{r\,\leq\,4\,\Delta x} \Delta\rho\,\Delta x^3 \;, \tag{B.11}$$

and the momentum of the sink particle is updated in a likewise manner that preserves global momentum conservation.

B.2.3 Verifying the Algorithm for the Accretion Rate

We have implemented this MHD sink particle algorithm in the `ORION2` code (Li et al. 2012). To test the method, we repeat the models of this work and Cunningham et al. (2012) on geometrically nested meshes having a base grid over $16r_B$ of at least 64^3 and enough AMR levels so that we coarsely resolve the accretion scales to be $r_B/\Delta x = 2,\ 8,\ 32,$ and 64 on the finest level. In Figure B.3 we show the comparison between the `ORION2` results and those from `RAMSES` . In general, we achieve accretion rates that are typically within a factor of two of the result obtained from high-resolution `RAMSES` models. In Figure B.4 we show the convergence properties in `ORION2` of a $\mathcal{M} = 1.41$, $\beta = 1$ model with a parallel field orientation. The parameters of this test case were chosen so that magnetic, thermal and ram pressure effects are all of comparable importance and that the influence of all of these effects on $\dot{M}$ converge with sufficient resolution. Note that at low resolution, the accretion rate is $\sim 40\%$ low and that the method transitions toward a pressure-less totally absorbing sphere by $r_{\rm B}/\Delta x > 64$, converging at high resolution to within 11% of the `RAMSES` model. The $\sim 11\%$ difference when `ORION2` is at comparable or higher resolution reflects intrinsic differences in the codes when resolving flows with a finite resolution.

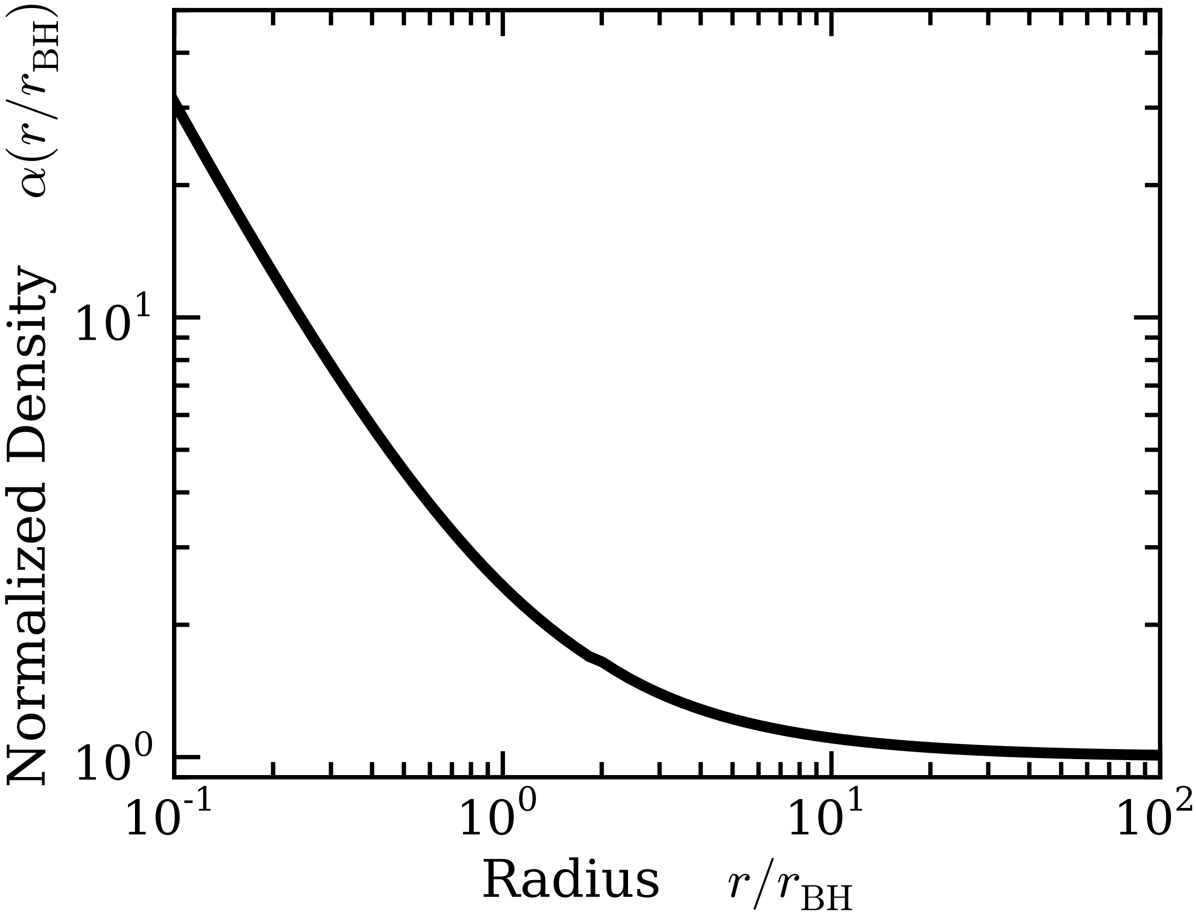

Figure B.1: Density profile from Krumholz et al. (2004b) for steady-state Bondi accretion onto a point source as a function of distance from the source r. The density is normalized to $\rho^\dagger = \rho(r \to \infty)$.

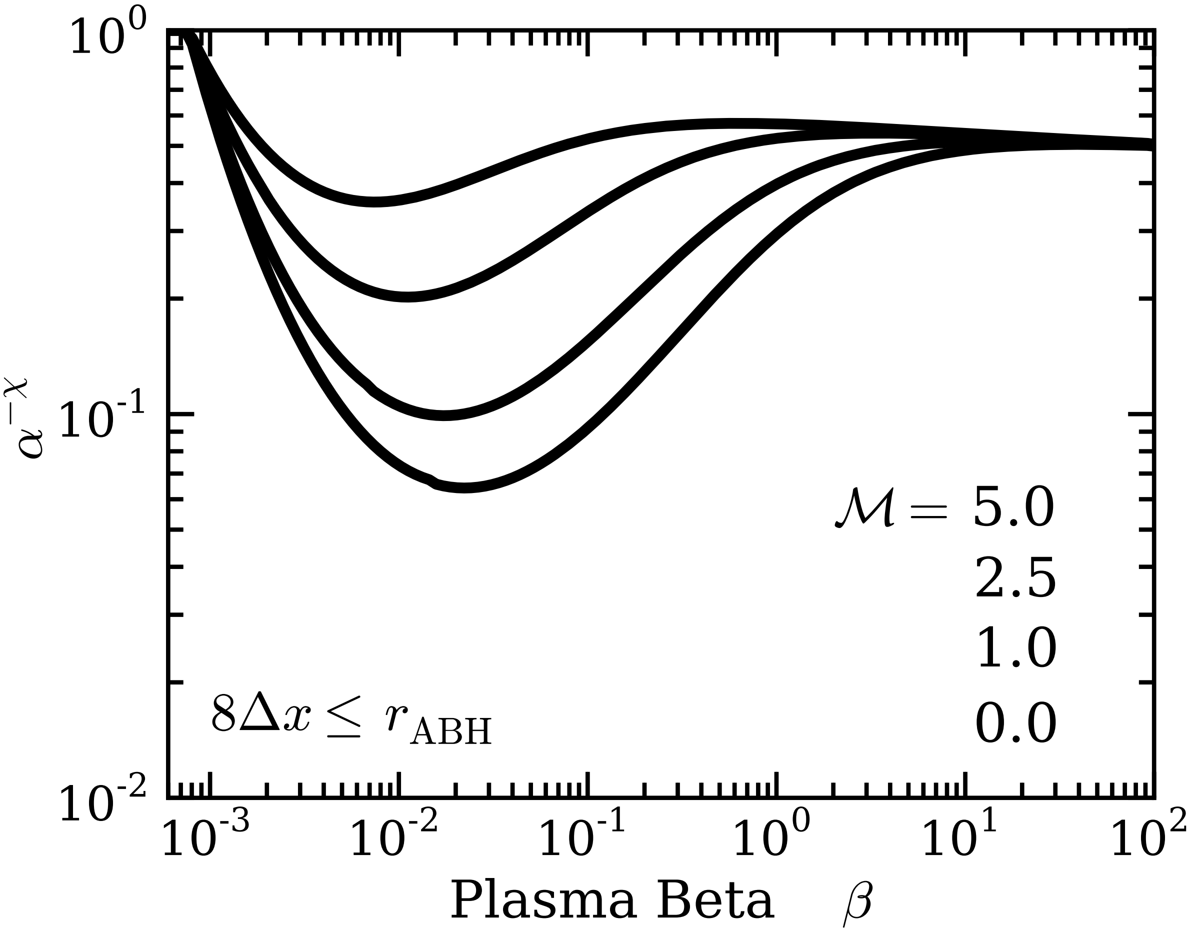

Figure B.2: The value of $\alpha^{-\chi}$ from Equation (B.6) for high resolution ($8\Delta x \leq r_{\rm ABH}$). Four different Mach numbers are considered, increasing from bottom to top. The function for $\chi(\beta)$ is given in Equation (B.8). For low resolution flows ($8\Delta x \geq r_{\rm ABH}$), $\alpha^{-\chi}$ is a monotonically decreasing function as Δx decreases. Therefore, these curves also give the minimum value of $\alpha^{-\chi}$ for a particular β and $\mathcal{M}$ in the low resolution limit.

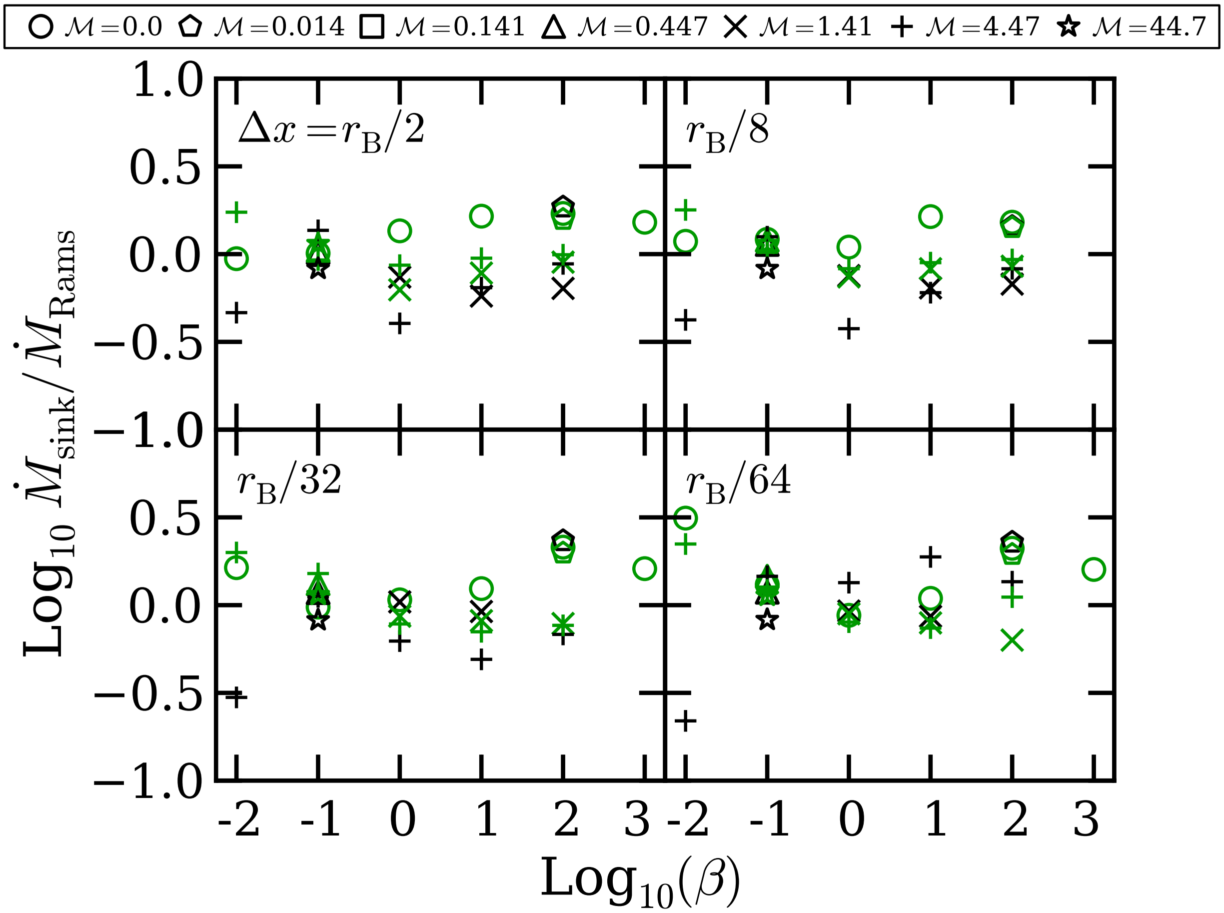

Figure B.3: Test of the magnetized sink particle algorithm implemented in `ORION2` . Plotted for each model is the ratio of steady-state accretion rates of `ORION2` ($\dot{M}_{\rm sink}$) and `RAMSES` ($\dot{M}_{\rm Rams}$) as a function of the smallest grid cell size Δx. Black points show the parallel orientation runs, green shows perpendicular. The Mach number of the run is given by the symbol.

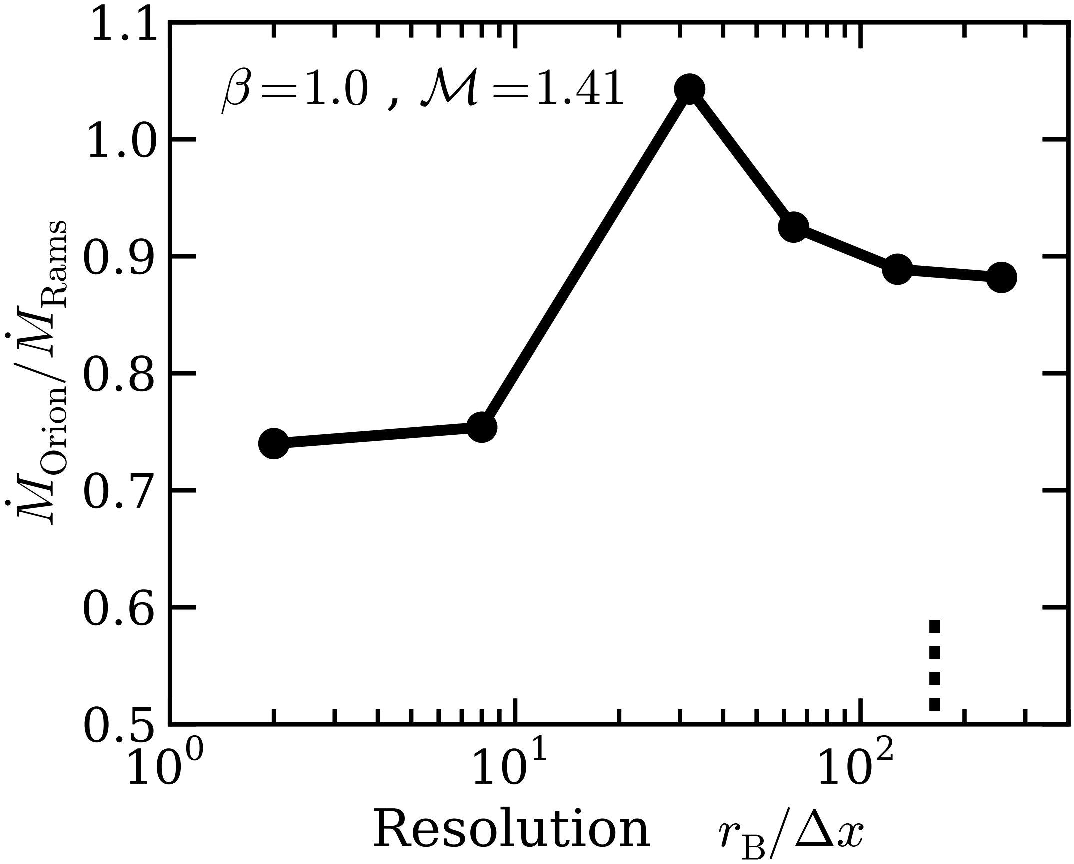

Figure B.4: Convergence study of the `ORION2` implementation as a function of the number of grid cells per $r_{\rm B}$. Plotted is the ratio of steady state accretion rates of `ORION2` versus `RAMSES` ($\dot{M}_{\rm Rams}$) for the $\mathcal{M} = 1.41$, $\beta = 1.0$ parallel case. For comparison, the default resolution `RAMSES` model had $r_{\rm B}/\Delta x \approx 164$, marked by the short dashed line.

Appendix C

Ray Tracing and Ionization

[1] Radiation from stars can strongly affect the surrounding environment. This radiation plays a crucial role in topics such as stellar atmospheres, the interstellar medium, star formation, galaxy formation, supernovae, and cosmology. Radiative transfer is a well-studied problem (e.g., Mihalas & Mihalas 1984; Rybicki & Lightman 1986), but its numerical treatment in multi-dimensional non-equilibrium calculations is difficult. Radiative processes depend on seven variables: three spatial, two angular, one in frequency, and one in time. This radiation is not uniform, and the nonlocal thermal and hydrodynamical response to intense radiative sources further adds to the difficulty of this problem.

Several methods have been developed to model radiation diffusing through optically thick gas. In this case where the radiation is strongly coupled to the hydrodynamics, flux-limited diffusion methods (Krumholz et al. 2007), M1 methods (first-moment methods; González et al. 2007), and variable Eddington tensor methods (Aubert & Teyssier 2008) have been employed, often using moment equations and assumptions of symmetry to reduce the number of independent variables. These flux-limited and moment methods have the advantage of being fast and independent of the number of radiation sources, although can become diffusive and result in incorrect solutions when inappropriately employed. An additional restriction is that these methods typically assume local thermodynamic equilibrium, which restricts the methods ability to capture scattering and multi-frequency radiative-hydrodynamics.

When the radiation is not strongly coupled to the hydrodynamics, long characteristic ray tracing methods are often used instead. These methods solve the radiative transfer equations along specific rays (e.g., Abel et al. 1999; Razoumov & Scott 1999; Susa 2006; Wise & Abel 2011). By tracing rays on a cell by cell basis, ray tracing can provide the highest accuracy for radiation emitted directly from point sources. The challenge with these methods is that they are often computationally expensive, since a large number of rays is typically necessary to hit distant cells or particles in a statistically significant manner. The problem does not

---

[1] Large portions of this appendix will eventually be published as Lee, A. T., McKee, C. F., Stacy, A. R., Rosen, A.R., Klein, R.I. Radiative Feedback in the Formation of Population III Stars I. HII Regions. in the Monthly Notices of the Royal Astronomical Society. Appropriate permissions to reproduce it here have been obtained from the co-authors.

benefit from parallelization algorithms as much as one would hope, especially when the rays are emitted from a few sources and travel only a small fraction of the domain's volume. In these situations, processors without rays idle while a small number of processors handle the entire ray tracing workload. The computational cost scales linearly with the number of sources, rays traced, and grid cells with which the rays interact. Unfortunately, standard load balancing algorithms often focus on balancing the total number of cells across processors and do not focus on equalizing the ray workload.

In Rosen et al. (2017), we recently implemented a highly-parallelizable long characteristic method to treat radiation from point sources. Section C.1 briefly describes our ray tracing method in ORION2, and Section C.2 describes how we have used this ray tracing method to trace the hydrogen-ionizing radiation from stars. Finally, Section C.3 discusses several tests we performed to assess the robustness of our method with regards to ionization.

C.1 Ray Tracing Method

We have implemented a long characteristic ray tracing method in the ORION2 adaptive mesh refinement code. Great care has been taken to optimize the parallel performance of the algorithm, which can operate on patched-based adaptive grids. In short, our parallel performance employs a completely asynchronous and non-blocking communication algorithm, which allows us to achieve near-perfect scaling up to $\mathcal{O}(10^3)$ processors on distributed memory machines.

Since the ray tracing is part of the AMR code ORION2, it is worthwhile to briefly describe the order of operations and domain decomposition in ORION2. The domain is decomposed into a set of nested grids with different cell sizes, with $l = 0$ denoting the coarsest level (the 'base-grid') and $l = l_{\max}$ denoting the finest level. ORION2 load balances using the knapsack algorithm, where each processor shares a roughly equal number of cells on each level. As a result neighboring and nested grids need not exist on the same processor. Grids on a given level are non-overlapping, and the grids are properly nested such that a cell on level l may have as its neighbor another level l cell or a cell of level $l-1$ or $l+1$, but not a cell of any other level. Each AMR level advances at a time step Δt_l, such that $\Delta t_l > \Delta t_{l+1}$. In ORION2, we use a refinement factor of 2, making $\Delta t_{l+1} = 0.5\Delta t_l$, and similarly for cell sizes. A given level with $l > 0$ synchs in time with the $l-1$ level every two time steps.

For simplicity of example, assume here that $l_{\max} = 2$; what is described below is easily generalizable to an arbitrary number of levels. In a simulation with only hydrodynamics, a single hydrodynamical coarse time step proceeds as follows. If the time at the beginning of the time step is t and the time step on the base grid is Δt_0:

1. Update: $l = 0$ grids are advanced Δt_0. Cells on level 0 are now advanced to $t + \Delta t_0$.

2. Update: $l = 1$ grids are advanced Δt_1 ($= \Delta t_0/2$). Cells on level 1 are now advanced to $t + \Delta t_1 = t + 0.5\Delta t_0$.

3. Update: $l = 2$ grids are advanced twice at Δt_2. Cells on level 2 are now advanced to $t + 2\Delta t_2 = t + \Delta t_1$.

4. Synchronize: The cells on levels 1 and 2 have now advanced to the same moment in time. The $l = 1$ grids with nested $l = 2$ grids are then synchronized; the finer grids are interpolated up to the $l = 1$ cells.

5. Update: $l = 1$ grids are advanced Δt_1. Cells on level 1 are now advanced to $t + 2\Delta t_1 = t + \Delta t_0$.

6. Update: $l = 2$ grids are advanced twice at Δt_2. Cells on level 2 are now advanced to $t + 4\Delta t_2 = t + \Delta t_0$.

7. Synchronize: The cells on levels 1 and 2 have now advanced to the same moment in time. The $l = 1$ grids with nested $l = 2$ grids are synchronized;

8. Synchronize: The cells on levels 0 and 1 have now advanced to the same moment in time. The $l = 0$ grids with nested $l = 1$ grids are synchronized.

9. The next coarse time step is determined and the process repeats until the desired end time is reached.

Synchronization interpolates information from the finer cells up to the coarser cells before the coarser cells are advanced again.

Ray tracing occurs during the finest level advances. In the example above, ray tracing is called $2^{l_{max}} = 4$ times for every coarse time step. As rays travel through the domain, they interact with the finest grid available at a given location.

One computational expense of previous ray tracing methods is that a large number of rays are needed adequately sample cells that are traversed by rays. When cells have a characteristic size Δx, then $\sim 4\pi(r/\Delta x)^2$ rays are needed from the onset to ensure that every cell within a radius r from the source is hit by at least one ray. In the optically thin limit, the $(r/\Delta x)^2$ factor can be a significant memory and computational request. Additionally, it is challenging to know *a priori* how far the rays will travel during a single ray trace. To ease the computational burden and assure ourselves that rays adequately sample the domain they traverse, we instead use a method similar to the method of Abel & Wandelt (2002) and Wise & Abel (2011), where we launch a smaller number of rays from each source, allowing the rays to divide as they travel. The initial directions of the rays are chosen using the Hierarchical Equal Area isoLatitude Pixelization of the sphere (HEALPix) to sample the full 4π steradians (Górski et al. 2005). Rays subdivide into four child rays once a division criterion is met.

We define a pixel level j for each ray, which increases as the rays divide. The minimum level, given by the user, determines the number of rays that are initially launched from the source. The pixel number $N_{\rm pix} = 12 \times 4^j$ designates the total number of rays that exist if all rays from a particular source are on the same pixel level j. Every time the ray subdivides, the level increases by one. Typically we set $j_{\mathrm{min}} = 3$ or 4, meaning that 768 or 3072 rays are initialized for each source. The solid angle sampled by a ray on level j is $\Omega_{\mathrm{ray}} = 4\pi/N_{\mathrm{pix}}(j)$. A given ray splits when the ratio

$$\frac{\Omega_{\mathrm{cell}}}{\Omega_{\mathrm{ray}}} = \frac{(\Delta x/r)^2}{4\pi/N_{\mathrm{pix}}(j)} < \Phi_c \ ,$$

where Φ_c is a fixed constant, which we set to four.

When the rays are initialized, the base luminosity is equally divided amongst the rays, and the current luminosity is equally split amongst child rays when splitting occurs. Rays do not merge when they overlap. As rays travel from cell to cell, the path-length Δl through each cell is used to calculate how much radiation is absorbed by the cell. Once the luminosity of the ray decreases below 0.01% the luminosity of the base ray, the ray is terminated. The reduction in luminosity from ray splitting is not included in this termination criterion. If rays make it to the edge of the domain, the rays are also terminated.

Our parallelization scheme addresses several serial bottlenecks and is described in detail in Rosen et al. (2017), see particularly Algorithm 1 of that paper. Since the number and position of sources and the distribution AMR grids in memory are not known in advance, the ray tracing workload is never equally spread among the processors, and a given ray may need to be communicated between several processors as it propagates. By using non-blocking and asynchronous communication, available since version 3.0 of the Message Passing Interface standard library, a processor can handle any rays that currently exist in its memory, communicating and receiving rays without blocking other processors from working on their own rays. This eliminates processors from idling because they are waiting for communication between other processors to complete. The accuracy of this adaptive ray tracing algorithm, along with the results of weak and strong scaling tests, are shown in Rosen et al. (2017). Below, we detail our implementation of ionization physics, which has not been discussed elsewhere.

C.2 Ionization

At a basic level, our ionization ray tracing scheme tracks the ionizing emission from stars, which are represented by sink particles. Each ray solves the radiative transfer equation

$$\frac{\partial I}{\partial t} + \nabla I \cdot \mathbf{n} = -\kappa I \ ,$$

where I is specific intensity as seen in the direction defined by $\mathbf{n}$. The right-hand side includes only a sink term with an absorption cross section κ, which accounts for the attenuation of the radiation field by the medium. For a ray traveling from a source in the radial direction r, it is convenient to write this equation in terms of photon flux P. Integrating over the solid angle of the ray with a closed surface containing the emitting source, we arrive at

$$\frac{1}{c}\frac{\partial P}{\partial t} + \frac{\partial P}{\partial r} = -\kappa P$$

(e.g., Rosen et al. 2016; Baczynski et al. 2015). This expression can be a function of wavelength; we will assume hydrogen-ionization at a single wavelength for now (see Rosen et al. 2017, for how this can be done).

Rays travel through the domain on a cell-by-cell basis, considering the highest level cell at any given location. Ray tracing methods typically consider each cell as a uniform block of gas with optical depth $\tau = \kappa\rho dl$, for the path length through the cell dl. Since each ray has with it an associated solid angle $\Omega_{\rm ray}$, the ray reaches a cell at a distance r with a cross sectional area $A = r^2\Omega_{\rm ray}$. We assume that the 'volume' of the gas being ionized by the ray is $A\,dl$, which is approximate when the ray hits the cell at an angle and when the ray passes near a cell corner. For photon-conserving schemes like our own, this introduced error is compensated for by an equal and opposite error in adjacent cells. Since rays are launched in random directions from the source, subsequent ray traces reduce the overall error statistically.

For a uniform cell of gas, a number flux P enters the cell and $Pe^{-\tau}$ leaves, and the number flux reduces by $dP = P(1 - e^{-\tau})$. The sum of dP for every ray that passes through the cell gives the total ionization rate for the cell. Equating $-\sum dP/\Delta x^3$ with $\dot{n}_{\rm H}^0$, the change in the neutral atomic hydrogen density is $\Delta n_{\rm H}^0 = \dot{n}_{\rm H}^0\Delta t$. As long as $n_{\rm H}^0 + \Delta n_{\rm H}^0 \geq 0$, this approach is a photon-conserving scheme. Knowing the energy released per ionization also allows us to calculate the total heating; in addition, knowing the direction of the original rays allows us to calculate the total momentum imparted to the gas.

Sub-cycling on the finest time step can be done to prevent abundances and the total energy of the cells from changing too much between hydrodynamic updates. For example, during a single ray trace, Wise & Abel (2011) used the previous ray trace to set the subsequent time step to be

$$\Delta t_{\rm ray} = \min\left(\frac{0.1 n_e}{|dn_e/dt|}, \frac{0.1 n_{\rm H}^0}{|dn_{\rm H}^0/dt|}, \frac{0.1 e}{|de/dt|}, \frac{\Delta t_{\rm hydro}}{2}\right)$$

where n_e is the electron number density, e is the gas's specific energy, and $dt_{\rm hydro}$ is the hydro time step on the finest level. This limits the change of relevant species or energies to no more than 10% and requires at least two ray trace calls per fine time step.

These approaches are not without their flaws. When calculating the optical depths of the cells, typically the original state is used for every ray that passes through the cell, thus overestimating the number of absorptions when multiple rays pass through. This risks breaking the photon conserving nature of the rays in cases where more ionizing photons are absorbed than there are neutral hydrogen atoms. In this case, to avoid negative number densities, the code resets the number density to zero, throwing away photons that would have continued to propagate. Furthermore, such sub-cycling requirements are typically over-restrictive, requiring an unnecessary number of sub-cycles for cells with low optical depths.

To alleviate these problems, we have developed a sub-grid model for our radiative transfer scheme. Other approaches use a method like the one above, where a net ionization rate is determined by the ray trace, using the original state of the gas. This rate is then multiplied by Δt to determine to the total number of ionizations. In reality, the state of the gas continuously changes over Δt. Our approach below addresses this issue.

C.2.1 Sub-Grid Model

In general the equation for the number density of neutral hydrogen $n_{\rm H}^0$ is

$$\frac{\partial n_{\rm H}^0}{\partial t} = -\nabla \cdot n_{\rm H}^0 \mathbf{v} + \alpha_B n_{\rm e} n_{\rm H}^+ - C_H n_{\rm e} n_{\rm H}^0 - \sum_{\text{rays } i} F_i \sigma_H n_{\rm H}^0 \ ,$$

where α_B is the recombination coefficient, assuming the on-the-spot approximation, C_H is the collisional ionization coefficient, F_i is the ionizing flux of ray i passing through the cell, $\sigma_{\rm H}$ is the ionization cross section, $n_{\rm e}$ is the electron number density, and $n_{\rm H}^+$ is the number density of ionized atomic hydrogen.

Considering only a single ray and that the time step is small enough so the various terms can be computed via an operator split technique, the ionization term becomes

$$\frac{\partial n_{\rm H}^0}{\partial t} = -F \sigma_{\rm H} n_{\rm H}^0 \ .$$

The flux here is the flux arriving at the cell after traveling some distance r from the source. We now integrate over the volume of the cell, setting $F = 0$ outside the ray. Recalling that the volume of the ray is $\sim A\,dl$, we arrive at[2]

$$\frac{\partial \mathcal{N}_{\rm H}}{\partial t} = -AF(1 - e^{-\tau}) \ ,$$

where $\tau = n_{\rm H}^0 \sigma_{\rm H} dl$ is the optical depth of the ray in the cell and $\mathcal{N}_{\rm H}$ is the total number of neutral hydrogen atoms inside the cell.

We now determine the change of neutral hydrogen over a time step $\Delta t_{\rm ray}$ without assuming that the change of neutral hydrogen is small. Since $n_{\rm H}^0 = \mathcal{N}_{\rm H}/\Delta x^3$, we have

$$\frac{\partial n_{\rm H}^0}{\partial t} = -\frac{AF}{\Delta x^3}\left(1 - e^{-\tau}\right) \ ,$$

so that

$$\frac{\partial \tau}{\partial t} = -\frac{AF\sigma_{\rm H} l}{\Delta x^3}\left(1 - e^{-\tau}\right) \ .$$

If τ_0 is the optical depth along the ray through the cell at time t and $\tau_0 + \Delta\tau$ is the optical depth at $t + \Delta t_{\rm ray}$, then one can integrate over time to obtain

$$\Delta\tau = \sigma_{\rm H} \Delta x \Delta n_{\rm H}^0 = \ln\left[e^{-\tau_0} + (1 - e^{-\tau_0}) e^{-\Delta t'}\right] \ , \tag{C.1}$$

---

[2] Since the rays are diverging, accuracy can be improved by replacing A with $A(1 + l/r)$.

where

$$\Delta t' = \left(\frac{AF\sigma_{\rm H} l}{\Delta x^3} \right) \Delta t_{\rm ray}$$

is the time measured in units of the mean ionization time of an atom in the cell. It is easy to see that the argument of the logarithm is less than unity so that $\Delta\tau < 0$, as it must be for ionization. One can further show that $\tau_0 + \Delta\tau > 0$, i.e., it is therefore impossible to over-ionize the cell if the neutral abundance is changed in accordance with this equation.

The number of photons removed from the ray exactly matches the number of ionizations. Subsequent rays see an altered cell upon arrival and calculate a reduced τ_0. Overall, the scheme remains photon conserving regardless of the size of the time step.

C.2.2 Algorithm

For each ray trace call, we perform the following:

1. Enter the ray trace routine with the current time t and the finest hydro time step Δt. Set the time needed in ray tracing to $t_{\rm ray} = \Delta t$. The variable $t_{\rm ray}$ represents the amount of time remaining before the ray trace is completed for the current hydrodynamical time step. A temporary array is created to store the total number of ionizations that occur in each cell.

2. Beginning of main loop: Particles store the value of the next ray trace time step $\Delta t_{\rm ray}$. If it exceeds $t_{\rm ray}$, set $\Delta t_{\rm ray} = t_{\rm ray}$. If this is the first time ray tracing is being called in the simulation, set $\Delta t_{\rm ray} = 0.1 t_{\rm ray}$. Particles calculate the ionizing luminosity, assuming their luminosity is constant over this $t_{\rm ray}$. ORION2 stores integrals of the Planck equation for rapid table lookups.

3. Subtract $\Delta t_{\rm ray}$ from $t_{\rm ray}$.

4. Particles cast their rays. Rays are moved via the ray tracing algorithm described above.

5. For every cell a given ray passes through, Equation (C.1) is used to calculate $\Delta n_{\rm H}^0$, and $\Delta n_{\rm H}^0 \Delta x^3$ photons are removed from the ray. Currently we assume a monochromatic spectrum and use a constant photoionization cross section $\sigma_{\rm H} = 6.3 \times 10^{-18}$ cm^2.[3] In the future, we will allow each ray to calculate $\sigma_{\rm H}$ based on the ray's average photon energy. The ray's current flux F and $\Delta t_{\rm ray}$ are used in calculating $\Delta t'$. The value of τ is calculated using the properties of the cell and the path length of the current ray. If the temporary array entry for this cell is non-zero, the ionizations from previous rays are incorporated when calculating τ.

---

[3] This value for $\sigma_{\rm H}$ is the cross section for ~ 13.6 eV photons. In general $\sigma_{\rm H} \propto (h\nu/13.6 \text{ eV})^{-3}$ until the energies become comparable to Compton scattering; for hydrogen, this occurs at $h\nu \approx 2.5$ keV (Draine 2011). For main-sequence stars below $\sim 20 M_\odot$, the average photon energy above 13.6 eV remains below 20 eV, where $\sigma_{\rm H}$ has dropped only to $0.3\sigma_{\rm H}(h\nu = 13.6 \text{ eV})$.

6. After the total number of ionizations are computed for a given ray passing through a given cell, the number of ionizations are added to the temporary array.

7. Secondary loop: Once the ray casting has finished, ionizations and recombinations are applied to change the actual state of the gas. Each processor loops over every cell it owns and applies ionization chemistry and heating. We assume that 2 eV is imparted to the electron gas per ionization. Ionizations are applied first, and then recombinations are applied.[4] The recombination coefficient is calculated as

   $$\alpha_B(T) = 2.59 \times 10^{-13} \left(\frac{T}{10^4}\right)^{-0.833-0.034\log(T/10^4)} \quad \mathrm{cm}^3/\mathrm{s}$$

   (Draine 2011), where the temperature of the gas is calculated using the resulting internal energy of the cell e after ionizations have been applied. Recombination cooling is incorporated assuming that

   $$\left(0.684 - 0.0416\log\left(\frac{T}{10^4}\right)\right) k_\mathrm{B}T \quad \mathrm{ergs}$$

   are lost per recombination. On average, this is $\sim 0.7 k_\mathrm{B}T$ but accounts for the fact that in hotter gas the higher energy electrons are harder to capture (Draine 2011). The net heating rate $\dot{e}$ for every cell is temporarily stored.

8. Compute the next ray trace time step $\Delta t_{\mathrm{ray},1}$ for each cell. The next time step for each cell as the minimum of either 10% of the recombination time scale $1/\alpha_B n_\mathrm{e}$ or 10% of the heating/cooling time scale $e/\dot{e}$. This concludes the second loop.

9. After the second loop is complete, the minimum value of $\Delta t_{\mathrm{ray},1}$ in the computational domain is the next ray trace time step Δt_ray. This time step is stored with the sink particles.

10. If $t_\mathrm{ray} = 0$, exit ray tracing. Else return to the top of the main loop, repeating the ray tracing routine until $t_\mathrm{ray} = 0$.

C.3 Tests

In this section, we detail two benchmark tests performed to assess the accuracy of the ionization ray tracing method. Additional tests not specific to ionization can also be found in Rosen et al. (2017). Our method can be directly compared to other ionizing ray tracing methods, for example, that of Wise & Abel (2011) and Baczynski et al. (2015), which perform similar tests.

---

[4] This approximation assumes that Δt_ray is small compared to t_rec. In the future we plan on improving the accuracy of this step by applying recombinations at half time-steps between incorporating ionizations.

C.3.1 Photon conservation and R-Type Ionization Front Expansion

Consider an initially uniform neutral medium where an embedded star suddenly turns on. The resulting ionization front initially expands much faster than the sound speed of the surrounding ambient gas, leaving no time for the photo-heated, overpressured gas to compress the surroundings. The front moves into neutral rarified gas and is commonly described as the R-Type expansion phase of the ionization front. Since this phase depends only on the state of the initial medium and the ionizing luminosity, this is an apt test to assess the ray tracing algorithm and the ionization and recombination rate calculations.

The ionization front will move outward as long as ionizing photons can make it to the front before being absorbed by recombining ionized gas inside the HII region. The radius where the recombination rate balances the ionization luminosity is the Strömgren radius

$$R_S = \left(\frac{4\pi Q}{3 n_{\rm H}^{+} n_{\rm e} \alpha_B}\right)^{1/3} \quad ,$$

where Q is the number of ionizing photons emitted per second (Strömgren 1939). The radius as a function of time can readily be derived, yielding

$$R(t) = R_S \left(1 - \exp(-t/t_{\rm rec})\right)^{1/3} \quad . \tag{C.2}$$

Here $t_{\rm rec} = (\alpha_B n_{\rm e})^{-1}$ is the recombination time. The exponential can be expanded to obtain

$$R(t) = \left(\frac{3Q}{4\pi n_{\rm H}^{0}}\right)^{1/3} t^{1/3} \tag{C.3}$$

as the leading term, which is the same expression you would derive for the location of the ionization front when $\alpha_B = 0$.

We perform a set of simulations for an isothermal gas with adiabatic index $\gamma = 1.0001$ that is comprised initially of neutral hydrogen. For these simple tests, the recombination coefficient is either set to 0 or 4.4×10^{-12} cm^3/s. The atomic hydrogen number density is set to $n_{\rm H}^0 = 100$ cm^{-3}. One source is placed at the center of the domain and emits $Q = 4 \times 10^{49}$ photons/sec. The domain has a base resolution of 64^3, which spans 12 pc in size. We set the heating and cooling rates in our ray tracing algorithm to zero. The simulation is run for 4000 years, which is a bit more than three recombination times. For these parameters, $R_S = 5$ pc.

Since the radiation time step is determined only by the number density, the evolution of the front should not depend on resolution. We have performed this test also with a base grid of 128^3 and with a base grid of 64^3, while allowing one or two levels of AMR refinement, and obtain identical results to those shown below.[5] Additionally, we have run this test by placing several particles within the innermost 8 grid cell at the center of the domain, adjusting the output luminosity so the total luminosity output equals $Q = 4 \times 10^{49}$ photons/sec. Once the

---

[5] For the AMR tests, refinement was applied where the gradient of the neutral hydrogen mass fraction exceeded 25%.

front's radius exceeds ~ 10 grid cells, the radial evolution of the front for this star cluster matches that of the single-star tests.

Figures C.1 shows example slices of the inner 10 parsecs of the domain, where the colorbar represents the neutral hydrogen fraction, for 64^3 and 128^3 grids. In general the ionization front is sharply defined and spans only a grid cell in length. The spherical symmetry depends on the total number of rays arriving at a certain radius; by adjusting the base number of rays launched from the source or by increasing the rate of ray refinement, the symmetry of the region can be further improved.

We use YT analysis tools (Turk et al. 2011b) to find the spherical radius where the ionization fraction first reaches 50%; this defines the ionization front. Figure C.2 plots the ionization front for different sets of simulation parameters. The left panel uses $\alpha_B = 0$ to test the photon-conserving nature of the ray trace. This is tested with different base resolutions and with three different ray refinement factors (2, 4–the default, and 8). We test our results against Equation (C.3) and find near-perfect agreement regardless of grid resolution and the refinement factor that determines when rays divide.

Second, we set α_B to the constant value 4.4×10^{-12} cm^3/s; these results are shown on the right panel of Figure C.2. Sub-cycling only depends on the recombination time scale criterion, which is initially set to $\Delta t_{\rm ray,max} = 0.1 t_{\rm rec}$. Put another way, this criterion requires ~ 10 ray traces per recombination time of a purely ionized cell. We consider more restrictive time stepping criterion up to requiring ~ 160 ray traces per recombination time. At early times, before recombinations begin to matter, all simulations agree with one another. Beyond $\sim 0.5 t_{\rm rec}$, our results generally overshoot the analytical solution but all asymptote to the correct value. If at least 20 ray traces are done per recombination, the ionization front is within one grid-cell of the analytical solution at all times. Similar results are found with higher resolution grids.

C.3.2 D-Type Ionization Front Expansion

Once the ionization front slows to twice the sound speed of the ionized gas c_i, the over-pressured gas launches a shock wave ahead of the ionization front, making the front move through dense post-shocked gas (this is called the "D-Type" expansion phase). Hosokawa & Inutsuka (2006) derived the solution for the ionization front during this phase, arriving at

$$R(t) = R_S \left(1 + \frac{7}{4} \sqrt{\frac{4}{3}} \frac{c_i t}{R_S} \right)^{4/7} . \tag{C.4}$$

This is similar to the solution originally derived by Spitzer (1978), which omits the square-root factor. Eventually, the expansion should come to rest once the pressure of the ambient medium matches the ever-decreasing pressure exerted by the ionized region (Raga et al. 2012). This D-Type phase of the ionization front expansion serves as a strong test for the hydrodynamical response to the overpressured HII region and the continuous driving of the shell by ionizing radiation.

This test was recently performed by several ray tracing codes in a benchmark study by Bisbas et al. (2015). We use initial conditions that match those described in this paper. To test the coupling of ray tracing and hydrodynamics, we measure our numerical solution against Hosokawa & Inutsuka (2006) at times before the pressure of the ambient gas becomes comparable to the ionized gas pressure, which Bisbas et al. (2015) refers to as the "early phase." To best reproduce the conditions assumed in the analytical solution, we assume that ~ 2 eV of heating occurs per ionization, and the recombination cooling removes $0.7kT$ per recombination. However, at the end of the ray tracing, before the hydrodynamic update is applied, we apply a ceiling to the temperature of the ionized gas at 10^4 Kelvin, which is the assumed constant temperature in Bisbas et al. (2015). This allows us to keep the ionized gas at a nearly constant temperature, which matches what is assumed in the analytical derivation of Hosokawa & Inutsuka (2006).

For these simulations, our domain is decomposed onto a 64^3 or 128^3 base grid that spans $8R_S$ in each dimension.[6] The gas is initially given a mass density of $\rho_0 = 5.21 \times 10^{-21}$ g/cm^3 with a temperature of 100 Kelvin. The central star emits $Q = 10^{49}$ ionizing photons/second. We run our simulations for 0.13 Myrs, which allows the ionization front to expand to $\sim 4R_S \sim 1.3$ pc in radius. At this point the shocked gas reaches the edge of the domain. Here we use both the recombination time scale and net energy injection time scale criterion when determining the ray trace time step. The location of the ionization front is determined using the same algorithm as the previous section.

Figure C.3 shows error between the numerical radius and the analytical radius of the ionization front as a function of time. Here, as is done in Bisbas et al. (2015), we define $t = 0$ as the time where the ionization front first reaches $R = R_S$ in size, which is approximately $2 - 3\ t_{\rm rec}$. Additionally, Figure C.3 shows the results of other 3D participating codes in Bisbas et al. (2015, taken from Figure 6 of that paper). The reader is encouraged to consult that paper for references and details regarding individual codes.

Overall, our code achieves excellent agreement with the analytical answer. For both resolutions we consider, the solution appears to asymptote to $98-99\%$ the radius calculated by Hosokawa & Inutsuka (2006). Figure C.4 shows example slices of the HII region at late times after the shock wave has propagated ahead of the ionization front. While we notice ray tracing artifacts in the ionization fraction from ray trace to ray trace, like what can be seen on the left panel of Figure C.4, the random orientation of the subsequent rays cast statistically reduce these errors and have little impact on the overall dynamics (even when the temperature ceiling is not applied). Furthermore, increasing the sub-cycling restriction from 10% to 25% or 50% of the recombination and heating time scales only increases the error by a few percent. In generally, our algorithm performs ~ 1000 ray traces per hydrodynamic time step[7] at early times, a result of an intense source suddenly heating neighboring cells of gas. Once the HII region begins to move outward, the ray trace time step is limited

---

[6] We note that the resolution in Bisbas et al. (2015) is equivalent to using a 256^3 grid, so we are simulating the front at a lower resolution.

[7] The hydrodynamical time steps are determined by the Courant condition: $\Delta t_{\rm hydro} \approx 0.5\Delta x / c_{\rm s}(T = 10^4\ {\rm K})$.

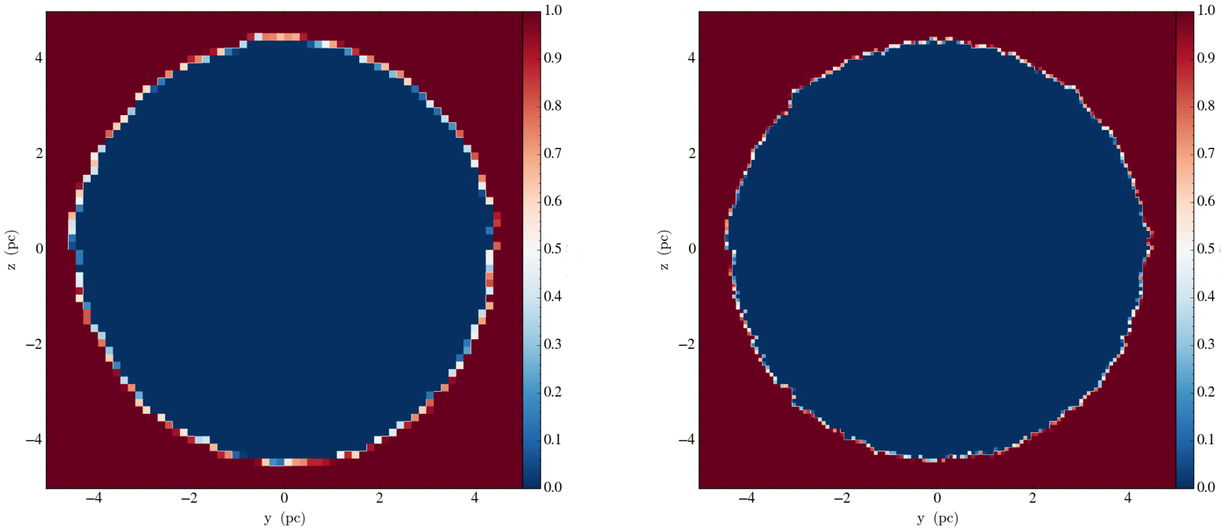

Figure C.1: Slices showing the neutral hydrogen mass fraction for (left) a 64^3 and (right) an 128^3 base grid. The width of the ionization front is approximately one grid cell. The spherical symmetry depends on the total number of rays arriving at a certain radius; by adjusting the base number of rays launched from the source or by increasing the rate of ray refinement, the symmetry of the region can be further improved.

primarily by the recombination time scale requirement. During this time, ~ 100 ray traces are performed per hydrodynamical time step for these coarse resolution runs.

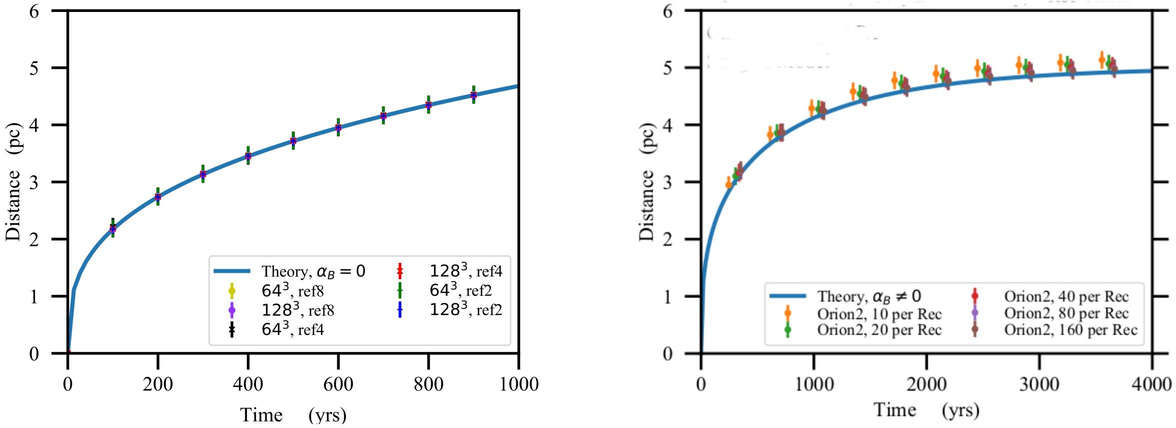

Figure C.2: Ionization front as a function of time. (Left) Without recombinations, our results nearly exactly match the analytical solution. The error bars show the length of one grid cell. The legend records the base grid resolution and the refinement factor ("ref") used in determining when the rays divide. (Right) With recombinations, our results are within one grid cell of the analytical solution at all times when at least 20 ray traces are done per recombination time. All runs are done with a 64^3 grid. For both plots, the ionization front is defined as the spherical radius where the ionization fraction first reaches 50%.

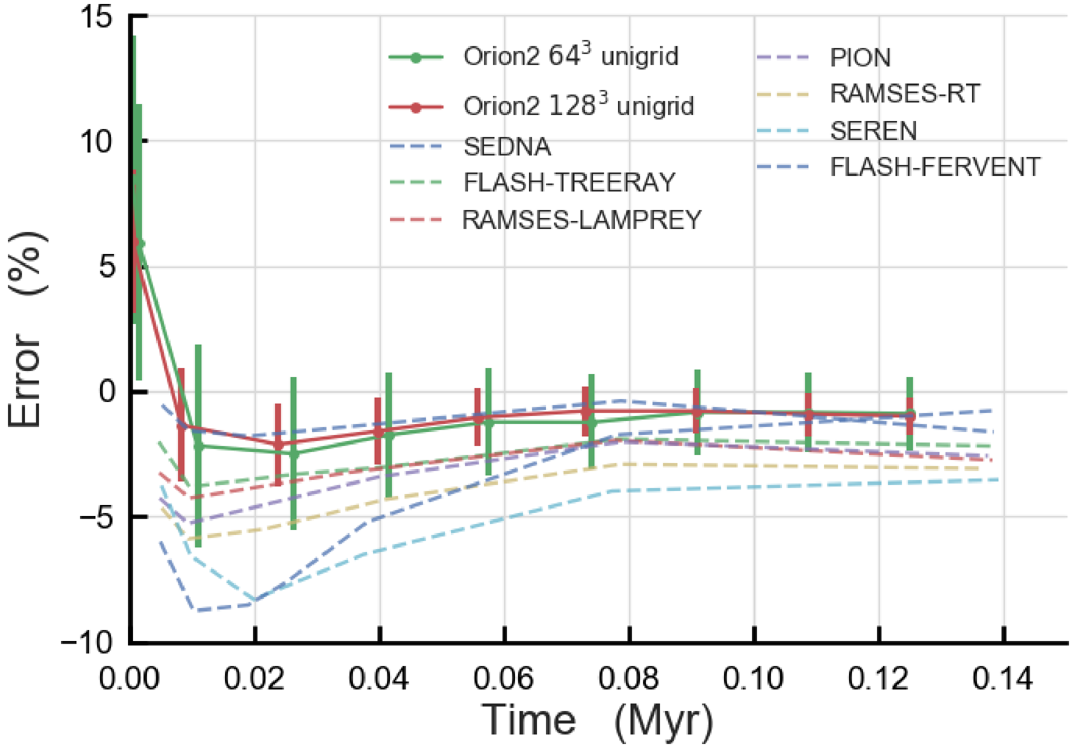

Figure C.3: Error between the numerical radius and the analytical radius of the ionization front as a function of time. The analytical solution is taken from Hosokawa & Inutsuka (2006), which is our Equation (C.4). Results of other codes that participated in Bisbas et al. (2015) benchmark test are also shown. The error bars assume the location of the front is known to within one grid cell. Overall, our numerical solution asymptotes to $98-99\%$ the analytical solution and generally agrees with the solution to better than 5%.

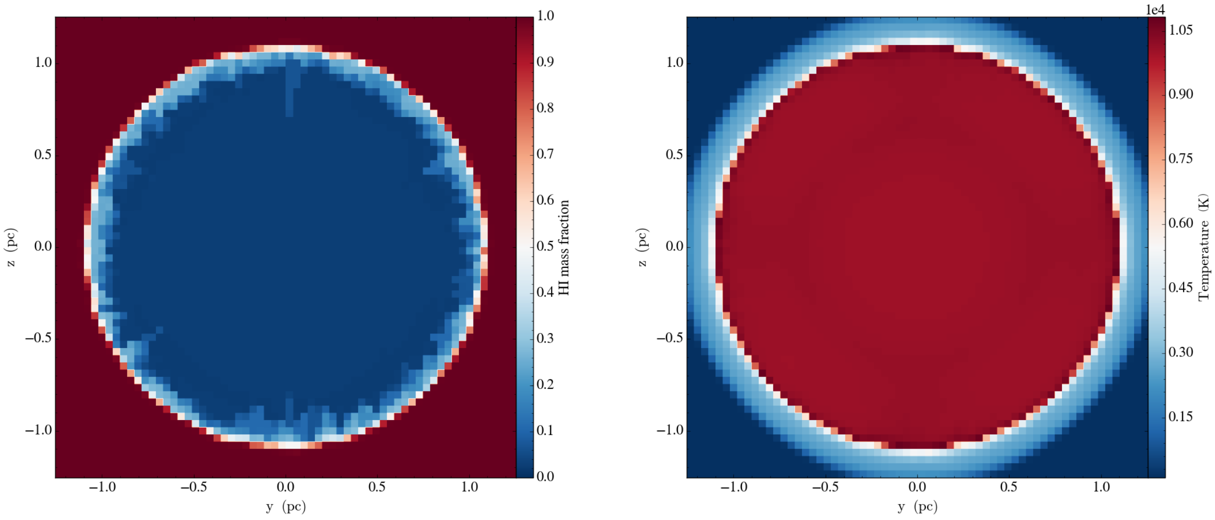

Figure C.4: Late-time slices showing the neutral H mass fraction (left) and the temperature (right) of the gas for one of our D-Type simulations. The ionized region maintains a nearly-constant temperature of 10^4 Kelvin, and a shock wave can clearly be seen propagating ahead of the ionization front. This example shows our 128^3 base grid run.

Appendix D

Monte Carlo Projection Code

[1] The major astronomical problems of computational hydrodynamics, including galaxy formation and evolution, the collapse of interstellar clouds and star formation, and the evolution of accretion disks, have been approached from both the Eulerian and Lagrangian viewpoint. The former utilizes a static grid, sometimes with the option of dynamically subdividing the grid if necessary; such an approach is often called Adaptive Mesh Refinement (AMR). The latter closely resembles the classic N-body problem and tracks 'particles' of fluid as they move through the computational domain. These particles represent a discrete bundle of the fluid's mass, and extend N-body methods by bestowing each particle with gas properties (e.g., pressure); such an approach is called Smoothed Particle Hydrodynamics (SPH). Each approach has its advantages and disadvantages, and the best computational approach depends on the given problem. For example, SPH codes are by design self-adaptive and do not require complex methods to provide high resolution in regions of high density, though the artificial imbibing of gas properties to the particles makes the treatment of shocks and turbulence considerably more difficult (e.g., Price & Federrath 2010).

In Chapter 4 of this thesis, we used the Eulerian code Ramses to simulate magnetized gas around young protostars. Chapter 5 uses the Eulerian code Orion2 (Li et al. 2012), which now incorporates a ray tracing module to track the ionizing radiation from stars (Appendix C), to study the formation of Population III stars. Future work on the formation of these stars will use initial conditions from the simulations of Stacy & Bromm (2013), which used the SPH code Gadget2 (Springel 2005). In this appendix, we describe a method to map the SPH results of Stacy & Bromm (2013) to uni-grid initial conditions for use in Orion2. Below we briefly describe some of the methods of SPH codes. Section D.1 discusses our projection routine, and Section D.2 describes our initial conditions and the results of the conservation tests we have performed.

SPH extends the N-body approach by assuming each particle represents not a point

---

[1] Large portions of this appendix will eventually be published as Lee, A. T., McKee, C. F., Stacy, A. R., Rosen, A.R., Klein, R.I. Radiative Feedback in the Formation of Population III Stars I. HII Regions. in the Monthly Notices of the Royal Astronomical Society. Appropriate permissions to reproduce it here have been obtained from the co-authors.

mass, but a smeared-out distribution of density

$$\rho_j(\mathbf{r}) = m_j\, W(|\mathbf{r} - \mathbf{r}_j|, h_j) \ , \tag{D.1}$$

where the density of particle j at some point $\mathbf{r}$ depends on the particle's total mass m_j, the location of the particle's center $\mathbf{r}_j$, and a scale parameter h_j (Price 2012). The functional form of the smoothing kernel W describes the extent to which the mass is spread out, and it should at least have the following properties: (1) W should be positive and have units of 1/volume, (2) W should monotonically decrease away from $\mathbf{r}_j$, (3) W should have smooth derivatives, (4) W should not be sharply peaked, and (5) the integral of ρ_j over the particle's volume returns m_j. The last property is equivalent to

$$\int_{V_j} W(|\mathbf{r} - \mathbf{r}_j|, h_j)\, dV = 1 \ . \tag{D.2}$$

At a given point in space, the total density is the summation over all $N_{\rm part}$ particles

$$\rho(\mathbf{r}) = \sum_{j=1}^{N_{\rm part}} m_j\, W(|\mathbf{r} - \mathbf{r}_j|, h_j) \ . \tag{D.3}$$

Property (4) prevents small changes in the position of nearby neighbors from affecting the value of $\rho(\mathbf{r})$, and Property (3) ensures that gradient of ρ is smooth. Since hydrodynamics is driven in part by ∇P, the functional form of W is a key component of SPH codes.[2] A natural choice for the kernel is the Gaussian function. This form satisfies the requirements above, and the flattened central region of the Gaussian helps to ensure that slight adjustments of nearby neighbors does not change the density gradient at the center of an SPH particle. However, this function has infinite extent, and therefore the sum in Equation (D.3) requires $\mathcal{O}(N_{\rm part}^2)$ operations to calculate the density and its derivative at each particle's location. Having a kernel with compact support, i.e., that models a Gaussian but truncates at a finite radius, is used instead. This reduces the sum to $\mathcal{O}(N_{\rm neigh} N_{\rm part})$, where $N_{\rm neigh}$ is usually $\sim 10^2$. The smoothing length of individual particles is chosen to ensure that each particle overlaps with a target number of neighbors.

For three-dimensional problems, a popular choice is the Monaghan & Lattanzio (1985) kernel:

$$\begin{aligned} W(r,h) &= \frac{1}{\pi h^3}\left[1 - \frac{3}{2}\left(\frac{r}{h}\right)^2 + \frac{3}{4}\left(\frac{r}{h}\right)^3\right] \ , 0 \le r/h \le 1 \\ W(r,h) &= \frac{1}{4\pi h^3}\left[2 - \left(\frac{r}{h}\right)\right]^3 \ , \qquad 1 \le r/h \le 2 \\ W(r,h) &= 0 \ , \qquad r/h \ge 2 \end{aligned} \tag{D.4}$$

---

[2] This is not meant to be an introduction to SPH codes. Evaluating pressure gradients involves more than just calculating ∇W for each particle. See Price (2012) for the details.

where $r = |\mathbf{r} - \mathbf{r}_j|$. Only neighbors within $2h$ contribute to the sum; the list of such particles can be quickly computed using tree methods (e.g., Hernquist & Katz 1989).

The choice of a Gaussian, or any other functional form of W, may appear arbitrary, especially when imagining a few lone particles. One may be inclined to scoff at the idea that a fluid is composed of a discrete set of Gaussian balls. The error in this thinking arises by failing to recognize that the power of SPH comes about when there are many particles overlapping each other at every point in space. Indeed, SPH codes set the smoothing length of particles so to ensure that each particle overlaps with a constant or minimally varying number of neighbor particles N_{neigh} (e.g., Springel 2005). Just as Fourier analysis discretizes a signal into a set of frequencies, SPH can be thought of as discretizing the local fluid into a collection of finite fluid elements; the macroscopic fluid quantities arise from the superposition of all of these particles.[3] Also, any well-functioning kernel in the limit of increased resolution ($h \to 0$) should begin to resemble a delta function and provide results independent of the exact form of W. In this limit, in order to ensure the N_{neigh} requirement, SPH codes would decompose the fluid into an infinite number of particles. This method has been extensively studied, with several kernels considered, and the Gaussian form becomes an ideal choice (Fulk & Quinn 1996). Below, when we describe how we map Gaussian shaped particles onto a grid, continue to bear in mind that every cell of the grid will receive contributions from many particles.

D.1 Mapping Method

Given a set of particles from an SPH simulation, each with its own mass, smoothing length, velocity, internal energy density, we wish to project these quantities onto a three-dimensional grid in such a way that conserves mass, momentum, and energy, as well as retains the large-scale structure that was present in the SPH simulation. We also can project any tracer species, quantities that are advected with the flow into a fixed set of cells. Currently our projection algorithm projects onto a uniform grid of cells–obtained by subdividing the domain equally in each dimension–but there is no reason why this could not also work with adaptive grids that have differing sized cells throughout the domain.

The density distribution of each particle depends on the kernel. For a given cell that overlaps with a given particle, the particle's kernel is not constant within that cell. A correct evaluation of the total mass of particle j within that cell is an integral over the cell

$$m_p \int_{\mathrm{cell}} W(r_{ij}, h_j)\, dV \ ,$$

where here $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $\mathbf{r}_i$ is a point within the cell. Our projection routine evaluates this integral using a Monte Carlo method.

---

[3] As an aside, this analogy extends to the shortcomings of this method as well: just as Fourier analysis struggles with discontinuities, the major shortcoming of SPH methods is the treatment of shocks.

For cell i, we compute the fraction f_{ij} of the kernel-weighted volume of particle j that overlaps with the cell. We do so by evaluating the kernel $W(r_{ij}, h_j)$ at N random points $\mathbf{r}_i$ in the cell and computing the sum

$$\langle W \rangle = \frac{1}{N} \sum_{i=1}^{N} W(r_{ij}, h_j) \ ,$$

so that

$$f_{ij} = \int_{\text{cell}} W(r_{ij}, h_j)\, dV \approx \langle W \rangle \int_{\text{cell}} dV = \langle W \rangle \Delta x^3 \ .$$

Here we have assumed each dimension has a length Δx. Assuming cube cells is not necessary for this method, however. The standard deviation of the mean for this method is

$$\sigma_M = \sqrt{\frac{\langle W^2 \rangle - \langle W \rangle^2}{N}} = \frac{1}{\sqrt{N}} \left(\frac{1}{N} \sum_{i=1}^{N} W^2(r_{ij}, h_j) - \left(\frac{1}{N} \sum_{i=1}^{N} W(r_{ij}, h_j) \right)^2 \right)^{1/2} \ .$$

For a given particle and cell, we track both $\langle W^2 \rangle$ and $\langle W \rangle$, re-evaluating σ_M every $N = 2,500$ samples. Once $\sigma_M < 0.001$, we exit the Monte Carlo routine for that particular particle-cell pair. We evaluate whether a given particle overlaps the cell before entering the Monte Carlo routine. Furthermore, we automatically set $f_{ij} = 1$ if the particle is entirely contained within the cell. Our choice of $N = 2,500$ was selected to reduce the chance that our random set of points would mis-calculate the kernel-weighted volume of a particle that barely overlaps with a given cell.[4]

To project the entire set of SPH particles onto the grid, we proceed using the following parallelized routine:

1. Input a set of SPH particles, the physical dimensions of the box to project into, and the grid imposed on that domain.
2. The center of the box is assumed to be (0,0,0). Adjust the positions of all the SPH particles so that the center of the densest particle's center is located at (0,0,0). Shift the velocities into the center of mass frame.[5]
3. Share the list of SPH particles with every processor. Each processor then gets a fraction of the grid cells. This eliminates the need for the processors to ever communicate with each other.
4. For each cell:

---

[4] This demand may be overkill, since those particles would barely contribute to the cell's ultimate state anyway.

[5] For these simulations, the densest particle's center and the simulations center of mass are nearly the same.

   (a) Loop through the particles and tag particles that overlap with that cell.
   (b) For each tagged particle, compute f_{ij} using the Monte Carlo routine. Particle j then contributes $\rho_{ij} = f_{ij} m_j / \Delta x^3$ to the cell.
   (c) The cell also receives momentum and a density-weighted internal energy: $\rho_{ij} \mathbf{v}_j$, and $\rho_{ij} e_j$. A similar quantity is calculated for any tracer fields included.
   (d) Once all tagged particles have been projected into the cell, the total density of the cell has been computed: $\rho_i = \sum_j \rho_{ij}$.
   (e) Divide the projected momentum and density-weighted internal energy to get the cell's velocity and internal energy: e.g., $\mathbf{v}_i = \sum_j \rho_{ij} \mathbf{v}_j / \rho_i$. Similarly for the tracer fields.

5. At this point, that cell is now completely initialized. This process is repeated for each cell the processor owns. Data can be outputted incrementally or all at once.

6. Once all processors have completed these tasks, the individual files from each processor can be stitched together to give a full set of initial conditions on the grid.

Richardson et al. (2013) used a similar method to map SPH particles to a grid. If a given particle overlapped the cell center $\mathbf{x}_i$, they projected $\rho_{ij} = m_j W(|\mathbf{r}_j - \mathbf{x}_i|, h_j)$ into the cell. The projected mass need not equal the actual mass of the particle. They define a correction factor c_j as the ratio of the particle's actual mass and the total mass projected from that particle.[6]. They apply this correction factor to the densities deposited in each cell from that particle to enforce mass conservation. The remaining particle's quantities are projected in the same vein as above. With sufficient resolution, the gradient of W will be negligible across each cell, and the correction factor will be close to unity. Otherwise, this method tends to concentrate mass near the center of each particle when $c_j > 1$ and smooths out the particle when $c_j < 1$. This becomes an issue only when $h_j/\Delta x$ is ~ 10 or smaller.

D.2 Example & Conservation Test

Future work will use initial conditions drawn from Halo 9 of Stacy & Bromm (2013). Their simulations commenced at $z = 100$ from cosmological initial conditions that are in accordance with a Λ-cold dark matter cosmology ($\Omega_\Lambda = 0.7, \Omega_M = 0.3, \Omega_B = 0.04, \sigma_8 = 0.9$, and $h = 0.7$). Their box was 1.4 Mpc [comoving] in size using Gadget-2 (Springel 2005). As the simulation progressed, they continued to refine ten regions that formed dark matter mini-halos, each tens of parsecs in size. We will simulate the central $(\text{pc})^3$ [physical] region of their Halo 9, starting at redshift $z = 24.71$. Our initial conditions project any particle whose physical extent overlaps with the intended domain, including particles with centers outside of our box. We project only the baryonic matter since within this region the dark matter

---

[6] We note that they incorrectly define this quantity as the inverse ratio in their paper.

distribution is nearly uniform and does not contribute to the gradient of the gravitational potential.

In order to test how well our projection method conserves mass, momentum, and energy, we perform this projection again, using the same domain but projecting only the particles that lie entirely within it. Since each particle has a well-determined mass, momentum, and energy, prior to the projection we can sum over these entirely contained particles to calculate the total mass, momentum, and energy to be projected. After projecting, these can be compared to the sum over the cells of the same quantities. While this projection is not identical to our actual initial conditions, this allows us to assess the global accuracy of the Monte Carlo routine. Doing so, we find that mass, momentum, and energy are all conserved to better than 0.01% for a 512^3 grid. We note, however, that even with coarser grids (down to 64^3), we find that all three quantities are conserved to better than 1%. Figure D.1 displays a density slice of our example initial conditions.

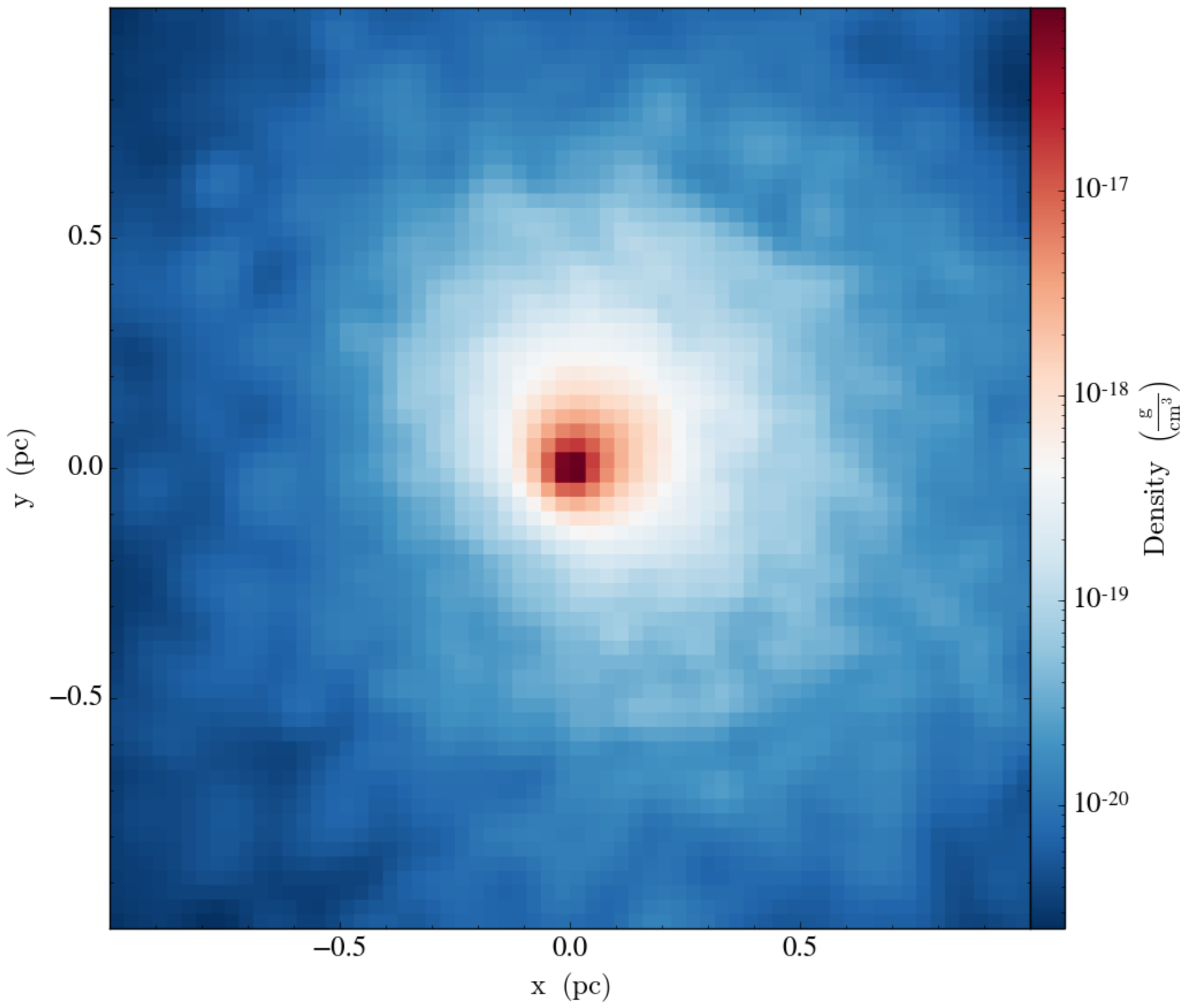

Figure D.1: Density slice through the center of the box of our example initial conditions. SPH data from Stacy & Bromm (2013) was used to project onto a 128^3 grid; future work will instead utilize a 512^3 grid for our initial conditions.

Bibliography

Abel, T., Anninos, P., Zhang, Y., & Norman, M. L. 1997, New Astron., 2, 181

Abel, T., Bryan, G. L., & Norman, M. L. 2002, Sci, 295, 93

Abel, T., Norman, M. L., & Madau, P. 1999, ApJ, 523, 66

Abel, T., & Wandelt, B. D. 2002, MNRAS, 330, L53

Adachi, I., Hayashi, C., & Nakazawa, K. 1976, Progress of Theoretical Physics, 56, 1756

Alvarez, M. A., Bromm, V., & Shapiro, P. R. 2006, ApJ, 639, 621

Andre, P., Ward-Thompson, D., & Greaves, J. 2012, Science, 337, 69

André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102

Applied Numerical Algorithms Group, & LBNL Computational Research Division. 2012, Chombo: Adaptive Solutions of Partial Differential Equations, Astrophysics Source Code Library, ascl:1202.008

Armitage, P. J. 2011, ARA&A, 49, 195

Aubert, D., & Teyssier, R. 2008, MNRAS, 387, 295

Baczynski, C., Glover, S. C. O., & Klessen, R. S. 2015, MNRAS, 454, 380

Bai, X., & Goodman, J. 2009, ApJ, 701, 737

Bai, X., & Stone, J. M. 2010a, ArXiv e-prints, arXiv:1005.4982

—. 2010b, ApJS, 190, 297

—. 2010c, ArXiv e-prints, arXiv:1005.4981 [astro-ph.EP]

Balbus, S. A. 2009, ArXiv e-prints, arXiv:0906.0854

Barkana, R., & Loeb, A. 2001, Phys. Rep., 349, 125

Barranco, J. A. 2009, ApJ, 691, 907

Barranco, J. A., & Marcus, P. S. 2000, in Studying Turbulence Using Numerical Simulation Databases, 8. Proceedings of the 2000 Summer Program, p. 97, 97

Barranco, J. A., & Marcus, P. S. 2005, ApJ, 623, 1157

—. 2006, Journal of Computational Physics, 219, 21

Bate, M. R., Bonnell, I. A., & Price, N. M. 1995, MNRAS, 277, 362

Batygin, K., & Stevenson, D. J. 2010, ApJ, 714, L238

Bisbas, T. G., Haworth, T. J., Williams, R. J. R., et al. 2015, MNRAS, 453, 1324

Blitz, L., & Rosolowsky, E. 2005, in Astrophysics and Space Science Library, Vol. 327, The Initial Mass Function 50 Years Later, ed. E. Corbelli, F. Palla, & H. Zinnecker, 287

Blum, J., & Wurm, G. 2008, ARA&A, 46, 21

Bodenheimer, P. H. 2011, Principles of Star Formation

Bondi, H. 1952, MNRAS, 112, 195

Bondi, H., & Hoyle, F. 1944, MNRAS, 104, 273

Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 1997, MNRAS, 285, 201

—. 2001, MNRAS, 323, 785

Bonnor, W. B. 1956, MNRAS, 116, 351

Bromm, V. 2013, Rep. Prog. Phys., 76, 112901

Bromm, V., Coppi, P. S., & Larson, R. B. 2002, ApJ, 564, 23

Bromm, V., & Larson, R. B. 2004, ARA&A, 42, 79

Bromm, V., & Yoshida, N. 2011, ARA&A, 49, 373

Brown, R. L., Wild, W., & Cunningham, C. 2004, Advances in Space Research, 34, 555

Chabrier, G. 2003, PASP, 115, 763

Chandrasekhar, S. 1981, Hydrodynamic And Hydromagnetic Stability, 1st edn. (Dover Publications, New York)

Chiang, E. 2008, ApJ, 675, 1549

Chiang, E., & Youdin, A. 2010, Annual Reviews of Earth and Planetary Science, 38

Chiang, E. I., & Goldreich, P. 1997, ApJ, 490, 368

Clark, P. C., Glover, S. C. O., Klessen, R. S., & Bromm, V. 2011, ApJ, 727, 110

Coradini, A., Magni, G., & Federico, C. 1981, A&A, 98, 173

Crutcher, R., Heiles, C., & Troland, T. 2003, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 614, Turbulence and Magnetic Fields in Astrophysics, ed. E. Falgarone & T. Passot, 155

Crutcher, R. M. 1999, ApJ, 520, 706

—. 2012, ARA&A, 50, 29

Cunningham, A. J., McKee, C. F., Klein, R. I., Krumholz, M. R., & Teyssier, R. 2012, ApJ, 744, 185

Cuzzi, J. N., Dobrovolskis, A. R., & Champney, J. M. 1993, Icarus, 106, 102

Cuzzi, J. N., Hogan, R. C., & Shariff, K. 2008, ApJ, 687, 1432

Dale, J. E. 2015, New A Rev., 68, 1

D'Alessio, P., Calvet, N., & Hartmann, L. 2001, ApJ, 553, 321

D'Alessio, P., Calvet, N., Hartmann, L., Franco-Hernández, R., & Servín, H. 2006, ApJ, 638, 314

Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2014, Protostars and Planets VI, 3

Dominik, C., & Tielens, A. G. G. M. 1997, ApJ, 480, 647

Dopcke, G., Glover, S. C. O., Clark, P. C., & Klessen, R. S. 2013, ApJ, 766, 103

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium

Draine, B. T., & McKee, C. F. 1993, ARA&A, 31, 373

Drazin, P. G., & Reid, W. H. 2004, Hydrodynamic Stability, 2nd edn. (Cambridge University Press, Cambridge)

Dullemond, C. P., & Dominik, C. 2005, A&A, 434, 971

Dullemond, C. P., Hollenbach, D., Kamp, I., & D'Alessio, P. 2007, Protostars and Planets V, 555

Dunham, M. M., Stutz, A. M., Allen, L. E., et al. 2014, Protostars and Planets VI, 195
Eaton, J. K., & Fessler, J. R. 1994, International Journal of Multiphase Flow Supplemental, 20, 169
Ebert, R. 1955, ZAp, 37, 217
Edgar, R. 2004, New A Rev., 48, 843
Evans, II, N. J. 1999, in Science with the Atacama Large Millimeter Array (ALMA)
Federrath, C., Banerjee, R., Clark, P. C., & Klessen, R. S. 2010, ApJ, 713, 269
Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102
Forrey, R. C. 2013, Phys. Rev. A, 88, 052709
Foster, P. N., & Chevalier, R. A. 1993, ApJ, 416, 303
Fulk, D. A., & Quinn, D. W. 1996, Journal of Computational Physics, 126, 165
Gammie, C. F. 1996, ApJ, 462, 725
—. 2001, ApJ, 553, 174
Garaud, P., & Lin, D. N. C. 2004, ApJ, 608, 1050
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, in Proc. SPIE, Vol. 6265, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 62650N
Gaspari, M., Ruszkowski, M., & Oh, S. P. 2013, ArXiv e-prints, arXiv:1301.3130 [astro-ph.CO]
Ghirlanda, G., Salvaterra, R., Ghisellini, G., et al. 2015, MNRAS, 448, 2514
Gibbs, J. W. 1898, Nature, 59, 200
Gilman, P. A., & Glatzmaier, G. A. 1981, ApJS, 45, 335
Glover, S. 2005, Space Sci. Rev., 117, 445
Glover, S. 2013, in Astrophysics and Space Science Library, Vol. 396, The First Galaxies, ed. T. Wiklind, B. Mobasher, & V. Bromm, 103
Glover, S. C. O., & Abel, T. 2008, MNRAS, 388, 1627
Glover, S. C. O., & Brand, P. W. J. L. 2001, MNRAS, 321, 385
Goldreich, P., Lithwick, Y., & Sari, R. 2004a, ARA&A, 42, 549
—. 2004b, ARA&A, 42, 549
Goldreich, P., & Lynden-Bell, D. 1965, MNRAS, 130, 125
Goldreich, P., & Ward, W. R. 1973, ApJ, 183, 1051
Gómez, G. C., & Ostriker, E. C. 2005, ApJ, 630
Gonzalez, G. 1997, MNRAS, 285, 403
González, M., Audit, E., & Huynh, P. 2007, A&A, 464, 429
Goodman, J., & Pindor, B. 2000, Icarus, 148, 537
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Gough, D. O. 1969, Journal of Atmospheric Sciences, 26, 448
Gould, R. J., & Salpeter, E. E. 1963, ApJ, 138, 393
Gray, W. J., & Scannapieco, E. 2010, ApJ, 718, 417
Greif, T. H., & Bromm, V. 2006, MNRAS, 373, 128
Greif, T. H., Glover, S. C. O., Bromm, V., & Klessen, R. S. 2010, ApJ, 716, 510
Greif, T. H., Johnson, J. L., Klessen, R. S., & Bromm, V. 2009, MNRAS, 399, 639
Guillot, T. 2005, Annual Review of Earth and Planetary Sciences, 33, 493
Guillot, T., Santos, N. C., Pont, F., et al. 2006, A&A, 453, L21

Guillot, T., & Showman, A. P. 2002, A&A, 385, 156
Haiman, Z., Rees, M. J., & Loeb, A. 1997, ApJ, 476, 458
Haiman, Z., Spaans, M., & Quataert, E. 2000, ApJ, 537, L5
Hartwig, T., Bromm, V., Klessen, R. S., & Glover, S. C. O. 2015, MNRAS, 447, 3892
Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2003, ApJ, 591, 288
Henyey, L. G., Lelevier, R., & Levée, R. D. 1955, PASP, 67, 154
Hernández, J., Hartmann, L., Calvet, N., et al. 2008, ApJ, 686, 1195
Hernquist, L., & Katz, N. 1989, ApJS, 70, 419
Hillenbrand, L. A. 2005, ArXiv Astrophysics e-prints, arXiv:astro-ph/0511083
Hirano, S., Hosokawa, T., Yoshida, N., et al. 2014, ApJ, 781, 60
Hogan, R. C., & Cuzzi, J. N. 2007, Phys. Rev. E, 75, 056305
Hollenbach, D., Johnstone, D., Lizano, S., & Shu, F. 1994, ApJ, 428, 654
Hopkins, P. F., Narayan, R., & Hernquist, L. 2006, ApJ, 643, 641
Hosokawa, T., Hirano, S., Kuiper, R., et al. 2016, ApJ, 824, 119
Hosokawa, T., & Inutsuka, S.-i. 2006, ApJ, 648, L131
Hosokawa, T., Omukai, K., & Yorke, H. W. 2012a, ApJ, 756, 93
Hosokawa, T., Omukai, K., Yoshida, N., & Yorke, H. W. 2011, Science, 334, 1250
Hosokawa, T., Yorke, H. W., Inayoshi, K., Omukai, K., & Yoshida, N. 2013, ApJ, 778, 178
Hosokawa, T., Yorke, H. W., & Omukai, K. 2010a, ApJ, 721, 478
—. 2010b, ApJ, 721, 478
Hosokawa, T., Yoshida, N., Omukai, K., & Yorke, H. W. 2012b, ApJ, 760, L37
Hoyle, F., & Lyttleton, R. A. 1939, Proceedings of the Cambridge Philosophical Society, 35, 405
Hubber, D. A., Walch, S., & Whitworth, A. P. 2013, ArXiv e-prints, arXiv:1301.4520 [astro-ph.IM]
Hummel, J. A., Stacy, A., Jeon, M., Oliveri, A., & Bromm, V. 2015, MNRAS, 453, 4136
Ilgner, M., & Nelson, R. P. 2006, A&A, 445, 223
Ishitsu, N., & Sekiya, M. 2003, Icarus, 165, 181
Jeans, J. H. 1902, Royal Society of London Philosophical Transactions Series A, 199, 1
Jeon, M., Pawlik, A. H., Greif, T. H., et al. 2012, in American Institute of Physics Conference Series, Vol. 1480, American Institute of Physics Conference Series, ed. M. Umemura & K. Omukai, 325
Jijina, J., Myers, P. C., & Adams, F. C. 1999, ApJS, 125, 161
Johansen, A., Henning, T., & Klahr, H. 2006, ApJ, 643, 1219
Johansen, A., Oishi, J. S., Low, M., et al. 2007, Nature, 448, 1022
Johansen, A., Youdin, A., & Mac Low, M. 2009, ApJ, 704, L75
Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010, PASP, 122, 905
Johnson, J. L., Greif, T. H., & Bromm, V. 2007a, ApJ, 665, 85
—. 2007b, ApJ, 665, 85
Karlsson, T., Bromm, V., & Bland-Hawthorn, J. 2013, Reviews of Modern Physics, 85, 809
Karlsson, T., Johnson, J. L., & Bromm, V. 2008, ApJ, 679, 6
Keller, S. C., Schmidt, B. P., Bessell, M. S., et al. 2007, PASA, 24, 1

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kitayama, T., Yoshida, N., Susa, H., & Umemura, M. 2004, ApJ, 613, 631
Klein, R. I., Inutsuka, S.-I., Padoan, P., & Tomisaka, K. 2007, Protostars and Planets V, 99
Klessen, R. S., Krumholz, M. R., & Heitsch, F. 2011, Advanced Science Letters, 4, 258
Knobloch, E., & Spruit, H. C. 1985, Geophysical and Astrophysical Fluid Dynamics, 32, 197
—. 1986, A&A, 166, 359
Kokubo, E., & Ida, S. 1996, Icarus, 123, 180
—. 2012, ArXiv e-prints, arXiv:1212.1558 [astro-ph.EP]
Kroupa, P. 2002, Science, 295, 82
Krumholz, M. R. 2014, Phys. Rep., 539, 49
Krumholz, M. R., Klein, R. I., McKee, C. F., & Bolstad, J. 2007, ApJ, 667, 626
Krumholz, M. R., & McKee, C. F. 2008, Nature, 451, 1082
Krumholz, M. R., McKee, C. F., & Klein, R. I. 2004a, ApJ, 611, 399
—. 2004b, ApJ, 611, 399
—. 2005, ApJ, 618, L33
—. 2006, ApJ, 638, 369
Lai, S.-P., Velusamy, T., & Langer, W. D. 2002, in Bulletin of the American Astronomical Society, Vol. 34, American Astronomical Society Meeting Abstracts, 1258
Larson, R. B. 1981, MNRAS, 194, 809
—. 2005, MNRAS, 359, 211
Lecoanet, D., Zweibel, E. G., Townsend, R. H. D., & Huang, Y. 2010, ApJ, 712, 1116
Lee, A. T., Chiang, E., Asay-Davis, X., & Barranco, J. 2010, ApJ, 718, 1367
Lee, A. T., Cunningham, A. J., McKee, C. F., & Klein, R. I. 2014, ApJ, 783, 50
Lee, A. T., & Stahler, S. W. 2011, MNRAS, 416, 3177
Li, P. S., Martin, D. F., Klein, R. I., & McKee, C. F. 2012, ApJ, 745, 139
Li, Z.-Y., & McKee, C. F. 1996, ApJ, 464, 373
Lissauer, J. J., Hubickyj, O., D'Angelo, G., & Bodenheimer, P. 2009, Icarus, 199, 338
Lithwick, Y. 2009, ApJ, 693, 85
Lodders, K. 2003, ApJ, 591, 1220
Mac Low, M.-M., & Klessen, R. S. 2004, Reviews of Modern Physics, 76, 125
Magnani, L., Blitz, L., & Mundy, L. 1985, ApJ, 295, 402
Martin, R. G., Colella, P. C., & Graves, D. 2008, JCP, 227, 1863
Maxey, M. R. 1987, J. Fluid Mech., 174, 441
McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565
McKee, C. F., & Tan, J. C. 2003, ApJ, 585, 850
—. 2008, ApJ, 681, 771
Mignone, A., Zanni, C., Tzeferacos, P., et al. 2012, ApJS, 198, 7
Mihalas, D., & Mihalas, B. W. 1984, Foundations of radiation hydrodynamics
Miles, J. W. 1961, Journal of Fluid Mechanics, 10, 496
Milosavljević, M., Couch, S. M., & Bromm, V. 2009, ApJ, 696, L146
Moeckel, N., & Throop, H. B. 2009, ApJ, 707, 268
Monaghan, J. J., & Lattanzio, J. C. 1985, A&A, 149, 135

Mori, M., Ferrara, A., & Madau, P. 2002, ApJ, 571, 40

Mouschovias, T. C. 1976, ApJ, 207, 141

Murray, N., & Chang, P. 2012, ApJ, 746, 75

Myers, A. T., Klein, R. I., Krumholz, M. R., & McKee, C. F. 2014, MNRAS, 439, 3420

Myers, A. T., McKee, C. F., Cunningham, A. J., Klein, R. I., & Krumholz, M. R. 2013a, ApJ, 766, 97

—. 2013b, ArXiv e-prints, arXiv:1211.3467 [astro-ph.SR]

Nakagawa, Y., Sekiya, M., & Hayashi, C. 1986, Icarus, 67, 375

Offner, S. S. R., Hansen, C. E., & Krumholz, M. R. 2009, ApJ, 704, L124

Ogura, Y., & Phillips, N. A. 1962, Journal of Atmospheric Sciences, 19

Oka, T., Hasegawa, T., Sato, F., et al. 2001, ApJ, 562, 348

Omukai, K., & Inutsuka, S.-i. 2002, MNRAS, 332, 59

Omukai, K., & Nishi, R. 1998a, ApJ, 508, 141

—. 1998b, ApJ, 508, 141

Omukai, K., & Palla, F. 2003, ApJ, 589, 677

Ormel, C. W., Spaans, M., & Tielens, A. G. G. M. 2007, A&A, 461, 215

Palla, F., Salpeter, E. E., & Stahler, S. W. 1983, ApJ, 271, 632

Perez-Becker, D., & Chiang, E. 2010, ArXiv e-prints

Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339

Price, D. J. 2012, Journal of Computational Physics, 231, 759

Price, D. J., & Federrath, C. 2010, MNRAS, 406, 1659

Price, D. J., Tricco, T. S., & Bate, M. R. 2012, MNRAS, 423, L45

Raga, A. C., Cantó, J., & Rodríguez, L. F. 2012, MNRAS, 419, L39

Razoumov, A. O., & Scott, D. 1999, MNRAS, 309, 287

Richardson, M. L. A., Scannapieco, E., & Thacker, R. J. 2013, ApJ, 771, 81

Ripamonti, E., Haardt, F., Ferrara, A., & Colpi, M. 2002, MNRAS, 334, 401

Rosen, A. L., Krumholz, M. R., McKee, C. F., & Klein, R. I. 2016, MNRAS, 463, 2553

Rosen, A. L., Krumholz, M. R., Oishi, J. S., Lee, A. T., & Klein, R. I. 2017, Journal of Computational Physics, 330, 924

Ruffert, M. 1994, ApJ, 427, 342

—. 1996, A&A, 311, 817

Rybicki, G. B., & Lightman, A. P. 1986, Radiative Processes in Astrophysics, 400

Safronov, V. S. 1969, Evolution of the protoplanetary cloud and formation of the Earth and planets (IPST Jerusalem)

Salvaterra, R. 2015, Journal of High Energy Astrophysics, 7, 35

Santos, N. C., Israelian, G., & Mayor, M. 2004, A&A, 415, 1153

Sekiya, M. 1998, Icarus, 133, 298

Shcherbakov, R. V. 2008, ApJS, 177, 493

Shima, E., Matsuda, T., Takeda, H., & Sawada, K. 1985, MNRAS, 217, 367

Shu, F. H. 1977, ApJ, 214, 488

Shu, F. H., Adams, F. C., & Lizano, S. 1987, ARA&A, 25, 23

Sokasian, A., Yoshida, N., Abel, T., Hernquist, L., & Springel, V. 2004, MNRAS, 350, 47

Spitzer, L. 1978, Physical processes in the interstellar medium

Springel, V. 2005, MNRAS, 364, 1105

Stacy, A., & Bromm, V. 2013, MNRAS, 433, 1094

Stacy, A., Bromm, V., & Lee, A. T. 2016, MNRAS, 462, 1307

Stacy, A., Bromm, V., & Loeb, A. 2011, MNRAS, 413, 543

Stacy, A., Greif, T. H., & Bromm, V. 2012, MNRAS, 422, 290

Stacy, A., Greif, T. H., Klessen, R. S., Bromm, V., & Loeb, A. 2013, MNRAS, 431, 1470

Stacy, A., Pawlik, A. H., Bromm, V., & Loeb, A. 2014, MNRAS, 441, 822

Stahler, S. W., & Palla, F. 2005, The Formation of Stars, 865

Stahler, S. W., Palla, F., & Salpeter, E. E. 1986a, ApJ, 308, 697

—. 1986b, ApJ, 302, 590

Strömgren, B. 1939, ApJ, 89, 526

Susa, H. 2006, PASJ, 58, 445

Susa, H., Hasegawa, K., & Tominaga, N. 2014, ApJ, 792, 32

Tan, J. C., & McKee, C. F. 2004, ApJ, 603, 383

Tegmark, M., Silk, J., Rees, M. J., et al. 1997, ApJ, 474, 1

Terebey, S., Shu, F. H., & Cassen, P. 1984, ApJ, 286, 529

Teyssier, R. 2002, A&A, 385, 337

Toomre, A. 1964, ApJ, 139, 1217

Toropina, O. D., Romanova, M. M., & Lovelace, R. V. E. 2012, MNRAS, 420, 810

Troland, T. H., & Crutcher, R. M. 2008, ApJ, 680, 457

Truelove, J. K., Klein, R. I., McKee, C. F., et al. 1997, ApJ, 489, L179

Turk, M. J., Clark, P., Glover, S. C. O., et al. 2011a, ApJ, 726, 55

Turk, M. J., Oishi, J. S., Abel, T., & Bryan, G. L. 2012, ApJ, 745, 154

Turk, M. J., Smith, B. D., Oishi, J. S., et al. 2011b, ApJS, 192, 9

Turner, N. J., Carballido, A., & Sano, T. 2010a, ApJ, 708, 188

—. 2010b, ApJ, 708, 188

Ward, W. R. 1976, in Frontiers of Astrophysics, 1

Ward, W. R. 2000, On Planetesimal Formation: The Role of Collective Particle Behavior (University of Arizona Press), 75

Ward-Thompson, D., Kirk, J. M., André, P., et al. 2010, A&A, 518, L92

Watson, D. M., Bohac, C. J., Hull, C., et al. 2007, Nature, 448, 1026

Weidenschilling, S. J. 1977, MNRAS, 180, 57

—. 1980, Icarus, 44, 172

Weidenschilling, S. J., & Cuzzi, J. N. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine, 1031

Whalen, D., Abel, T., & Norman, M. L. 2004, ApJ, 610, 14

White, J. A., Boley, A. C., Hughes, A. M., et al. 2016, ApJ, 829, 6

Wilner, D. J., D'Alessio, P., Calvet, N., Claussen, M. J., & Hartmann, L. 2005, ApJ, 626, L109

Wise, J. H., & Abel, T. 2011, MNRAS, 414, 3458

Wise, J. H., Turk, M. J., Norman, M. L., & Abel, T. 2012, ApJ, 745, 50

Woods, D. T., Klein, R. I., Castor, J. I., McKee, C. F., & Bell, J. B. 1996, ApJ, 461, 767
Woosley, S. E. 1993, ApJ, 405, 273
Yoshida, N., Abel, T., Hernquist, L., & Sugiyama, N. 2003, ApJ, 592, 645
Youdin, A. N. 2005a, ArXiv Astrophysics e-prints, arXiv:astro-ph/0508659
—. 2005b, ArXiv Astrophysics e-prints, arXiv:astro-ph/0508662
Youdin, A. N., & Chiang, E. I. 2004, ApJ, 601, 1109
Youdin, A. N., & Goodman, J. 2005, ApJ, 620, 459
Youdin, A. N., & Shu, F. H. 2002, ApJ, 580, 494
Zhao, B., Li, Z.-Y., Nakamura, F., Krasnopolsky, R., & Shang, H. 2011, ApJ, 742, 10
Zinnecker, H. 1982, in Astrophysics and Space Science Library, Vol. 98, IAU Colloq. 69: Binary and Multiple Stars as Tracers of Stellar Evolution, ed. Z. Kopal & J. Rahe, 115
Zuckerman, B., & Evans, II, N. J. 1974, ApJ, 192, L149