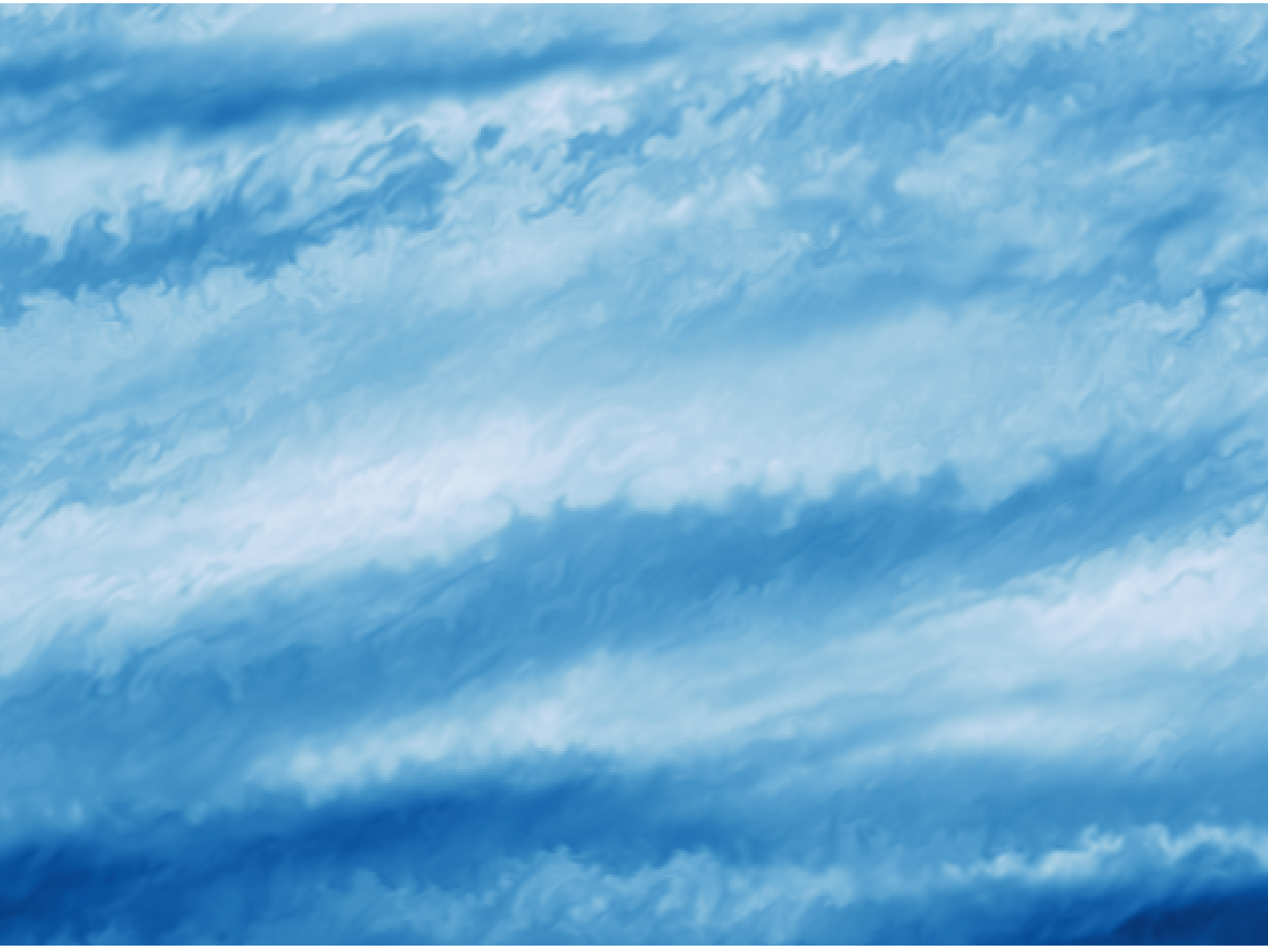
Formation Criteria and Initial Constraints on Objects Formed in Gravitationally Unstable Disks

Hans-Paul F. Baehr



Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Hans-Paul Frederick Baehr

born in Houston, Texas, USA

Oral examination: 5. July 2019

Formation Criteria and Initial Constraints on Objects Formed in Gravitationally Unstable Disks

Referees:
Apl. Prof. Dr. H. Hubertus Klahr
Apl. Prof. Dr. Henrik Beuther

Zusammenfassung: Bildungskriterien und Beschränkung anfänglicher Zustände von in gravitativ instabilen Scheiben gebildeten Objekten

Junge Protoplanetare Scheiben sind kühl und massiv, weshalb sie gravitativen Instabilitäten sowie Fragmentation unterliegen, was zur Bildung dichter Gasklumpen führen kann. Diese Fragmente sind massiv genug, um außerhalb von 50 astronomischen Einheiten Gasriesen und braune Zwerge zu bilden. Für klassische Planetenentstehungs-Szenarien ist es schwierig, planetare Kerne schnell genug entstehen zu lassen, um direkt beobachtete Planeten zu erklären. Ich habe hochaufgelöste dreidimensionale Hydrodynamik-Simulationen genutzt, um den Kollaps von selbstgravitativen Scheiben zu modellieren. Zielsetzungen dieser Simulationen waren, Einschränkungen für Bildungsregionen dieser Fragmente zu finden sowie deren anfängliche Gas- und Partikelmasse zu bestimmen, um letztlich mit direkt beobachteten Planeten und braunen Zwergen zu vergleichen. Ich fand heraus, dass das klassische Kriterium zur Kühlung der Scheibe—welche die Bildungsregion von Gasfragmenten auf die äußeren Scheibenregionen zwingt—in diesen Simulationen konvergiert und allgemein die Massen der massiven Gasriesen konsistent mit denen sind, die an braune Zwerge grenzen. Die Konzentration von Festkörpern in diesen Fragmenten führt zu einem Anstieg der gesamten Metallizität in diesen Regionen und zudem zu einem festen Kern mit einigen zehn Erdmassen. Um die Fragmentation in globalen Scheibensimulationen zu modellieren, habe ich ein Multigitter-Selbstgravitations-Lösungsalgorithmus im PLUTO Code implementiert, welcher adaptive Gitterverfeinerung nutzt, um sowohl die Scheibe wie auch Fragmente aufzulösen.

Abstract: Formation Criteria and Initial Constraints on Objects Formed in Gravitationally Unstable Disks

Early protoplanetary disks are cool and massive and thus subject to gravitational instabilites and fragmentation of the disk into dense clumps of gas. These fragments are massive enough to become gas giant planets and brown dwarfs in the distant regions of the disks beyond 50 au where traditional planet formation scenarios have trouble creating planetary cores fast enough to explain directly observed planets. I used high-resolution three-dimensional hydrodynamic simulations to model the collapse of self-gravitating disks to constrain the formation location of these fragments and characterize their initial gas and particle masses to compare to directly observed planets and brown dwarfs. I find the traditional cooling criterion, which constrains the formation location to the outer disk, is converged in these simulations and overall masses are consistent with massive gas giants bordering on brown dwarfs. The concentration of solid material in these fragments leads to an increase of the overall metallicity of the fragment and a solid core several tens of Earth masses. To model fragmentation with full disk simulations, I have also implemented a multigrid self-gravity solver in the PLUTO code which uses adaptive mesh refinement to resolve both the disk and fragments.

"[...] all the pieces matter."
Lester Freamon, *The Wire*, Season 1 Episode 6

Contents

Conventions

This thesis contains extensive use of units and conventions which may be unfamiliar to some readers unaccustomed to the scale-free shearing box simulations and terminology of astrophysical disks. Here I will clarify some of those conventions to provide more clarity about the results of this thesis.

All times in all figures are given in terms of the inverse of the orbital frequency Ω^{-1} and as such, a duration of $2\pi\Omega^{-1}$ should be interpreted as a full orbit for a given stellar mass M and orbital radius R.

$$\Omega^2 = \frac{GM}{R^3} \tag{0.1}$$

Similarly, all distances are reported in terms of the pressure scale height H which should be interpreted based on the aspect ratio of the disk H/R, which measures the how the height of the disk scales with radius and is a proxy for the disk temperature, and the orbital radius. Unless otherwise noted, one can assume for this thesis that the aspect ratio $H/R = 0.1$ and an orbital radius of $R = 100$ au.

Masses will often be quoted using the vertically-integrated volume density $\int \rho dz$, commonly known as the surface density Σ. While the volume density is often integrated over a Gaussian vertical profile which means $\Sigma = 2\pi H\rho$, a common approximation used in this thesis will assume $\Sigma \approx H\rho$.

The process by which disks gravitationally collapse goes by a number of different names with similar implications but different contexts, all of which will be used in this thesis. *Gravitational instability* or *disk instability* refers to any process where gravity dominates over other stabilizing forces and affect the equilibrium of the disk and will refer to all possible outcomes whether it be gravitoturbulence or collapse into a bound object. *Disk fragmentation* on the other hand refers only to the outcome where the disk collapses to form a large clump.

1 Introduction

The past few decades have seen a tremendous increase in the amount of available information about extrasolar stellar systems and their formation and evolution. Currently operating satellites and telescopes like *Kepler*, the *Atacama Large Millimeter/submillimeter Array (ALMA)* and the *Very Large Telescope (VLT)* among others continually provide new data and images that have both altered and refined our understanding of planet formation. Future projects and missions with equal promise include the *James Webb Space Telescope (JWST)* and the *European Extremely Large Telescope (E-ELT)* and well as the survey missions *Transitting Exoplanet Survey Satellite (TESS)* and *Planetary Transits and Oscillations of stars (PLATO)*. This wealth of observational data provides theorists with valuable constraints and system archetypes for disk evolution and planet formation models. Using only our own solar system as a basis, one might neglect the existence of gas giants with very small orbital periods, but the results of the *Kepler* and *K2* missions show the rich variety of exoplanet systems. From this ever increasing sample, the role of the theorist in planet formation is to explain both the diversity of planet and system populations, while still considering the constraints on both sets of information. This means using theoretical models based on our current understanding of physics to construct and test models of processes that may lead to planet formation and help astronomers better evaluate their observations. The following chapter will give an introduction to modern planet and star formation theory and its relevance to observational constraints. The emphasis will be on the formation of gas giant planets and low-mass stars since the focus of this thesis, gravitational instability (GI), also known as disk instability, only forms large gaseous objects and not smaller rocky planets.

The aim of this thesis is to investigate the formation conditions of gravitational instabilities in circumstellar disks using high-resolution hydrodynamic simulations and establish observational constraints on the objects formed via these formation conditions. In Chapter 2, this means looking into the non-convergence problem of numerical simulations of circumstellar disks, which affects the cooling constraint of objects formed by gravitational instabilities. Uncertainty in this criterion results in weak constraints on the formation location of fragments and the work of Baehr & Klahr (2015) aimed to resolve this discrepancy. Expanding upon those results, Chapter 3, expands upon the 2D local models to do 3D models of disk fragmentation (Baehr et al., 2017), which aim to solidify the formation criterion and establish initial fragment sizes. Then in Chapter 4, I add solid particles to the preceding 3D simulations to study the enrichment of fragments and what further observational constraints to solid and gas mass and atmospheric metallicities can be established. Finally in Chapter 5, I detail my efforts to run high-resolution full disk simulations by using adaptive mesh refinement methods, which required the implementation of a multigrid self-gravity solver. The thesis concludes with a discussion of the state of the field and the significance of these results therein in Chapter 6. The appendices which follow the thesis elaborate on the theoretical framework (Appendix A) and numerical tools (Appendix B) used to carry out the thesis.

1.1 Protoplanetary Disks

Accretion disks are commonly observed objects in the universe, occurring around a variety of objects (i.e. stars and stellar remnants, black holes and active glalctic nuclei) and on scales ranging from a few astronomical units (au) to a few parsecs. As the name implies these are flat, rotating, elliptical entities around a central potential which may contain a number of substructures and processes which lend them to unique phenomena. While this thesis focuses entirely on the circumstellar disks which form in the aftermath of star formation processes, some features, particularly those related to gravitational effects have very similar analogs on disks of galactic scales (Toomre, 1964; Goldreich & Lynden-Bell, 1965). As star formation proceeds, not all material of a collapsing gas cloud falls directly on the protostar, but coalesces as an envelope which slowly flattens into a disk and accretes onto the young star over time. The classic flat disk shape is a result of the slow transport of angular momentum compared to the vertical gravitational settling of material. Disks are not a guaranteed result of star formation, as ideal magnetohydrodynamics predicts that magnetic braking will strip envelope material of its angular momentum and a disk will not be supported (Allen et al., 2003). If non-deal terms are included however, the magnetic braking catastrophe can potentially be averted (Wurster et al., 2018).

Provided a disk forms, it's early evolution is regulated by the net accretion onto the disk considering both accretion losses onto the star and the infall onto the disk from the surrounding envelope. A positive net flow of material onto the disk will result in a massive disk susceptible to gravitational instabilities which will significantly impact the structure, accretion rate and evolution of the disk and thus influence the disk's later planet formation capabilities. The fragmentation of these young disks into dense gas objects which are potential percursors to gas giant planets are the primary focus of this thesis. In an effort to provide adequate background to the theory of protoplanetary disks and planet formation, the following sections give a broad overview of protoplanetary disks, including their formation, structure and properties.

The Young Protoplanetary Disk

The *solar nebular disk model* is the current best understanding of the formation of protostars and the disks that surround them, first used to describe the co-planar orbits of the solar system (Kant, 1755). Early observations of extrasolar systems showed the model was extensible to other systems as well (Pollack, 1984). Initially, vast quantities of gas and dust collect in large structures called nebulae, stretching up to thousands of parsecs in length. The mass is dominated by light gases and some heavier elements and molecules, but particularly molecular hydrogen. According to the solar nebula disk model, when a roughly spherical region of the nebula has enough mass, it will become unstable to gravitational collapse and contract until the core is hot and dense enough to begin nucleosynthesis. The radiation pressure from nucleosynthesis is enough to balance out gravity, creating a marginally hydrostatic object, and thus beginning the protostar phase (Kennicutt & Evans II, 2012; McKee & Ostriker, 2007).

While the majority of the initial cloud mass falls directly onto the protostar during the initial collapse, more continues to accrete onto the forming protostar and circumstellar disk from the surrounding nebula. As a protostar grows and its luminosity increases, solar winds break through the infall at the polar areas, creating narrow outflow jets. The outflows become wider with time, eventually separating the disk and star from the original nebula, revealing the new system with the star beginning to evolve along the main sequence (Adams & Lin, 1993). At the same time, the rest of the envelope begins to settle at the

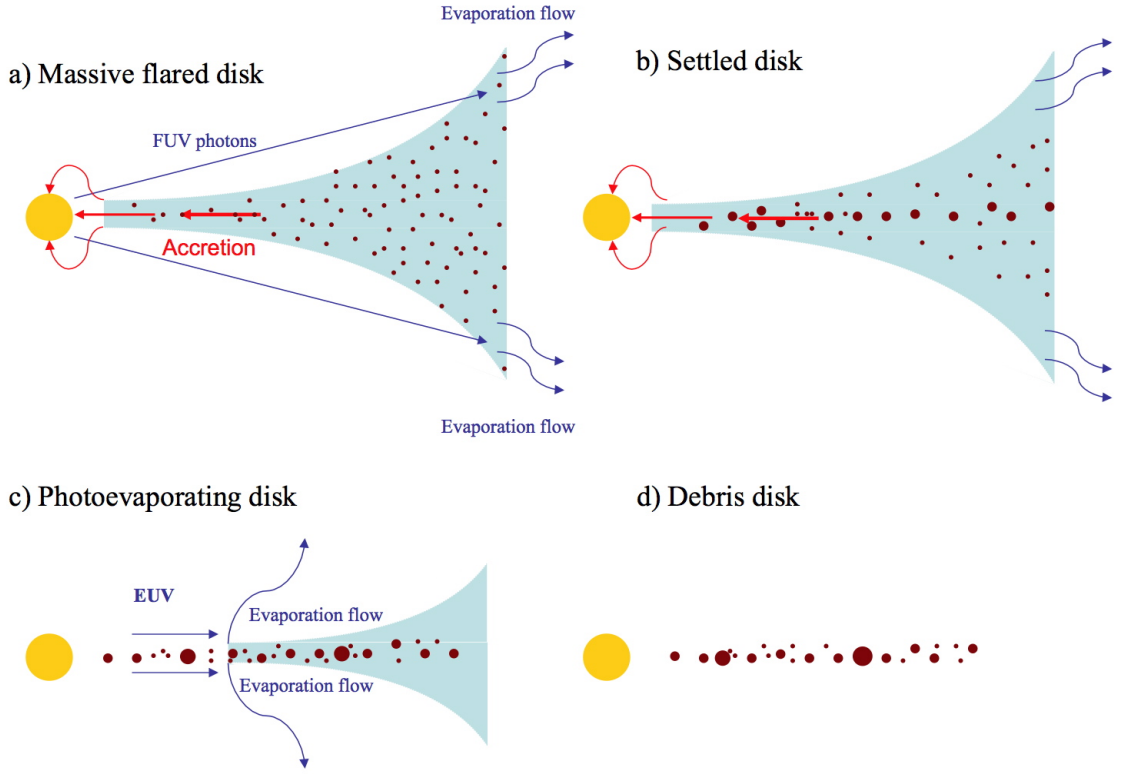


Figure 1.1: Stages of protoplanetary disk evolution beginning with a gaseous disk and minimal envelope, filled with small solid particles and accreting along the disk midplane. Gas (in light blue) is depleted from the disk through accretion, evaporation due to extreme ultraviolet (EUV) and far ultraviolet (FUV) irradiation from the central star and by the formation of planets in the disk. Original figure from Williams & Cieza (2011).

equatorial regions. While gravity will efficiently damp and settle vertical motions to the disk midplane, angular momentum is not so easily transferred from one annulus of the disk to the next (e.g. in the radial direction) and the result is the shape of disk. For disks to evolve and eventually dissipate in accordance with observations, theorists invoke processes which efficiently transfer angular momentum, which can arise from many different effects and properties of the disk (Williams & Cieza, 2011).

Disk Structure

By the time the envelope is no longer accreting on to the protoplanetary system, the disk will have flattened and the protostar becomes visible in the optical and infrared (McKee & Ostriker, 2007). A protoplanetary disk is generally said to have a thickness described by the pressure scale height, or the vertical distance over which the pressure decreases by a factor of e. In an isothermal disk the thickness can be derived assuming that all forces are in balance, also known as vertical hydrostatic equilibrium

$$\frac{\nabla P}{\rho} - \Omega^2 z = 0, \tag{1.1}$$

and the isothermal equation of state $P = \rho c_s^2$ for pressure P, gas density ρ, orbital frequency Ω and sound speed c_s. Simplifying the problem to the vertical direction z and

substituting with the ideal equation of state with assumed constant density

$$\frac{1}{P}\frac{dP}{dz} = \frac{\Omega^2}{c_s^2} z, \tag{1.2}$$

resulting in a Gaussian vertical pressure/density distribution

$$\int d\ln P = \int_0^z \frac{\Omega^2}{c_s^2} z' dz' \rightarrow \ln P = \frac{\Omega^2}{c_s^2}\left[\frac{z^2}{2}\right] + C, \tag{1.3}$$

which has a solution

$$P(z) = P_0 e^{-z^2/2H^2}. \tag{1.4}$$

The width of the Gaussian, $H = c_s/\Omega$ is the definition of the disk pressure scale height.

Sufficiently evolved disks will exhibit prominent radial flaring in the outer regions, described by an aspect ratio H/R and shown in Figure 1.1 (Williams & Cieza, 2011). The disk height increases at larger radial distances due the weaker vertical gravitational pull of the central star. Also shown in Figure 1.1 are some important disk processes such as accretion, both through the midplane and along magnetic field lines near the stellar surface, grain settling, and photoevaporation. While typical circumstellar envelopes might initially extend up to $\sim$ 1000 au, based on their spectral energy distributions, the characteristic size of protoplanetary disks is limited to a few hundred au (Adams & Lin, 1993). Photoevaporation of the outer disk by far-ultraviolet radiation from the central protostar, creates an outer truncation radius, shown in Figure 1.1 (a) and (b). During the later stages of protoplanetary development, an inner gap forms when the disk is no longer replenishing mass lost due to accretion onto the star and photoevaporation due to high energy ultraviolet radiation. The UV radiation may originate from a nearby star which bathes the entire disk or from the hot accreting protostar itself (Williams & Cieza, 2011).

The initial cloud is assumed to have some initial net rotation, so once it collapses into a comparatively smaller protostar plus disk, conservation of angular momentum holds that the latter system will have a higher rotation rate. If the protostar is significantly larger than the surrounding disk, the disk will rotate with a Keplerian profile, where the rotation frequency as a function of orbital radius R and stellar mass M is given by $\Omega = \sqrt{GM/R^3}$. There are however a number of disk features and substructures which disturb this uniformity and a fair amount of uncertainty over their origins and role in disk evolution. Radial gaps are one of the most ubiquitous features of disk observations (i.e HL Tau (ALMA Partnership et al., 2015) and TW Hydrae (Andrews et al., 2016)), but they often appear without any obvious planets to carve these gaps, opening speculation to other means of producing gaps, including ice lines causing a build up of material (Stammler et al., 2017), sintering (Okuzumi et al., 2016) and particle trapping at deadzone or gap edges (Ruge et al., 2016; Zhu et al., 2014). Disk asymmetries potentially resulting from vortices were discovered in the disk of Oph IRS 48 (van der Marel et al., 2013). The presence of warped disks are inferred from observations of HD 142527 (Casassus et al., 2015) and HD 135344B (Stolker et al., 2016), but the mystery of their formation remains.

Dust, in the form of silicates and metals form a small amount of the total mass of the disk, but make a larger contribution to the disk opacity than the gas, making the observation of dust the simplest way to infer disk masses using the ISM gas-to-dust mass ratio of $\sim$ 100 (Williams & Cieza, 2011; Bell & Lin, 1994). Disk masses are observed to be in the range of 0.001 to $0.1 M_\odot$, and usually significantly smaller than their parent star, but even these may be overestimates due to uncertainties in the dust opacities and CO

isotopologue ratios (Ansdell et al., 2016).

Little evidence has been found to suggest a difference in composition between matter that collapsed into the protostar and the material that became part of the nebula, so the protostar and surrounding disk are expected to have similar metallicities (Lissauer, 1993). In the case of our solar system, this leads to the idea of a *Minimum Mass Solar Nebula* (MMSN), the lowest mass for a nebula that could have formed the current configuration of solids which compose the planets in the solar system. The MMSN is estimated by taking the solid mass of each planet in the solar system and spreading it out over an annulus ranging halfway to its nearest neighbors (Lissauer, 1993). Gas is then added until the disk reaches solar composition, which yields radial distributions of mass ($\Sigma \sim R^{-3/2}$) and central temperature ($T_c \sim R^{-1/2}$) (Pollack, 1984; Bodenheimer & Lin, 2002).

With the ever-expanding catalog of exoplanet systems, this can be extended to a larger population, which show a similar power law relation of the surface density, although with a more massive nebula (Chiang & Laughlin, 2013).

Accretion onto the Star

The protostar continues to grow by gravitational and viscous accretion from the nearby medium, the strength of which is determined by the radius at which the material cannot escape the gravitational pull of the star, known as the Bondi radius and accretion rate (McKee & Ostriker, 2007). Mass transport in the disk is primarily driven by outward angular momentum transport; conservation of momentum dictates that the loss of angular momentum by some mass (and thus inward movement) must be accompanied by the increase in angular momentum of another parcel of mass. This loss of mass from the disk onto the star through accretion and photoevaporative winds is what regulates the lifetime of a disk which can vary significantly from system to system, even for similar stellar types. Typical disk lifetimes still tend to be around 5 to 10 million years (Ribas et al., 2015), after which the remnants are known as debris disks.

The relative strength of accretion is often described in terms of a dimensionless quantity of the disk known as α. According to the description of Shakura & Sunyaev (1973), α measures the efficiency of local angular momentum transport on length scales comparable to the scale height H and with velocities on the order of the sound speed c_s. For a thorough derivation and discussion of the α parameters and its caveats, see Appendix A.3. Typical values in observed disks are low, up to $\alpha =\sim 10^{-3}$ (Flaherty et al., 2015, 2017; Teague et al., 2016), corresponding to mass accretion rates on the star of $\dot{M} = 10^{-8} - 10^{-7} M_{\odot}/yr$. Theoretical estimates for various turbulent processes expect values ranging from $\alpha = 10^{-4} - 10^{-2}$ for the magneto-rotational instability, up to $\alpha = 10^{-1}$ for gravitational instability (Rafikov, 2005; Gammie, 2001).

Disk Thermodynamics

Disks are heated both internally through the dissipation and thermalization of kinetic energy due to accretion and viscous stresses or from external irradiation either from the central star or from another nearby source (McKee & Ostriker, 2007; Pollack, 1984). The turbulent processes described above give rise to an effective viscosity which generates heat as mass is accreted onto the star

$$Q_+ = \frac{9}{8}\nu\Sigma\Omega^2, \tag{1.5}$$

where ν is the effective viscosity and Ω is the Keplerian frequency defined earlier. Most studies of self-gravitating disks consider the disk to be thin, and the surface density Σ

is the vertically integrated density from one pressure scale height below ($-H$) to one above ($+H$): $\Sigma \approx 2\rho H$. The disk is considered nonmagnetized in the outer regions and turbulence is generated solely by the hydrodynamics and self-gravity of the disk.

Irraditation from the central star becomes the dominant energy source in outer regions of the disk due to the differing radial dependencies of accretion heat and irradiation (Kratter & Murray-Clay, 2011). Irradiation depends on the stellar flux and so decreases with $\sim R^{-2}$ while the heat generated by viscous accretion (Equation (1.1)) decreases according to $\sim R^{-3}$ meaning that as one moves from the inner disk to the outer disk, there is a radius at which the energy regulation in the disk transitions from dissipation of accretion to irradiation. Where irradiation is dominant, the disk may become vertically isothermal as the constant radiation bath penetrates to the disk midplane to establish a uniform temperature (Kratter & Murray-Clay, 2011).

Heat generated at or near the disk midplane by turbulent viscosity escapes the disk primarily through the disk atmospheres to the top and bottom. In radiating through disk material between the midplane and the surface, the resultant temperature and energy at the surface is dependent on how much energy was absorbed by the dust and gas opacities of the disk material. As an optically thick medium, energy that is absorbed by the gas and dust in the disk limits cooling and helps the disk retain heat. Once reaching the photosphere of the disk, it will be radiated away. If a disk cannot radiate heat away efficiently, it will maintain higher local sound speeds, which disrupts the ability of the gas to settle together (Hubeny, 1990; D'Angelo et al., 2010).

Hydrodynamic Instabilities

After initial infall of material onto the star and disk, accretion needs to be driven by various turbulent processes, as molecular viscosity is too weak to explain the disk lifetimes observed. Hydrodynamic (HD) and magnetohydrodynamic (MHD) instabilities are the most commonly invoked phenomena to describe efficient angular momentum transport and may arise from a wide variety of conditions (Pollack, 1984; McKee & Ostriker, 2007). The most prominent hypothesis for the generation of turbulence in a disk is the *magnetorotational instability* (MRI), where the sheared magnetic coupling between ions and the subsequent rebound of the magnetic field result in "fingers of high and low angular momentum to interpenetrate" (Balbus & Hawley, 1991), however, the inclusion of non-ideal MHD effects leads to significant turbulent deadzones which inhibit the effectiveness of MRI turbulence (Lesur et al., 2014).

If magnetohydrodynamic instabilities do not result in turbulence, there are several hydrodynamic processes which could lead to efficient angular momentum transport. The baroclinic instability, where radial disk convection is induced by a radial entropy gradient, creates vortices as gas is heated at warmer regions closer to the star and buoyantly migrating outwards before cooling off due to thermal relaxation and migrating back towards the star and repeating the process all over again (Klahr & Bodenheimer, 2003; Klahr, 2004). The vortices generated by this instability may trap the solid particles necessary for planetesimals formation and thus be crucial for accelerating planet formation. When there are vertical and radial gradient of both the angular momentum and entropy, a vertical shear instability (VSI) arises when parcels of gas move upwards as they cool, but do so along lines of constant angular momentum, which are not vertically straight but slightly parabolic (Goldreich & Schubert, 1967; Fricke, 1968; Nelson et al., 2013).

Dust and Gas Co-evolution

While dust particles only make up a small part of the disk by mass, as a fundamental component of planet formation, how solids of varying sizes and compositions evolve over the course of the disk lifetime is extremely important to understanding planet formation and disk observations (Shi et al., 2016; Gibbons et al., 2012; Pohl et al., 2016). There are a few fundamental physical differences between solids and gas which result in special behaviors. For one, unlike the gas, particles do not feel the effect of gas pressure gradients, meaning the gas and dust will have different velocities. These velocity differences will result in a headwind or tailwind on the particles to move towards pressure maxima regions (Weidenschilling, 1977). Additionally, solids and gas interact with each other through mutual drag forces, causing them to concentrate to reduce their collective drag, known as the streaming instability (Youdin & Johansen, 2007; Schreiber & Klahr, 2018). Dust in a protoplanetary disk also sediments significantly and will create a thin, dense particle layer with a scale height H_d smaller than that of the gas (Dubrulle et al., 1995)

$$H_d = \left(1 + \frac{\mathrm{St}}{\alpha}\right)^{-1/2} H, \tag{1.6}$$

dependent on the particle size St and turbulent viscosity α.

Solids will eventually grow to a size which drifts inwards fast enough that further growth is negligible before destruction by accretion on to the star, or the so-called fragmentation barrier. There are however, many potential ways to trap particles in rings or lopsided disk structures (Birnstiel et al., 2013; Raettig et al., 2015), which are apparent in observations (van der Marel et al., 2013; Pohl et al., 2017). Since observations of solids larger than $\sim$ 1m are challenging if not impossible, how these solid concentrations form and how they grow to become a planetesimal distribution that forms planets remains an open question (Lenz et al., 2019).

1.2 Planet Formation

Planet formation is broadly broken down into two categories: cold-start or bottom-up planet formation, wherein small particles grow from micron-sized grains to pebbles to planetesimals to planetary cores which accrete surrounding gas to form atmospheres, known most commonly as *core accretion* (CA). A hot-start or top-down planet formation scenario is one where the rapid gravitational collapse of any of the gas or solid constituents of a primordial planet, known as *gravitational instability* (GI) or disk fragmentation.

Protoplanetary disks fragment when a region of gas is unstable to its own gravity and forms an overdense region of gas within the disk. Giant planet formation by CA and GI have very different formation regimes, with only the latter able to form gas giant planets in wide orbits around young stars, as in the systems of HR8799 (Marois et al., 2010), Fomalhaut (Kalas et al., 2008), and HD 95086 (Rameau et al., 2013). Disk fragmentation may lead to the formation of brown dwarfs in addition to large gaseous planets, so the term *companion* is sometimes used to describe its products.

The Core Accretion Scenario

Core accretion forms planets by slowing aggregating larger and larger solid objects with one another within the disk, growing from around molecular scales to pebbles to planetesimals. When enough planetesimals collect together they form planetary embryos and

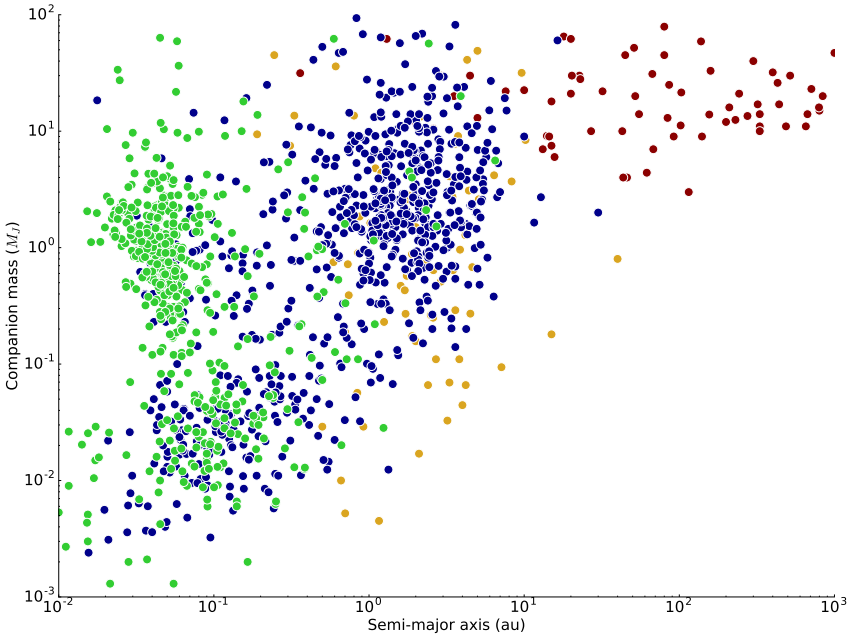


Figure 1.2: Plot of observed planets by mass and semi-major axis. Planet color indicates detection method: blue for radial velocity, green for transit spectroscopy, yellow for microlensing and red for direct imaging. Made using data from exoplanet.eu.

cores, which then begin slowly accreting gas from the surrounding disk (Lissauer, 1993). This is currently the best explanation for the formation of rocky planets and some gas planets, particularly those at radii within 10 au (Mordasini et al., 2012). Core accretion is particularly efficient at radii less than 10 au due to the high amount of solid material and low relative velocities (Safronov, 1969).

$$\frac{dM_{\text{solid}}}{dt} = \pi R_{\text{capt}}^2 \sigma_s \Omega F_s \tag{1.7}$$

This assumes that a planetesimal grows only through direct collections with the geometrical cross-section πR_{capt}^2 for a surface density of smaller solid objects σ_s. Planetesimals roughly larger than a kilometer will further gravitationally focus material into collisions, warranting the inclusion of a focusing factor F_s.

While core accretion might be the typical mode of planet formation (Janson et al., 2011) it is a slow process in outer disk regions, and core growth timescales are longer than the lifetime of the disk because solid material is not as well concentrated. Thus, explaining gas giant planets which are discovered far beyond 30 au is problematic without reliable outward planet migration. However, for gravitational instability the timescale at large radii is on the order of a few orbital timescales, corresponding to $10^3 - 10^4$ years, which allows for an alternate formation explanation.

Pebble Accretion

One possible solution that falls within the core accretion paradigm is that of pebble accretion. Assuming planetesimals form within the gaseous disk, their ability to accrete additional solid material is greatly enhanced by the gas drag in addition to the gravitational potential of the planetesimal. Solids which are small enough to be slowed considerably by the drag with the gas in the vicinity of the planetesimal will spiral onto it (Ormel & Klahr,

2010). Larger objects, such as other planetesimals, will pass by without being accreted. Population synthesis models, including those by Bitsch et al. (2015) have shown this is a very efficient method of producing gas giant planets out to $30 - 50au$, but growth in the extremely distant regions beyond 50 au is still too slow to form planets *in situ*.

Fragmentation via Self-Gravity

Gravitational instabilities occurs when regions of a gaseous protoplanetary disk become unstable to their own gravity and collapse (Kuiper, 1951). At distances relatively close to the young star, sufficient mass is available but temperatures are too high and relaxation rates too slow for gravitational contraction to overcome the local pressure support. Thus it is at distances beyond $\sim$ 50 au, where disks are cooler, where protoplanetary disks are expected to became gravitationally unstable when they have large enough mass to collapse (Boss, 1997). When still young, outer regions of the disk still replenish gas from the surrounding envelope and can maintain high enough surface densities for gravitational collapse to occur (Armitage, 2011). Thus discoveries of planets with large orbital radii (Rameau et al., 2013; Marois et al., 2008) suggest that GI may be a feasible formation mechanism within this region.

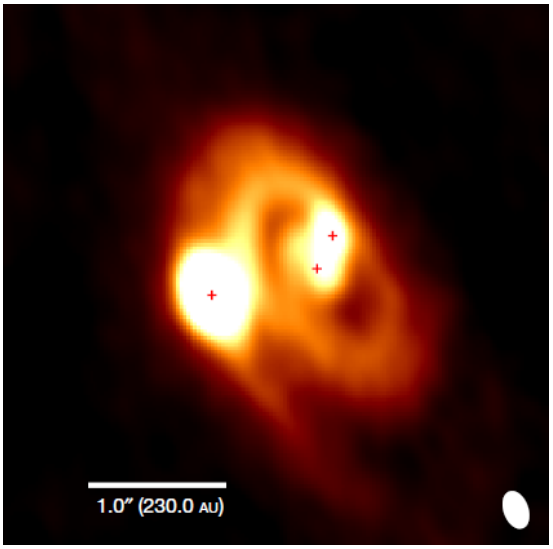


Figure 1.3: ALMA image at 1.3 mm of the system L1448 IRS3B, which appears to show the formation of a fragment (leftmost feature) around a circumbinary disk around a pair of small young stars (Tobin et al., 2016a).

As a theory, GI has difficulties explaining the metallicity of giant planets, which are host sizable solid cores (Bolton et al., 2017) and are typically more metal-rich than the interstellar medium (Kreidberg et al., 2014), which can be well-explained by core accretion. As a process which is more similar to star formation, GI is expected to produce objects with similar compositions to stars (Helled & Bodenheimer, 2010). Futhermore, GI requires very massive disks, which are rare, and planet formation by GI does not explain the existence of moons, asteroids and other small rocky objects (Lissauer, 1993; Bodenheimer & Lin, 2002). Combined with the inability of core accretion to form giant planets at large distances, neither hypothesis provides a complete description of planet formation based on observed planet distributions.

Gravitational disk fragmentation can however explain massive, large-radii planets far better than core accretion since the formation timescale is much shorter at large distances from the star. Given their early formation, fragments then have plenty of time to migrate inwards to locations which would affect their metallicity and stability (Mercer & Stamatellos, 2017; Ilee et al., 2017). GI also provides an explanation of eccentric orbits and has fewer problems with the short orbital decay timescale of planet migration processes (Bodenheimer & Lin, 2002).

Therefore while the exact process of giant planet formation is not perfectly described by one scenario or the other, in tandem they cover the deficiencies of the other (Boley, 2009).

This only considers that planets do not move during or after formation which is generally not the case (Kley & Nelson, 2012). Migration in self-gravitating disks in particular is not well-characterized and results vary as to how fast and in which direction migration will occur (Baruteau & Masset, 2008; Stamatellos & Inutsuka, 2018).

1.3 Gravitational Instabilities

Among the non-magnetic hydrodynamic disk instabilities, gravitational instability is special for its particularly strong but short-lived turbulence, driving strong accretion and migration through large scale spiral arms which efficiently redistribute angular momentum. GI becomes a significant factor in the disk when enough cold gas is present for strong self-gravitation (Toomre, 1964; Goldreich & Lynden-Bell, 1965) and the disk cools efficiently (Gammie, 2001; Durisen et al., 2007; Malygin et al., 2017). These two conditions are measured by the Toomre parameter Q and a simple cooling timescale t_c, respectively. Here I present order-of-magnitude estimations and qualitative descriptions for both Q and t_c.

The Poisson Equation

At the heart of any description of a self-gravitating medium is the calculation of the potential ϕ from the underlying spatial distribution of the density $\rho(x, y, z)$. In the Newtonian limit, this takes the form of

$$\nabla^2\phi(x, y, z) = 4\pi G\rho(x, y, z). \tag{1.8}$$

Strategies to solve the Poisson equation come in many varieties, a few of which will be discussed in greater detail in later chapters depending on the computational method employed. Fourier methods will be used in the periodic domains of the shearing boxes used in Chapters 2 through 4 while the implementation of a multigrid solver will be the focus of Chapter 5. Suffice to say for the moment, solutions to the Poisson equation are added to the equation of motion as an additional source term $-\nabla\phi$.

Toomre Instability

The Toomre stability parameter is an important quantity for understanding what happens during the gravitational collapse of matter in a disk. It represents the ratio of counteracting centripetal and gravitational forces when a small quantity of matter condenses, which can then be easily extended to larger amounts of matter (Toomre, 1964). Starting from the linearized equations of motion and the Poisson equation in a 2D disk (see the derivation of shearing sheet coordinates in Appendix A.2)

$$\frac{\partial u}{\partial t} + \Omega\frac{\partial u}{\partial \theta} - 2\Omega v = \frac{1}{\Sigma}\frac{\partial P}{\partial r} - \frac{\partial \phi}{\partial r}\bigg|_{z=0} \tag{1.9}$$

$$\frac{\partial v}{\partial t} + \Omega\frac{\partial v}{\partial \theta} - \frac{\kappa^2}{2\Omega}u = \frac{1}{r\Sigma}\frac{\partial P}{\partial \theta} - \frac{\partial \phi}{\partial \theta}\bigg|_{z=0} \tag{1.10}$$

$$\frac{\partial \Sigma}{\partial t} + \Omega\frac{\partial \Sigma}{\partial \theta} + \frac{1}{r}\frac{\partial}{\partial r}[r\Sigma u] + \frac{\Sigma}{r}\frac{\partial v}{\partial \theta} = 0 \tag{1.11}$$

$$\frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2} = 4\pi G\Sigma\delta(z) \tag{1.12}$$

To describe the disk with a dispersion relation, the velocities u_x and u_y, surface density Σ and gravitational potential ϕ can be represented in the form of waves

$$u = u_0 e^{i(m\phi + kr - \omega t)} \tag{1.13}$$

$$v = v_0 e^{i(m\phi + kr - \omega t)} \tag{1.14}$$

$$\Sigma = \Sigma_0 e^{i(m\phi + kr - \omega t)} \tag{1.15}$$

$$P = P_0 e^{i(m\phi + kr - \omega t)} \tag{1.16}$$

$$\phi = \phi_0 e^{i(m\phi + kr - \omega t)} \tag{1.17}$$

where ω is the angular wave frquency, k is the radial wave number, and m is the azimuthal wave number. When inserted into the hydrodynamic equations ((1.9)-(1.12)) under the WKB approximation ($kr \gg 1$), and using the Fourier solution of the gravitational potential ϕ,

$$-i\omega u + \Omega i m u - 2\Omega v = -ik(\phi + P) \tag{1.18}$$

$$-i\omega v + \Omega i m v - \frac{\kappa^2}{2\Omega}u = -\frac{1}{r}im(\phi + P) \tag{1.19}$$

$$-i\omega\Sigma + \Omega i m\Sigma + ik\Sigma u + \frac{\Sigma}{r}imv = 0 \tag{1.20}$$

$$\phi = \frac{2\pi G\Sigma}{k} \tag{1.21}$$

one can reproduce a Lin-Shu dispersion relation (Lin & Shu, 1964; Cossins et al., 2009)

$$(\omega - m\Omega)^2 = \kappa^2 - 2\pi G\Sigma k + c_s^2 k^2, \tag{1.22}$$

which is a linear description of gravitational instability. Considering the most unstable azimuthal mode is $m = 0$ and Keplerian rotation $\kappa = \Omega$, Equation (1.22) simplifies to

$$\omega^2 = \kappa^2 - 2\pi G\Sigma k + c_s^2 k^2. \tag{1.23}$$

When $\omega^2 < 0$ the exponents of Equations (1.13)-(1.16) are real and lead to stationary solutions which grow and become unstable (D'Angelo et al., 2010). The most unstable radial wavenumber k is found by finding the minimum of Equation 1.23, e.g $\partial\omega^2/\partial k = 0$, resulting in

$$k = \frac{\pi G\Sigma}{c_s^2} \rightarrow \lambda_{\rm crit} = \frac{c_s^2}{\pi G\Sigma}, \tag{1.24}$$

which leads to the familiar Toomre Q when assuming the critical wavelength is on the order of a scale height $\lambda_{\rm crit} \sim H$ (Toomre, 1964; Goldreich & Lynden-Bell, 1965).

$$Q = \frac{c_s\Omega}{\pi G\Sigma} \tag{1.25}$$

When the Toomre parameter is below one in a region of a razor-thin disk, that region is expected to condense and form an overdensity. Here, $\Omega = \sqrt{GM/R^3}$ is the orbital Keplerian frequency of the disk, which stabilizes the disk to large wavelength density perturbations and the local speed of sound c_s, stabilizes the disk to shorter density perturbations (Durisen

et al., 2007).

An important distinction needs to be made between what is called the critical wavelength $\lambda_{\rm crit}$, which is the fastest growing wavelength of the linear instability and the Toomre wavelength $\lambda_{\rm T}$, which is the largest unstable wavelength. As long as there is not just one unstable value of k, these wavelengths will be different with $\lambda_{\rm T} > \lambda_{\rm crit}$. The latter has greater meaning the context of the local simulations presented here, as the radial length of the box must at least contain $\lambda_{\rm T}$ to encapsulate Toomre instability in its entirety.

Equation (1.25) can be classified into three regimes based on the value of Q. When $Q \gtrsim 1.5$, some gravitoturbulence may arise but non-axisymmetric structures analogous to spirals are unlikely to form. Especially for values above $Q = 2$, local self-gravity is less than the gravitational potential of the central object. Thus the regime of $Q \gtrsim 1.5$ is of minute importance to a discussion of self-gravity and particularly fragmentation.

More important is the regime of gravitoturbulence between $1 \lesssim Q \lesssim 1.5$ where non-axisymmetric structures begin to form and generate the strong shocks and waves that are associated with mass and angular momentum transport. This is a result of growing unstable wavelengths for higher order modes ($m \geq 1$) of the dispersion relation of Equation (1.23) resulting in various multi-arm spiral disks (Lodato & Rice, 2005; Cossins et al., 2009). The strength of this gravitoturbulence has been well-characterized in terms of an *alpha*-viscosity and is typically quite strong compared to other forms of hydrodynamic turbulence, even for long cooling timescales t_c (Gammie, 2001)

$$\alpha = \frac{4}{9}\frac{1}{\gamma(\gamma-1)t_c\Omega}. \tag{1.26}$$

Gravitoturbulence has many fascinating applications, including planetesimal formation (Gibbons et al., 2012; Shi et al., 2016), accretion and

Thermal Saturation

As demonstrated by the dispersion relation of Equation (1.23), regions of the disk with Toomre values $Q \leq 1$ are linearly unstable to axisymmetric ($m = 0$) collapse as well as to nonaxisymmetric perturbations. The non-linear evolution of gravitational instabilities show that the axisymmetric perturbations will contract into bound clumps, while a region with $Q > 1$ will remain stable to gravitational collapse (Toomre, 1964).

The disk tends to stay close to $Q = 1$ because when above this value the disk can cool leading to instabilities. When below the disk will develop strong shocks that heat and disrupt the disk (Balbus & Papaloizou, 1999). A gravitoturbulent situation will be one where these effects balance each other and the simulation settles. While planets will not be formed when the disk is in this steady state, significant viscosities due to the trailing density structures will efficiently transport angular momentum which will lead to accretion and disk evolution.

Disk Fragmentation

Once a clump has formed it must cool fast enough so that it can collapse with the free-fall timescale, which means that the cooling timescale should be on the same order as the shearing timescale $\beta = (t_c/t_{shear}) \simeq 1$. A cooling timescale longer than the shearing timescale means the fragment will be pressure supported and may not collapse before being disrupted by shear in the disk. The shearing timescale is the same as the dynamical

timescale Ω^{-1}, yielding a cooling prescription

$$t_c = \beta\Omega^{-1}. \tag{1.27}$$

Thus a short cooling time ($\beta_{\rm crit} \approx 1$) is the second necessary condition for creating an unstable fragmenting disk (Kratter et al., 2010). The critical value which separates fragmentation from thermal saturation was first established by Gammie (2001) with a value of $\beta_{\rm crit} = 3$. Subsequent studies have shown a slight dependence of this critical cooling parameter to the adiabatic index, which may increase $\beta_{\rm crit}$ to roughly 6 (Rice et al., 2005), but the real problem will be the non-convergence of the cooling parameter with resolution which will be discussed further in the following section.

Sufficient cooling in outer regions of the disk has recently been a major issue to the viability of GI (Rafikov, 2005). It was first shown shown that for short enough cooling times it is possible to form unstable spiral arms and gravitationally bound clumps that survive the shear of the disk (Tomley et al., 1991, 1994). A region of a disk that cools fast enough will reduce the thermal pressure support and allow for gravitational contraction of the gas overdensity.

After initial collapse, a fragment must still survive the disruptive effects of shock and compression heating and tidal forces. The former may slow down the collapse long enough to allow tidal shear to pull the fragment apart (Kratter et al., 2010). Tidal forces are the result of the force gradient due to the central mass on an extended fluid body. Combined with the differential rotation in the disk, the effect is to stabilize the disk against gravitational collapse.

Once formed, fragments represent pre-secondary collapse gas giants, brown dwarfs or even low-mass stars, equivalent to the first Larson core in the case of star formation (Larson, 1969). As will be shown in Chapter 4, fragments at this stage have a very star-like structure, with a Bonner-Ebert-like density profile and roughly isothermal temperature profile. What becomes of the fragment ultimately comes down to how much mass is enclosed and how how the interior becomes. High internal temperatures and pressures will result in the dissociation of hydrogen molecules and a mush softer equation of state, which effectively cools the core and results in a second collapse (Bhandare et al., 2018). This change in equation of state and the secondary collapse is not modeled in the hydrodynamic models used here and is therefore beyond the scope of this thesis.

1.4 Relevance to Planet and Disk Observations

Disk and Planet Observations

Early images of disks showed signs of spiral arms (Muto et al., 2012; Grady et al., 2013), a potential indicator of gravitational instability although as scattered light images, these spirals could be visible by temperature, and thus pressure scale height, variations and not density waves. Later observations of the same systems with ALMA showed that these spirals were indeed density waves down to the midplane. Recent observations have observed the occurrence of fragmentation around small and large stars (Tobin et al., 2016a; Ahmadi et al., 2018; Ilee et al., 2018) or the spiral arms of self-gravitating disks (Pérez et al., 2016) in not just scattered light images, but also millimeter dust emission.

Since the self-gravitating disk regime is very short-lived, it is often best to look at discovered planets as possible indicators of gravitational instability. Theoretical constraints can be used to create a region where gas giant planets are expected to be formed and

observations of large planets far from their star can test the applicability of these constraints. Among the best examples of companions formed by gravitational instability are those around HR 8799 (Marois et al., 2010; Maire et al., 2015), Fomalhaut (Kalas et al., 2008), and HD 95086 (Rameau et al., 2013) among others (Bonnefoy et al., 2014; Maire et al., 2018; Bonnefoy et al., 2018; Cheetham et al., 2018). The ultimate goal is to understand what role gravitational instability plays in the formation of gas giant planets in the circumstellar disks around young stars.

Core accretion is the generally accepted planet formation hypothesis, that pairwise mergers of solid objects leads to the formation of planetesimals which in turn form rocky planets within 30 au of a young star. Particularly massive cores will be able to accumulate significant gaseous envelopes to become gas giant planets provided a gaseous disk still exists. However, this does not explain how to form large planets in outer regions of the disk without invoking significant outward migration or scattering, and thus disk instability offers a niche formation mechanism that can form planets in these regions given the disk is cool, massive relative to the star and cools rapidly. While these are difficult conditions to meet and making disk fragmentation rare, observations also indicate that there are not many planetary mass stellar companions which are likely to be formed through disk fragmentation (Janson et al., 2011, 2012; Vigan et al., 2017).

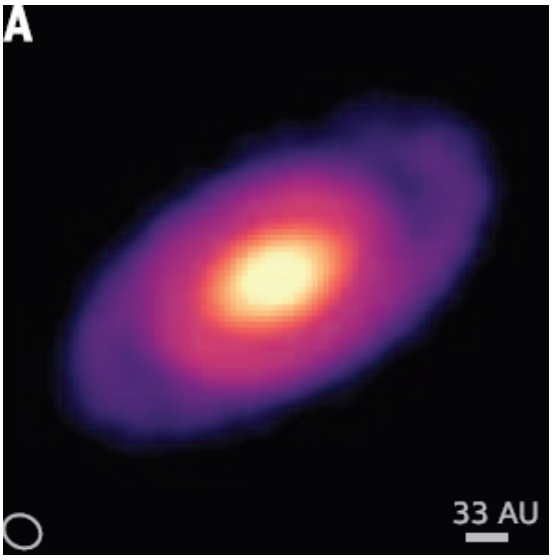


Figure 1.4: Thermal dust emission at 1.3 mm of the disk Elias 2-27, showing the presence of large spiral structure, perhaps indicative of disk instability caused by self-gravity (Pérez et al., 2016).

There is however reason to believe that disk fragmentation may also form stellar mass companions, particularly binary systems with a separation of approximately ~ 100 au (Tobin et al., 2013, 2016b). Observations of young massive stars indicate the existence of Keplerian disks and gravitationally unstable regions within (Ahmadi et al., 2018; Beuther et al., 2018; Ilee et al., 2018). This has been anticipated by theorists, since the initial fragment mass may be in the planet or brown dwarf regime, later accretion will provide additional mass (Kratter et al., 2010).

Constraints on Giant Planet Formation

A niche regime of planet formation by gravitational instability can be defined by the conditions established by Equation 1.25 (Toomre, 1964; Goldreich & Lynden-Bell, 1965) and Equation 1.27 (Gammie, 2001) and result in a formation region for gravitational instabilities (Clarke, 2009; Janson et al., 2012). The Toomre parameter is used to establish a minimum mass below which there is not enough mass within a few disk scale heights H for collapse by self-gravitation to proceed. This mass can roughly be estimated as the mass within a Toomre wavelength (Forgan & Rice, 2011).

$$M_{\text{frag}} \approx \lambda_T^2 \Sigma = (2\pi H)^2 \Sigma, \tag{1.28}$$

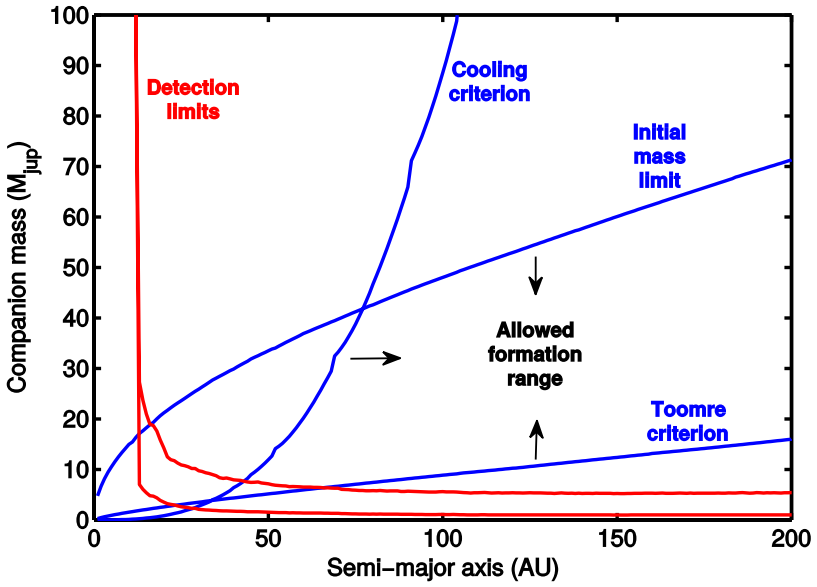


Figure 1.5: Mass and distance formation constraints of companions by gravitational instability. The Toomre parameter is extrapolated to a minimum mass limit and the cooling criterion is dependent on the radius. Detection limits are for the direct imaging of companions around host stars and the more restrictive limit represents younger, brighter stars which make it more difficult to detect companions. Figure from Janson et al. (2012).

where λ_T is the local Toomre wavelength, the largest unstable mode in linear Toomre theory, and Σ is the surface density of the disk. This is an upper limit on the initial mass of a fragment, although initial estimates may vary by up to an order of magnitude, depending on the assumptions used (Boley et al., 2010; Kratter & Lodato, 2016).

The cooling criterion leads to a limit on the semi-major axis of a planet due to the decrease in thermal relaxation time with radius (Malygin et al., 2017). With larger radii, thermal relaxation times become shorter and only in the furthest reaches of the disk are they short enough to meet the condition of Equation (1.27) and allow for gravitational collapse (Johnson & Gammie, 2003; Rafikov, 2005). Therefore one expects to find GI-formed giant planets beyond a certain radius where the disk can cool efficiently.

The most reliable technique for detecting planets potentially formed by disk instability is direct imaging, since it is biased towards objects which are massive enough to be detected by extremely sensitive photometry and far enough away from their parent star that they can still be resolved as separate objects (Marcy & Butler, 1998).

These constraints are combined in Figure 1.5 (Janson et al., 2011, 2012) to create a mass-radius formation region for giant planets by gravitational instability. While this is a large range and is not restricted by detection biases, there are very few companions found in this allowed formation range, suggesting that even if planet formation by disk fragmentation is possible, but unlikely (Vigan et al., 2017).

The possibility that the cooling timescale required for gravitational instability could be longer than shown here might mean that GI forms planets at shorter orbital radii and thus making disk instability less restrictive than previously thought.

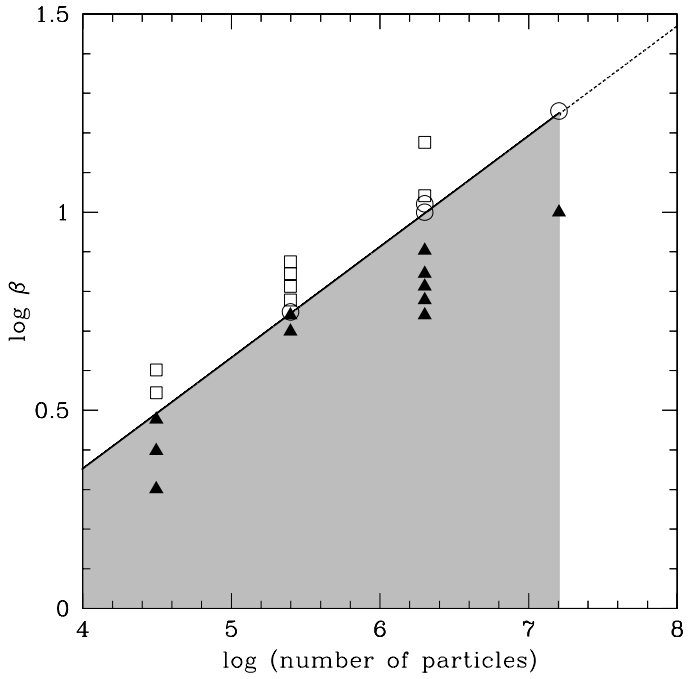


Figure 1.6: Cooling parameter β as a function of number of particles, the SPH measure of resolution similar to grid cells in finite difference methods. Open squares represent stable non-fragmenting simulations, solid triangles are fragmenting cases and open circles are borderline situations. (Meru & Bate, 2011)

Non-convergence of the Critical Cooling Criterion

Complications arise when as Meru & Bate (2011) found, $\beta_{\rm crit}$ changes with resolution, initially for global smoothed particle hydrodynamic (SPH) simulations and later followed by local finite-difference shearing sheet models of Paardekooper (2012). These results indicate that the critical cooling timescale may be significantly longer, up to $t_c = 20\Omega^{-1}$ (see Figure 1.6), extending the GI formation domain to regions of the disk potentially within 30 au (see Figure 2 of Janson et al. (2012)). This conflicts with the current understanding of formation regimes of core accretion and gravitational instability, suggesting that GI may be more common than detected by current observations (Janson et al., 2011, 2012). Additionally, this is a significant challenge to the applicability of previous studies as a resolution dependent solution implies that other simulations with $\beta_{\rm crit} = 3$ were using an inappropriate value.

As will be detailed in the second introductory chapter, Chapter 2, the β-prescription (1.27) cools every location at the same rate, failing to account for the effects of optical depth on cooling efficiency or the increased strength of surface density fluctuations with resolution. To account for the varying optical depth with density, I altered the cooling prescription to include a linear dependence on the local surface density in the disk. This will suppress the formation of fragments from strong density fluctuations which will cool slower and be supported against collapse by higher internal temperature. Conversely, underdense regions will cool faster and prone to collapse, perhaps leading to convergence of the fragmentation boundary with resolution.

Other studies into the effect of thermodynamics on disk stability have focused on different methods, such as radiative cooling (Boley et al., 2006; Cai et al., 2006) and modifications to the equation of state (Laughlin et al., 1998) or have concentrated on different

sources of heat such as irradiation from external sources (Rice et al., 2011; Kratter & Murray-Clay, 2011; Vorobyov et al., 2015). The most consequential effects on how disk fragmentation scales with resolution turns out to be not the cooling prescription, but the artificial viscosity (Deng et al., 2017) and the inadequacies of approximating a 3D disk with a 2D model (Müller et al., 2012; Young & Clarke, 2015).

This thesis will proceed by introducing the reader to the simulations I used to investigate the non-convergence of this critical cooling parameter. I proposed a new cooling prescription to maintain the limited formation range of planets by gravitational instability without resorting to expensive full radiative transfer hydrodynamics. Flux-limited diffusion provides a more rigorous treatment of radiative transfer, but the simple cooling prescription employed here allows for a direct comparison of the relevant timescales to determine a fragmentation criterion.

2 Gravitational Instabilities in a Razor-thin Local Approximation

In the following introductory chapter, the work presented is based on simulations from the author's master thesis (Baehr, 2015). Results of simulations after submission of the thesis changed the ultimate conclusion of the project and the final results are expounded upon here and in the publication Baehr & Klahr (2015). Nevertheless, this work provides an important segue and provides a strong motivation for the actual thesis work in the subsequent chapters.

The linear analysis of gravitational instabilities is limited to the vertically-integrated thin disks. Self-gravitating disks are naturally quite flat because the already thin hydrostatic balance imposed by a central potential is further flattened by the potential of the relatively dense gas.

Being able to reduce the vertical dimension to a integrated quantity is useful for attaining high radial and azimuthal resolutions, which is necessary to avoid spurious and artificial fragmentation due to the growth of truncation errors (Truelove et al., 1997; Nelson, 2006). Thus, many local simulations which study gravitoturbulence and fragmentation have limited themselves to 2D studies (Gammie, 2001; Johnson & Gammie, 2003; Gibbons et al., 2014; Paardekooper, 2012) in an effort to focus on the radial and azimuthal scales more relevant to gravitational instabilities, as well as on other relevant physics, such as particle-gas interaction, radiative transfer and magnetic fields.

As this chapter will demonstrate, 2D simulations are useful for attaining high resolutions, but must carefully consider the evaluation of the Poisson equation in doing so. The solution to the Poisson equation does not always translate perfectly from a three dimensional problem to a two dimensional solution without caveats. Thus the differences between 2D and 3D models of gravitationally unstable disks is one of the primary motivations of this thesis. The current chapter will establish the limitations of 3D models so that beginning with Chapter 3, the primary thesis work can explore the value of 3D models.

2.1 Hydrodynamic Models

All simulations were run using the finite difference code Pencil, which solves the compressible hydrodynamic equations in a shearing box with a sixth-order explicit integration scheme. A shearing box simulation means the hydrodynamic equations are fully linearized and expanded around an orbital distance from a central star R in a co-rotating, sheared frame (see derivation and discussion in Appendix Section A.2), and the mass, momentum,

and energy conservation equations respectively take the following form

$$\frac{\partial \Sigma}{\partial t} - q\Omega x \frac{\partial \Sigma}{\partial y} + \nabla \cdot (\Sigma \boldsymbol{u}) = f_D(\Sigma) \tag{2.1}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} - q\Omega x \frac{\partial \boldsymbol{u}}{\partial y} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} = -\frac{\nabla P}{\Sigma} + q\Omega v_x \hat{\boldsymbol{y}} - 2\Omega \times \boldsymbol{u} - \nabla \phi + f_\nu(\boldsymbol{u}) \tag{2.2}$$

$$\frac{\partial s}{\partial t} - q\Omega x \frac{\partial s}{\partial y} + (\boldsymbol{u} \cdot \nabla) s = \frac{1}{\Sigma T} \left(2\Sigma \nu \boldsymbol{S}^2 - \Lambda + f_\chi(s) \right), \tag{2.3}$$

where $\Sigma \approx \rho H$ is the mass surface density, $\boldsymbol{u}$ is the gas velocity in the Cartesian co-moving frame, Ω is the orbital frequency, q is the shear parameter ($q = 3/2$ for a Keplerian disk), P is the pressure, ϕ is the self-gravitational potential, s is the entropy, the thermodynamic variable, T is temperature, $\boldsymbol{S}$ is the rate-of-strain tensor, ν is the viscosity and Λ is the simple cooling prescription (see Section 2.2). The source terms $f_D(\Sigma), f_\nu(\boldsymbol{u}), f_\chi(s)$ are hyperdissipative terms which are described in more detail in Appendix A.3.

Since this shearing frame and the equations (2.1) - (2.3) above are expanded around a co-rotating point this class of simulations is called local because they model the physics around R, but do not model the central object or any global gradients. Further details of the Pencil code and the methods used can be found in the following chapter and in Appendix Section B.2. In this chapter I focus on the cooling prescription (Λ in Equation (2.3)), how it affects fragmentation, and whether it can be modified to converge the fragmentation criterion.

2.2 Cooling Parametrization

Since fragments become especially sharp density perturbations which become worse with improved resolution, one possible solution is use a modified cooling prescription to suppress fragmentation. This works because the cooling prescription typically used for shearing sheets is constant throughout the domain, failing to match the physical reality of increasing optical depths with higher densities. Thus the change to the cooling prescription implemented here makes the cooling time longer in regions denser than the initial condition and shorter in underdense regions, with the expectation that the shorter cooling time will result in higher temperatures and internal pressures which will keep a fragment from collapsing.

More physically accurate models currently used is radiative transfer cooling, which typically requires three dimensional ray-tracing schemes with detailed opacities (see Boss (2001), Boley et al. (2006), (Kuiper et al., 2010) and Cai et al. (2006), among others). Simulating hydrodynamics with coupled radiative transfer is a complicated task that requires a significant amount of computational resources and results are inconclusive as to whether cooling is sufficient to reduce pressure support and form fragments by gravitational instability (Mejía et al., 2005; Boss, 2004; Durisen et al., 2007; Boley et al., 2007). Here, simulations are run with a cooling law such that each grid cell loses an amount of heat per time given by the cooling law in the form $\Lambda = U/t_c$

$$\Lambda = -\frac{\Sigma(c_s^2 - c_{s,0}^2)}{\gamma(\gamma - 1)t_c}. \tag{2.4}$$

where $U = \Sigma(c_s^2 - c_{s,0}^2)/\gamma(\gamma - 1)$ is the two-dimensional energy density with a non-zero background irradiation term $c_{s,0}^2$. The majority of the heat generation in a disk will come

from the internal sources, but the background irradiation term has an effect on the onset of fragmentation. A disk with no irradiation will allow the gas to cool to zero, which is not realistic of a disk and reduces the pressure support against gravitational collapse and potentially promotes fragmentation without a strong density perturbation, an undesired effect.

The cooling time, given by $t_c = \beta\Omega^{-1}$, is the focus of these simulations because in its default implementation, it is the constant for every region in the domain and is too simple to account for large changes in density and optical depth. This assumes heat escapes a dense gas region with the same efficiency as a thinner region. However, the optical depth of the gas medium, or roughly how many interactions a photon has over a particular distance l, depends on the density of the material to determine how much radiation is absorbed or reflected. Therefore, a denser object should absorb more radiation that passes through it and thus cool slower i.e. with a higher effective cooling constant β.

Thus a new cooling prescription should have a different effective cooling timescale for each grid cell in the simulation domain, cooling denser regions slower and less dense regions faster, thus offering a more realistic approach to disk cooling.

Assuming all energy is radiated away vertically from the surface of the disk, heat is lost at a rate per unit surface area Λ according to the Stefan-Boltzmann law

$$\Lambda \sim 2\sigma T_e^4, \tag{2.5}$$

with Stefan-Boltzmann constant σ and effective surface temperature T_e. The factor of two is due to radiative losses from both the top and bottom surfaces of the disk. From the Hubeny (1990) treatment of an radiative transfer in an optically thick disk by the diffusion approximation, one can express the effective surface temperature in terms of the midplane temperature T_{mid} and the Rosseland mean optical depth τ of the intervening disk material

$$T_e^4 = \frac{8}{3}\frac{T_{\rm mid}^4}{\tau}. \tag{2.6}$$

From this one can estimate the heat loss through the vertical extent of the disk in terms of the midplane temperature (Johnson & Gammie, 2003), as heat generated through viscous processes will primarily originate from the midplane

$$\Lambda = \frac{16}{3}\frac{\sigma T_{\rm mid}^4}{\tau}. \tag{2.7}$$

From this, one can construct a cooling timescale as a ratio of the disk internal energy U and the cooling rate

$$t_c = \frac{U}{\Lambda} = \frac{3}{16}\frac{U\tau}{\sigma T_{\rm mid}^4} \approx \frac{3}{16}\frac{U\Sigma\kappa}{\sigma T_{\rm mid}^4}, \tag{2.8}$$

where the optical depth is estimated in the optically thick regime by

$$\tau = \int_0^z \rho\kappa\, dl \approx \Sigma\kappa, \tag{2.9}$$

when the gas density ρ is assumed constant from the midplane to the disk surface z. Since the opacities κ in the cool regions of the disk are dominated by ice grains and metal grains ($a = 0$), Λ has no additional Σ dependence (Bell & Lin, 1994)

$$\kappa \propto \Sigma^a T^b \tag{2.10}$$

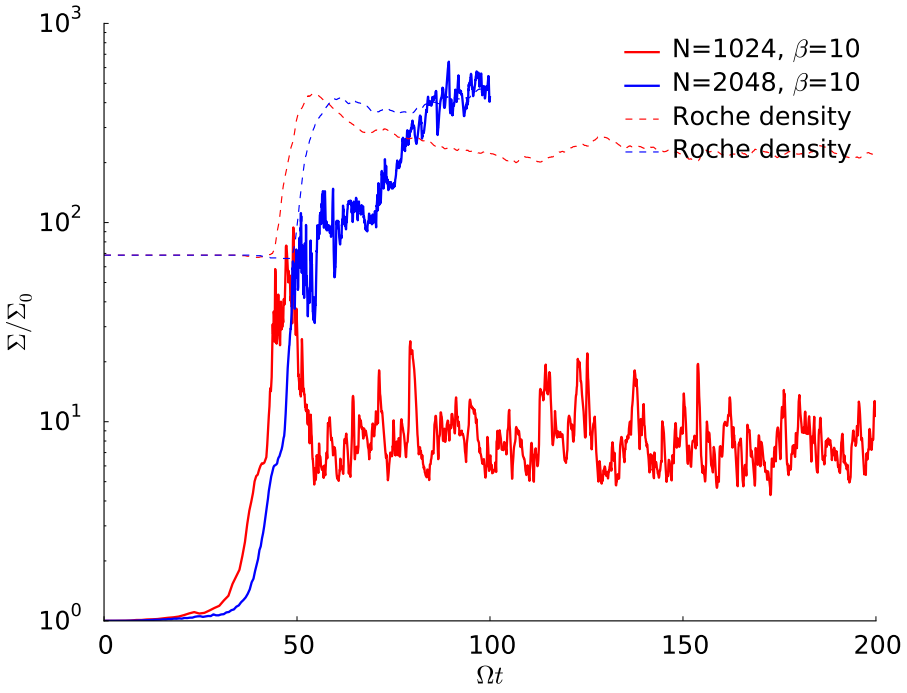


Figure 2.1: Evolution of the maximum surface density (solid lines) in two simulations which have the same initial conditions and run parameters except for the grid resolution. Dashed lines are a Roche density fragmentation threshold, which indicate the formation of a fragment when the maximum density is above this line for more than a few dynamical timescales. Figure from Baehr & Klahr (2015).

Furthermore κ and by extension t_c does not have an extraordinary dependence on temperature (compared to the opacities of hydrogen-scattering and evaporating metal grains higher temperatures) or change dramatically below 1000 K (See Table 3 and Figure 9 of Bell & Lin (1994)). Thus the standard β cooling prescription can accommodate an additional dependence on the local disk surface density, and it will be approximated by the deviation from the initial surface density Σ/Σ_0

$$t_c = \beta \, (\Sigma/\Sigma_0) \, \Omega^{-1}. \tag{2.11}$$

The old cooling prescription is specific to constant large optical depths (Johnson & Gammie, 2003) and this new cooling law will take into consideration the changing opacities in the disk due to local surface densities over the course of the simulation. In this way, Equation (2.11) is similar to the flux-limited diffusion (FLD) approach to radiative transfer (Levermore & Pomraning, 1981), although in a simpler, easier to implement form which is parametrized for easier comparison to disk timescales. Previous studies using the simple cooling law, notably Gammie (2001), have found a critical value of $\beta \simeq 3$, where disks tend to fragment for β values below this critical value and will not fragment above it. It is important to note that this critical value is found for simulations with 1024×1024 grid cells, and will differ with increasing resolution, such is the focus of this investigation.

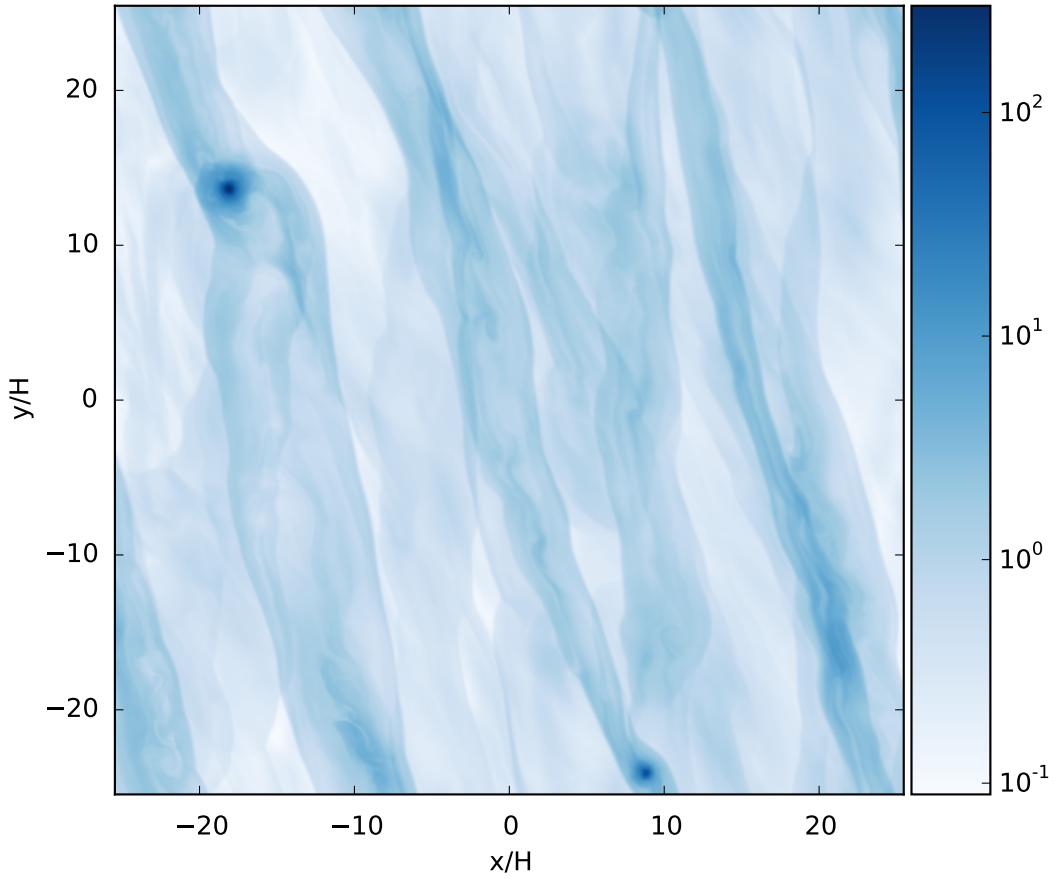


Figure 2.2: Map of the surface density in a fragmenting simulation using the unaltered cooling prescription ($\beta = 2$). Regions with densities above $\Sigma \sim 100$ (dark blue) are above Roche density and will be stable against destruction by shear and are thus considered fragments.

2.3 Fragmentation

Initial simulations run with $N_x = 1024$ and $N_y = 1024$ are established with a uniform surface density distribution and minute velocity perturbations which instigate the growth of instabilities. The linear growth phase begins after a few tens of orbital timescales and the linear collapse is balanced by shocks that develop from trailing density structures and cooled with the constant cooling prescription. As discussed in Gammie (2001), the initial small random velocity perturbations develop into non-linear fluctuations in surface density, velocity and gravitational potential before settling to a steady state in the non-fragmenting case or continuing to grow in the fragmenting case.

This is demonstrated in Figure 2.1 by the solid red line, which shows that the maximum surface density in the simulation did not go above the Roche density (corresponding dashed line in red). The Roche density is point at which a density perturbation has become strong enough to resist being shorn apart by tidal shear

$$\rho_{\text{Roche}} = 3.5\frac{M_*}{R^3}, \tag{2.12}$$

which is expressed in the 2D scale-free local frame as

$$\Sigma_{\text{Roche}} = 7\frac{c_s^2}{HG}. \tag{2.13}$$

In the stable gravitoturbulent case clumps are continuously forming but are torn apart before they are allowed to reach Roche density where they will be able to withstand further disruption. Fragmentation is shown for identical simulation as for the stable case, but with four times the number of grid cells, in Figure 2.1 with the solid blue line. This time fragmentation is clear as the maximum density surpasses the density threshold.

This shows the essence of the problem brought up in Meru & Bate (2011): merely changing the resolution alters the outcome of the simulation. This is a problem for multiple reasons. First, by the Lax-Richtmyer stability theorem, a finite difference scheme converges if and only if it is stable and consistent (LeVeque, 2007). Stable means that the simulation does not develop discontinuities and anomalous values which cause the code to crash. Consistency means the the error approaches zero with increasing resolution. Second, violation of Lax-Richtmyer stability indicates that the problem is possibly poorly posed, likely a result of incorrect assumptions, and thus the results might be unreliable. Third, since nature has 'infinite' resolution, if the critical cooling parameter β_{crit} continues to increase with resolution, at what point does it stop? For these reasons it is important to determine the source of convergence issues and fix the problem. A more detailed discussion of convergence and its consequences can be found in Appendix Section A.8.

In the gravitationally unstable case, fragments cool faster than they can be torn apart by tidal shear and collapse into one or more overdensities that may continue to grow or merge through collisions. Such behavior is expected for simulations with resolution $N = 1024$ and cooling parameter $\beta \lesssim 3$, as a clump will be able to collapse to a compact density before being sheared apart over a few dynamical timescales. A disk is considered to fragment when it has surpassed the Roche surface density.

The difference between the two simulations shown in Figure 2.1 is that the total number of grid cells is quadrupled, which should not lead to such a drastic shift towards fragmentation. This is consistent with simulations by Meru & Bate (2011) and Paardekooper (2012), which indicate the critical cooling criterion of Gammie (2001) may be as high as $\beta_{\text{crit}} = 10$. Thus a constant cooling parameter does not adequately scale with resolution and a new approach to cooling is perhaps needed to observe a convergent fragmentation boundary.

Results with Adjusted Cooling

The results here use parameters identical to the simulations in the previous section, besides the modification to the cooling law described in Section 2.2 by equation (2.11). Figure 2.4 shows the case where cooling is inefficient ($\beta = 10$) is shown for four different resolutions, where all but the highest resolution show no fragmentation. These simulations are a direct comparison to the two shown in Figure 2.1 and in particular, the significant change between the behavior of the $N = 1024$ and $N = 2048$ cases shows the effect of altering the cooling prescription.

This is expected because defining the cooling according to Equation (2.11) creates an effective cooling time for each grid cell. When the disk cools according to the old cooling law $t_c = \beta\Omega^{-1}$, all regions in the disk lose the same amount of heat at the same timescale t_c. However, since one expects a clump to have a higher optical depth $\tau \approx \Sigma\kappa$ cooling efficiency should change from one location to another depending on the surface density.

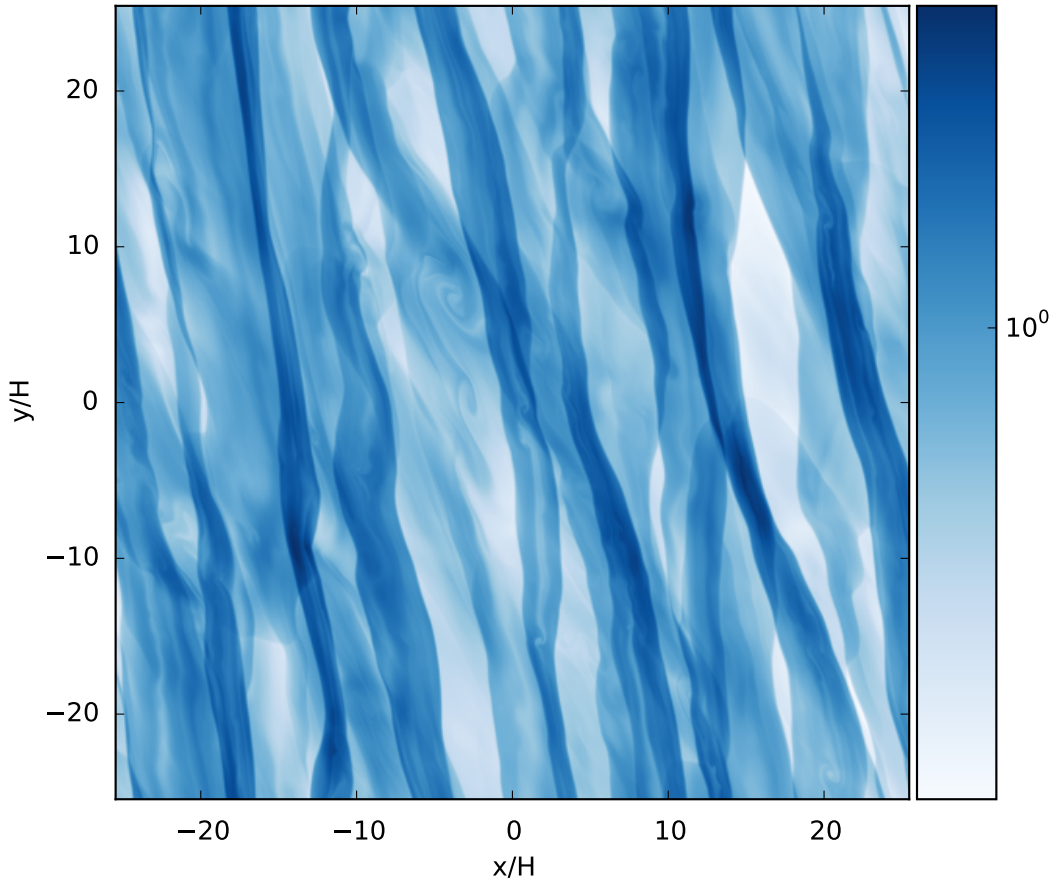


Figure 2.3: A simulation with the same setup as Figure 2.2 ($\beta = 2$), but with the modified cooling timescale designed to inhibit fragmentation. Fragments do not form and what remains is stable gravitoturbulence with non-axisymmetric structures.

This motivates the alteration to the cooling timescale, which causes denser regions of the disk to have a higher optical depth and retain their heat, stabilizing to gravitational collapse. On the other hand, underdense regions will have a lower relative optical depth, cool faster and clump into dense structures more easily.

For resolutions $N \leq 2048$, stability continues to lower cooling parameter values as well, Figure 2.5 showing the results of three simulations with different resolutions at $\beta = 2$. Previously these simulations would have been expected to fragment, but here all show consistent steady gravitoturbulence. This begins to show the shift of the fragmentation boundary towards shorter cooling timescales. Figure 2.6 shows the results of simulations at even short timescales with a new fragmentation boundary anticipated at around $\beta = 0.5$ were it not for the simulations at $N = 4096$ resolution. At this point there is a complete breakdown of the modified cooling prescription and shows that the inability of fragmenting simulations to converge with resolution cannot be solved in this way. The modified cooling prescription manages to delay fragmentation until the resolution is increased by a factor of two, but eventually gravity prevails even for inefficient cooling.

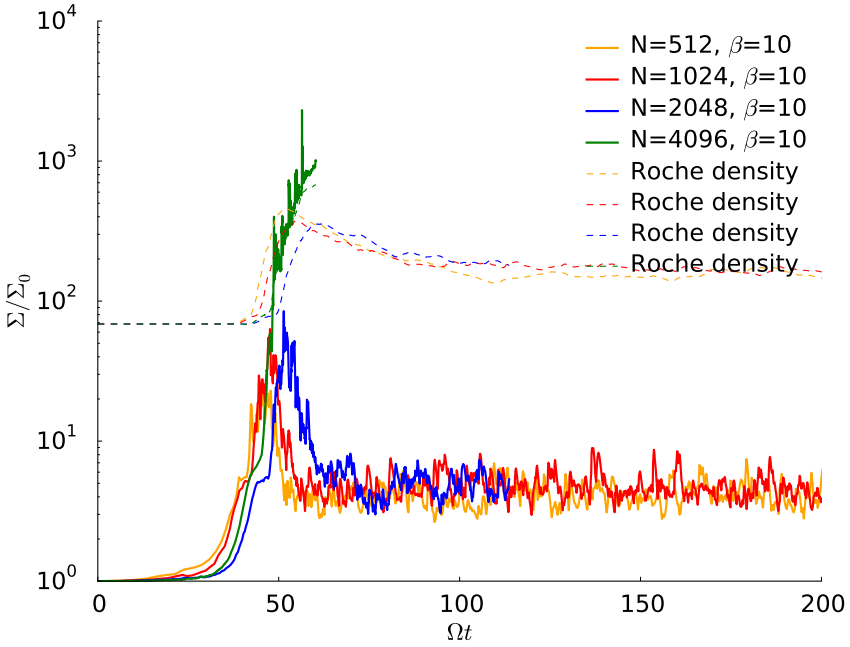


Figure 2.4: The maximum surface densities of the two simulations shown in Figure 2.1 in addition to higher and lower resolutions in green and yellow, but with cooling prescription changed to account for the varying optical depths in the disk. The dashed lines are the corresponding Roche densities which are a fragmentation threshold.

2.4 Validity of the Thin Disk Approximation

Simulations of gravitational instabilities are reliant on sufficient resolution of the Toomre wavelength λ_T by at least 4 grid cells to avoid artificial fragmentation (Truelove et al., 1997; Nelson, 2006). Therefore, 2D simulations have underpinned a large number of studies on the fragmentation criteria of self-gravitating disks. A significant drawback to the 2D approximation for self-gravitating disks is that they do not consider potentially drastic changes in the disk vertical scale height as regions of extreme overdensities and underdensties form (Müller et al., 2012). This was illustrated in Meru & Bate (2011) and Paardekooper (2012) for both smoothed particle hydrodynamic codes and grid codes, respectively, and in both cases the critical cooling criterion of fragmentation was affected by resolution.

This effect was tested in my simulations and is shown in Figure 2.1, which show the evolution over time of the maximum surface density in a simulation. The solid red line shows the expected outcome for a simulation with inefficient ($\beta = 10 > \beta_{\rm crit}$) cooling: heat is not released fast enough from a locally collapsing overdensity and the thermal pressure is enough to prevent the formation of a object dense enough to resist break up by tidal shear. When the linear resolution is increased by a factor of two, as in the case of the solid blue line of Figure 2.1, the same inefficient cooling is now insufficient to keep an overdensity from fragmenting.

Smoothing Self-Gravity

It was ultimately best shown in Young & Clarke (2015) that smoothing self-gravity on the order of a scale height H keeps fragmentation from diverging at high resolutions. A smoothing length works by adding an additional smoothing factor $\epsilon \sim H$ to the two-

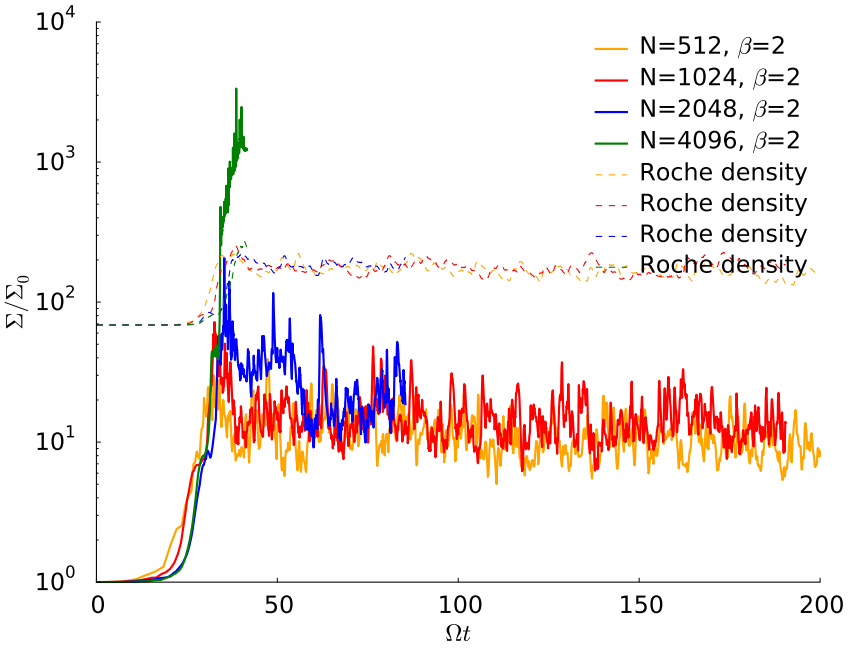


Figure 2.5: Maximum density evolution of simulations with $\beta = 2$. Under the fragmentation criterion of Gammie (2001) these simulations should all fragment, but are instead gravitoturbulent with the exception of the highest resolution.

dimensional self-gravitational potential, written in the Plummer form (Plummer, 1911; Müller et al., 2012)

$$\phi = -\iint \frac{G\Sigma}{\sqrt{s^2 + \epsilon^2}} dA, \tag{2.14}$$

where s is the distance between two points projected onto the 2D radial-azimuthal plane. This is explained by Huré & Pierens (2009) as the result of a 2D gravitational potential integrating differently than a 3D potential. Equation (2.14) without the smoothing term looks like

$$\phi = -\iint \frac{G\Sigma}{\sqrt{s^2}} dA = -\iint \frac{G\Sigma}{s} dA, \tag{2.15}$$

which can be directly integrated, yielding the potential as a natural logarithm of the distance s between two points. This is problematic because the natural logarithm exaggerates the force at small scales compared to the solution to Equation (2.14), which can produce singularities in the potential if gravitational instabilities are able to grow fast enough. Thus it is necessary to include a smoothing length when treating self-gravity in a 2D simulation.

Pencil uses a Fourier self-gravity solver, so in order to add a smoothing length, one limits the wavenumbers k, excluding all values k which correspond to high frequency or small scale density perturbations. The Fourier solver and the inclusion of a smoothing length is described in more detail in Appendix Section A.1.

2.5 Discussion

The results here show that the convergence issue of fragmentation in protoplanetary disks can not be approached by merely refining the cooling prescription as a work around to the application of more sophisticated physics. As simulations reach higher resolutions, they

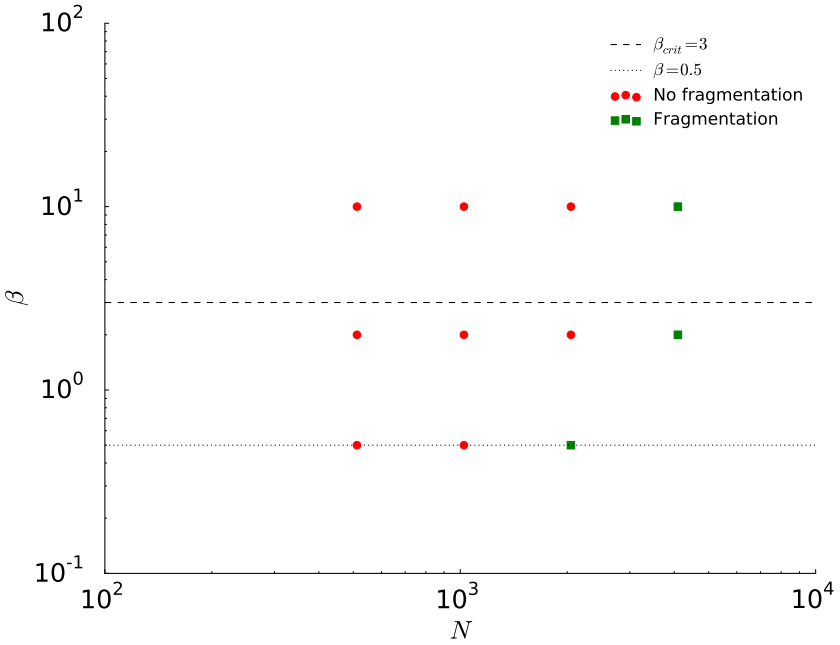


Figure 2.6: Plot of cooling parameter β against number of grid cells in one direction N for all simulations run with the new cooling prescription. The dashed line at $\beta_{\rm crit} = 3$ is the fragmentation boundary as defined by Gammie (2001). The dotted line was the proposed new fragmentation boundary based on the simulations before the pair of $N = 4096$ simulations fragmented and disproved the hypothesis.

will likely need improved physical models to approach convergence.

Convergence

Whereas the standard cooling prescription showed drastic differences in fragmentation behavior between $N = 512$ and $N = 1024$ as seen in Figure 2.1, these two resolutions are consistent over similar cooling timescales, cooling reaching a gravitoturbulent state at similar rates and settling at similar densities. This can be attributed to the sensitivity to density of the new cooling method employed here. When overdensities cool slower than other regions, they retain more heat and are more likely to be disrupted, decreasing the fragment density below the Roche density threshold and are suppressed from further fragmentation.

However this does not mean a new fragmentation boundary has been attained. Figure 2.6 shows different fragmentation behavior for varying cooling times and resolutions and the lack of convergence is noticeable when $\beta = 0.5$. The $N = 1024$ simulation appears closer to fragmentation than the $N = 512$ case and might be considered borderline fragmentation, where a clump surpasses the Roche density, but is sheared apart in less than an orbit ($2\pi\Omega^{-1}$) (Meru & Bate, 2011). This is not the case for $N = 512$ as a clump here never reaches Roche density and at the highest resolution studied ($N = 2048$) the disk fragments. This may be due to the fact that assuming a simple linear relation in surface density does not fully capture the dependence of the cooling timescale on surface density. Also, as the cooling rate is a function of orbital frequency Ω, surface density Σ and temperature T there may be an additional dependence on temperature that must be explored in the future.

Paardekooper (2012) confirmed the non-convergence shown in SPH simulations with a

finite difference code, which led to the assertion that fragmentation might be a stochastic process in circumstellar disks. The implication is that planet formation by GI is inevitable and the only reason that GI is not more prevalent is because the timescale for clumps in weakly cooling disks to achieve fragmentation is longer than the lifetime of the disk. This assumes that once fragmented, clumps do not fall apart and there is no process in the disk that could lead to the disruption of a successfully fragmenting clump. The new cooling implementation becomes weaker for a fragment as it increases in density, offering the necessary resistance to the stochastic formation of growing overdensities such that fragmentation is no longer an eventuality.

Consider the case of a clump which hovers very close to its Roche density, such as the solid blue line in figure 2.5. In this case the simulation forms a clump which should have a better chance of crossing the threshold into becoming a fragment and remaining so. However, even as it manages to form a clump which crosses this threshold once, it still returns to a gravitoturbulent state at simulation time $t = 50\Omega^{-1}$, a result of the local cooling time. For this reason, these simulations do not indicate fragmentation is a strictly stochastic process independent on the strength of fluctuations.

Meru & Bate (2011) and Lodato & Clarke (2011) have suggested that the lack of convergence may be an effect of the numerical setups used, however Rice et al. (2014) did not find an issue with the artificial viscosity. Rice et al. (2014) does however alter they way in which an SPH method cools, using kernel smoothing to spread released heat around to neighboring particles. At high densities this has a similar effect to the cooling used here, with clusters of particles able to share their heat among each other so that dense clumps retain heat and resist collapse. At lower densities cooling is unchanged and this shows in their resulting critical cooling criterion which increases to $7 \leq \beta \leq 9$ compared to the reduction of the criterion in this study. Additionally, the ability of this cooling modification to remain consistent with particle number is uncertain, as it introduces a parameter, smoothing length, which should be scaled with resolution.

Rice et al. (2005) suggests that fragmentation is the result of the disk being unable to withstand the combined Reynolds and gravitational stresses which results in a fragmentation boundary at $\alpha \approx 0.1$. The simulations here do not support a fragmentation boundary at this value as some disks remain stable at values as high as $\alpha = 1$. For simulations with the altered cooling time it is expected that the disk can remain stable to higher stresses because the localized cooling time stabilizes fragments and the disk as a whole.

Limitations

In this investigation, only a small modification has been made to the physics of the circumstellar disk and should not be expected to be a final solution to the convergence issue regarding gravitational instability. There are still some drawbacks to this approach though, as numerous assumptions and simplifications were made for the sake of efficient computation of high resolution physics and these might influence the evolution of the disk simulation as well as the fragmentation criteria.

Radiative transfer is a more physically complete description of the cooling in the disk, but due to its relation to realistic opacities and the need for an additional dimension for effective simulation makes it a complicated option. Implementing radiative transfer in addition to adding a vertical computational direction significantly increases the amount of processing power needed for resolutions similar to what is implemented here. Using a simple cooling timescale in 3D leads to the same fragmentation behavior as Gammie (2001), but including radiative transfer has not been shown to lead to consistent fragmentation,

with cooling by radiative transfer too slow to form fragments as in Cai et al. (2006) and Boley et al. (2006), but not in Boss (2001).

There are still some improvements which can be made to simulations which use simple cooling. A stronger dependence of the cooling timescale on surface density is possible due to an additional factor of Σ in the energy density U of equation (2.8). While opacity remains independent of surface density at low temperatures, the temperature dependence varies greatly at low temperatures (Bell & Lin, 1994) and is difficult to model with this simple cooling prescription. For this reason I have only considered surface density in the cooling timescale and a more complete description would handle the changing temperature dependence of opacity.

This inability of 2D models to capture the small scale interactions and full structure of fragments provides the motivation for the following chapters on fragmentation of gravitationally unstable disks. Chapter 3 will use high-resolution 3D models to constrain the cooling criterion, suggesting fragmentation is limited to the distant reaches of a young disk, and provide estimates for initial minimum fragment masses, which affect whether observers should expect gas giant planets or brown dwarfs. Chapter 4 will show that the inclusion of solids to the process of fragmentation produces stellar companions which have significant solid cores and enhanced overall solid fractions. This is an unusual characterization of objects formed by gravitational instabilities which are expected to have more star-like compositions and very minute solid fractions. Finally, in Chapter 5 I detail my efforts to implement a self-gravity solver for an adaptive mesh code so that it can be used to model the fragmentation process in a full disk with high-resolution of the overdensities and fragments which form.

3 Gravitational Instability in Disks of Finite Thickness

The following chapter is focused on the results of the publication Baehr et al. (2017).

Following the results of Baehr & Klahr (2015) and Young & Clarke (2015) it became apparent that two dimensional simulations would always be limited by the need to smooth the self-gravitational potential on small scales to roughly the disk scale height H. This is necessary to avoid a self-gravitational potential which is too strong at the small scales and exaggerates the growth of density perturbations. Thus, it was important to conduct 3D simulations with a vertically stratified density distribution, which would eliminate the need to smooth self-gravity and allow one to include the effects of small scales, whatever they may be.

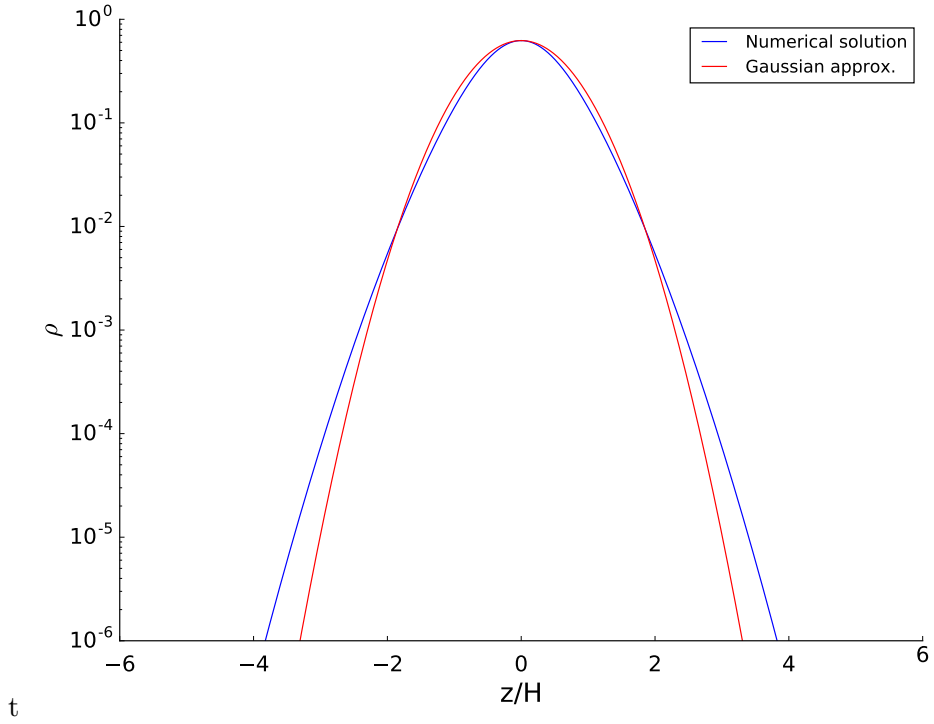
This meant sacrificing one of the principle strengths of a 2D model of gravitational instability, high spatial resolution, but the resolution would still keep the simulations of this thesis well above the limits of (Truelove et al., 1997) and (Nelson, 2006). This disadvantage was partially offset by reducing the simulation domain, maintaining the same grid spacing as the 2D simulations, so that these simulations have the same effective resolution and are more directly comparable.

3.1 Vertical Stratification

In both global and local 2D simulations, it has been well-established both in literature (Meru & Bate, 2011; Paardekooper, 2012) and in this thesis (see Chapter 2) that there is a problem converging the critical cooling parameter for both smoothed particle methods and grid-based methods. Potential resolutions to this issue have varied from changes to the implementation of viscosity (Rice et al., 2014; Klee et al., 2017), to varied cooling methods (Rice et al., 2012; Baehr & Klahr, 2015), but one likely reason for the problem lies in the smoothing of the self-gravity potential (Young & Clarke, 2015). Without smoothing self-gravity to approximately the scale height of the disk (Huré & Pierens, 2009; Müller et al., 2012), self-gravity on small scales is exaggerated in 2D, the effect being more pronounced with increasing resolution, resulting in easier runaway collapse with each increase in resolution. Thus, 3D models are needed to clarify the critical cooling boundary for disk fragmentation including what value the boundary takes and whether or not it converges in 3D with a simple cooling law.

The origin of the need for a smoothing length in two-dimensional simulations can be seen in the form of the self-gravity potential in the case of a razor-thin disk and disk of finite thickness. In the previous chapter, the equations of the self-gravitational potential had no vertical component, but now include the vertical displacement from the midplane

$$\phi = -\iiint \frac{G\rho}{\sqrt{s^2 + (z - z')^2}} dA dz' . \tag{3.1}$$



t

Figure 3.1: The analytic solution to Equation (3.3) which gives the density stratification in a self-gravitating disk for $Q_0 = 1$, (blue) and the Gaussian profile (red) which is used as a simple approximation to the solution.

This renders the need for a gravitational smoothing length moot, as the disk material is no longer condensed to an infinitesimally thin layer, but spread over the vertical extent. This prevents overdensities from growing into singularities which lead to the resolution dependent behavior of the previous chapter.

Global 3D studies of disk fragmentation are not uncommon (Mejía et al., 2005; Boley et al., 2006; Mayer et al., 2007; Michael et al., 2012; Lichtenberg & Schleicher, 2015) but because of the disk scale, cannot focus on the proper microphysics of the clump/fragment, instead focusing on the disk and its interaction with embedded objects. In contrast, *local* 3D simulations are rarer and do not focus on the formation of fragments, but on properties of the gravitoturbulent state, such as structure (Mamatsashvili & Rice, 2010; Shi & Chiang, 2014) and dust concentration (Shi & Chiang, 2013). Shi & Chiang (2014) came the closest to establishing a link to the critical cooling threshold of Gammie (2001) using 3D local simulations, but stopped short of modeling clump collapse or conducting simulations of the fragmentation threshold over multiple resolutions.

A proper implementation of a 3D shearing box simulation requires vertical stratification of the gas density. If this is not correctly initialized, the simulation will adjust on its own, causing undesired rapid heating and cooling as the disk swells up or contracts. Thus we assume the disk begins in hydrostatic equilibrium, and including self-gravity, the equation reads

$$\frac{1}{\rho}\frac{dP}{dz} = -\Omega^2 z - 4\pi G \int_0^z \rho(z')dz'. \tag{3.2}$$

The initial vertical density distribution is determined by rewriting the above as a dimen-

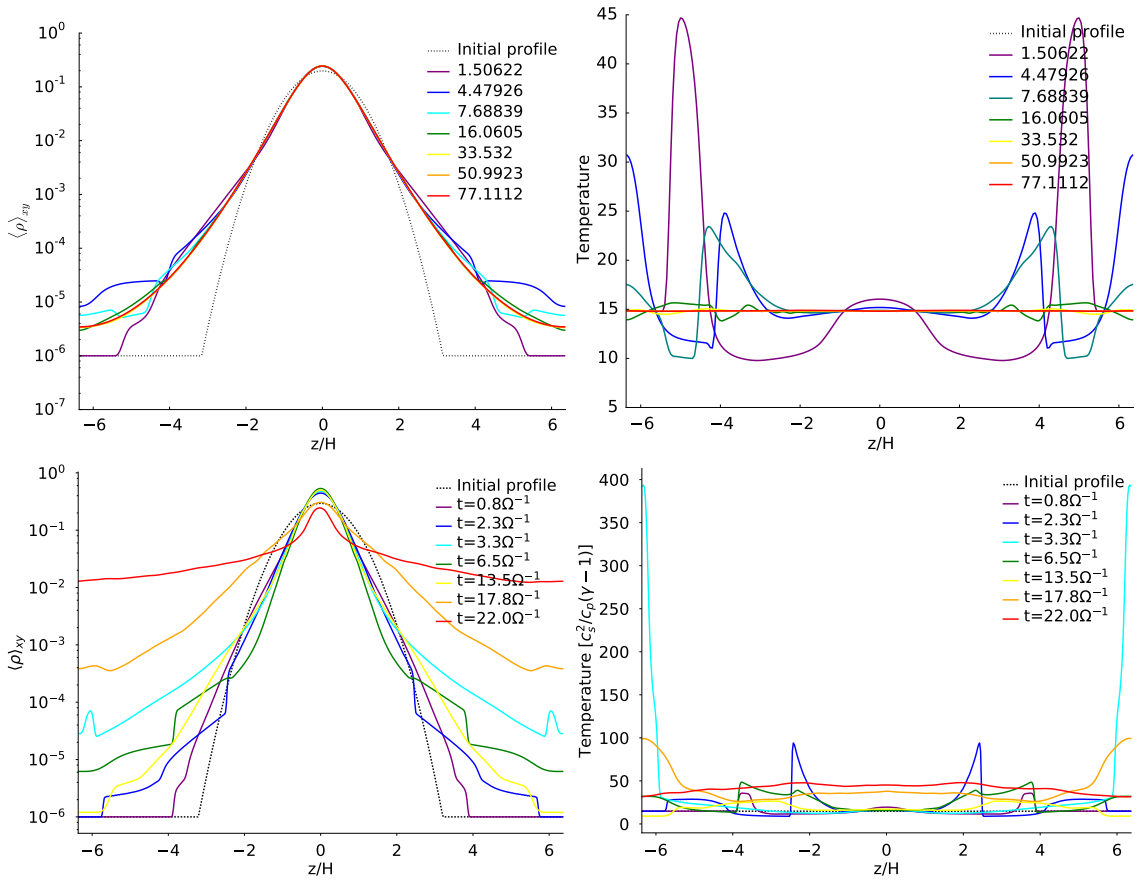


Figure 3.2: Top row: Average vertical density (*left*) and temperature (*right*) structure of a disk at various times (colored lines), initialized with $Q_0 = 1$. The black dotted line is the initial distribution which, in the case of the density, is as shown in Figure 3.1. As expected for a stable configuration, after minor heating due to the simulation adjusting to equilibrium, the temperature profile returns to its original flat vertical profile and the density settles to a stable. Bottom row: Same as top row, but for a simulation which is Toomre unstable. In the density structure dense object is clearly present at the midplane while the temperature shows the dissipation of some strong shocks which raise the overall temperature beyond the initial constant profile. Colored lines indicate different times expressed in units of Ω^{-1}.

sionless, differential equation, following Shi & Chiang (2014)

$$\frac{d^2\tilde{\rho}}{d\tilde{z}^2} + \frac{\gamma-2}{\tilde{\rho}}\left(\frac{d\tilde{\rho}}{d\tilde{z}}\right)^2 + Q_0^2\tilde{\rho}^{2-\gamma} + \frac{2}{h}\tilde{\rho}^{3-\gamma} = 0. \tag{3.3}$$

I use an isothermal approximation to the polytropic equation of state for which Equation (3.3) was derived ($\gamma = 1, K = c_s^2$), which means solving the equation for vertical hydrostatic equilibrium including self-gravity

$$\frac{d^2\tilde{\rho}}{d\tilde{z}^2} - \frac{1}{\tilde{\rho}}\left(\frac{d\tilde{\rho}}{d\tilde{z}}\right)^2 + Q_0^2\tilde{\rho} + \frac{2}{h}\tilde{\rho}^2 = 0, \tag{3.4}$$

where $\tilde{\rho} = \rho/\rho_0$, $\tilde{z} = z/(c_{s,0}^2/\pi G\Sigma_0)$, $Q_0 = c_{s,0}\Omega/\pi G\Sigma_0$ and $h = H/(c_{s,0}^2/\pi G\Sigma_0)$. The solution to the vertical density distribution for this equation is shown as the blue in line in Figure 3.1 along with a Gaussian distribution which is used to approximate the solution.

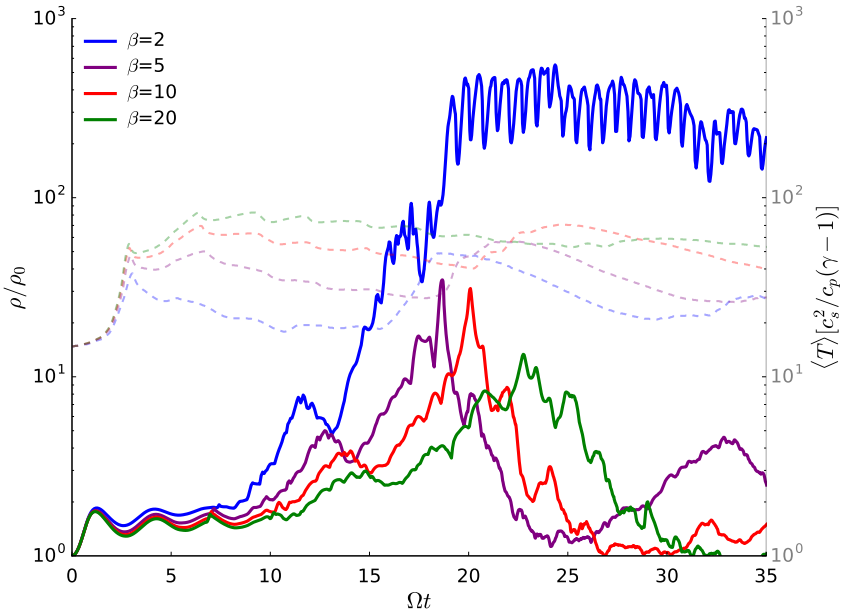


Figure 3.3: Evolution of the maximum density (solid colored lines) for simulations with varying cooling timescales at medium resolution ($N^3 = 512^3$). As expected for a critical cooling criterion of $\beta_{\rm crit} = 3$, simulations with cooling times greater than $\beta_{\rm crit}$ are stable against fragmentation, while the one with a shorter cooling time collapses. Also plotted in in semi-transparent dashed lines are the box-averaged temperatures.

Effect on Stability

The traditional Toomre parameter assumes a thin disk and vertically integrated or averaged quantities, so when determining stability in a vertically stratified disk, modifications to the Toomre parameter are necessary. The simple way is to replace the surface density with the midplane volume density approximated over the scale height of the disk $\Sigma \approx \rho H$ (Mamatsashvili & Rice, 2010)

$$Q_{\rm 3D} = \frac{\Omega^2}{\pi G \rho}. \tag{3.5}$$

However the critical stability value also changes, since the gas mass is no longer concentrated in a razor-thin 2D slice, but diluted over the entire vertical extent. Numerous formulations of a three-dimensional critical Toomre parameter, going back to one of the original disk instability papers Goldreich & Lynden-Bell (1965) and continuing throughout the decades (Vandervoort, 1970; Wang et al., 2010; Behrendt et al., 2015). No one has conclusively determined a 3D Toomre stability parameter, likely because it is a poorly posed problem with too many free variables to for which to account. Regardless, most derived values lie between $0.6 < Q_{\rm 3D,crit} < 0.8$, which will be shown later is in good agreement with the critical value in my simulations. Studies of similar 3D physics at galactic scales were conducted by Kim et al. (2002), which found a critical Toomre parameter $Q \approx 0.7$ but did not include thermodynamics or investigate the effect of resolution on the fragmentation behavior.

In my case, the value of the unstable Toomre value is affected by the comparatively high choice of the initial sound speed, which yields a disk aspect ratio $H/R \sim 0.1$, which is fairly high for typically cool self-gravitating disks. This means that more mass is required to concentrate locally to overcome the increased thermal pressure, resulting in a lower

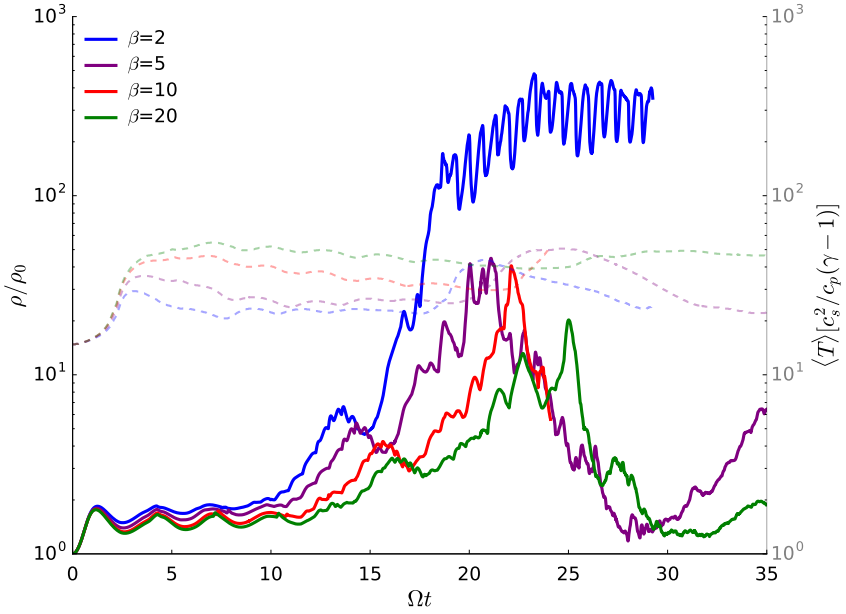


Figure 3.4: Same as Figure 3.3, but for the high resolution ($N^3 = 1024^3$) simulations. As in the case of the medium resolution simulations, fragmentation occurs according to the Gammie cooling criterion and shows no signs of divergence with the increase in resolution. Also plotted in in semi-transparent dashed lines are the box-averaged temperatures.

$Q(t = 0)$. Regardless, this does not change the ability of the simulations to fragment and only leads to slightly larger fragment masses.

The Gaussian approximation to the hydrostatic equilibrium shown in Figure 3.1 will settle to the true equilibrium shortly after the simulation begins and mass will shift around, evident in the top left panel of Figure 3.2. There is a slight increase of mass at the disk midplane, comparable deficits between $1H < |z| < 2H$ and additional surpluses beyond $|z| > 2H$. The amount of additional mass in these upper and lower disk layers is on the order $\mathcal{O}(10^{-3})$ and result in some minor shocks and heat generation, which is quickly dissipated, visible in the adjacent plot (top right) of Figure 3.2.

Consequences

The initial temperature distribution is not stratified but instead isothermal, equal to the background irradiation level, as distant regions of the disk will be dominated by external radiation from the star or other nearby sources, rather than accretion heating. This irradiation hits the surface of the outer disk and being warmer than the midplane, the top and bottom irradiated layers gradually heat up the interior layers until the entire vertical extent is the same as the temperature set by the irradiation. As can been seen in Figure 3.2, the onset of self-gravitating shocks generate significant heat, but fast enough cooling allows the vertical temperature profile to settle to its original flat distribution.

The vast difference in densities from the midplane to the vertical boundaries is a problem for the numerical integration by the code, causing high turbulent velocities and short timesteps which cannot be computed in a reasonable amount of time. Thus I introduced a density floor, which can be seen in the density profiles of Figure 3.2, which had to be carefully calibrated to avoid undesired mass loss. The settling of material onto the midplane layers not only removes mass from the upper layers as a whole, but individual cells also drop below the threshold and are filled at the end of the integration. This

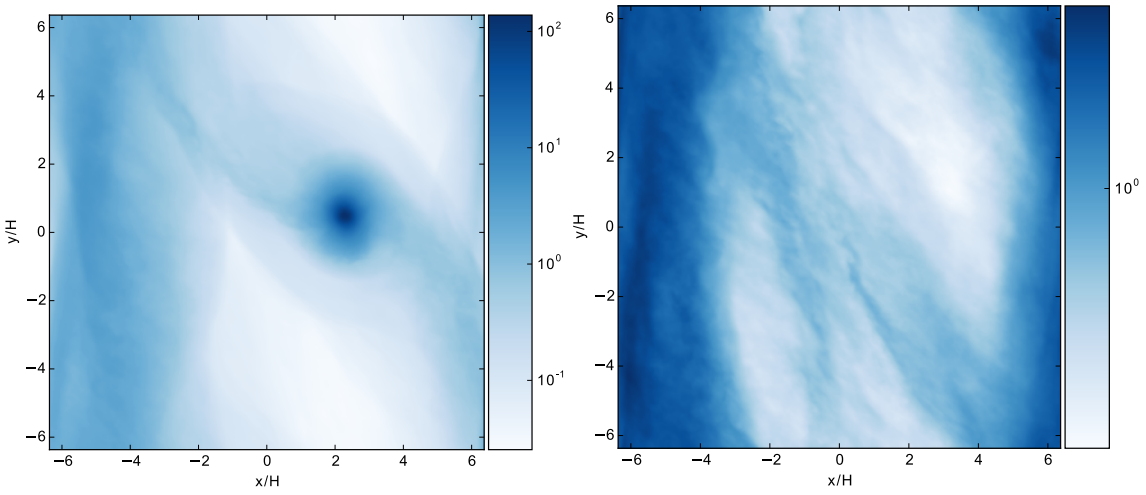


Figure 3.5: Left: Surface density map of a fragment formed in a simulation with a cubic domain and 512^3 grid cells with cooling $\beta = 2 < \beta_{\rm crit}$. Right: the gravitoturbulent case in the same domain when $\beta = 5 > \beta_{\rm crit}$.

creates mass out of nowhere, which piles up and affects the overall stability and total mass in the simulation. A floor value of $\rho_{\rm floor} = 10^{-6}$ was found to be sufficient such that over the duration of the simulation, mass gains were negligible and the timestep was not prohibitively small.

3.2 Three-Dimensional Models

For this chapter and the next, I conduct 3D hydrodynamic shearing box simulations of a self-gravitating disk with an Eulerian mesh. Local simulations simulations allow the Toomre wavelength $\sim H$ to be well-resolved and keep the boundary conditions periodic. Shearing box simulations use hydrodynamic equations which are linearized and transformed into co-rotating Cartesian coordinates, where $q = \ln\Omega/\ln R = 3/2$ is the shear parameter. The 3D equations are written out explicitly as

$$\frac{\partial \rho_{\rm g}}{\partial t} - q\Omega x \frac{\partial \rho_{\rm g}}{\partial y} + \nabla \cdot (\rho_{\rm g} \boldsymbol{u}) = f_D(\rho_{\rm g}) \tag{3.6}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} - q\Omega x \frac{\partial \boldsymbol{u}}{\partial y} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} = -\frac{\nabla P}{\rho_{\rm g}} + q\Omega v_x \hat{\boldsymbol{y}} - 2\Omega \times \boldsymbol{u} - \nabla\phi - \frac{\epsilon}{\tau_f}(\boldsymbol{u} - \boldsymbol{w}) + f_\nu(\boldsymbol{u}) \tag{3.7}$$

$$\frac{\partial s}{\partial t} - q\Omega x \frac{\partial s}{\partial y} + (\boldsymbol{u} \cdot \nabla) s = \frac{1}{\rho_{\rm g} T}\left(2\rho_{\rm g}\nu \mathbf{S}^2 - \Lambda + f_\chi(s)\right). \tag{3.8}$$

In equations (3.6) - (3.8), $\mathbf{u} = (v_{\rm x}, v_{\rm y} + q\Omega x, v_{\rm z})^T$ is the gas flow plus shear velocity in the local box, $\boldsymbol{w}$ is the particle velocity, which imparts a backreaction on to the gas proportional to the local dust-to-gas ratio ϵ, $\rho_{\rm g}$ is the gas density, and the thermodynamic variable is the gas entropy s. For now, no particles are added. Viscous heat is generated by $\mathcal{H} = 2\rho_{\rm g}\nu\mathbf{S}^2$, with rate-of-strain tensor $\mathbf{S}$. Heat is lost using a simple β-cooling prescription

$$\Lambda = \frac{\rho(c_{\rm s}^2 - c_{\rm s,irr}^2)}{(\gamma - 1)t_{\rm c}} \tag{3.9}$$

with t_c given by $t_c = \beta\Omega^{-1}$ and background irradiation term $c^2_{s,irr}$, which ensures fragmentation is a result of growing mass perturbations rather than a transient region of effectively no thermal pressure support.

Self-gravity is solved in Fourier space by transforming the density into Fourier space to find the potential at wavenumber k and transforming the solution back into real space. The solution to the Poisson equation in Fourier space at wavenumber k is

$$\phi(\boldsymbol{k}, t) = -\frac{2\pi G\rho(\boldsymbol{k}, t)}{\boldsymbol{k}^2}, \tag{3.10}$$

where ϕ is the gravitational potential of the gas.

Finally, I use an ideal equation of state, with internal energy ε, and a constant specific heat ratio $\gamma = 5/3$

$$P = (\gamma - 1)\rho\varepsilon. \tag{3.11}$$

3.3 Fragmentation

Preliminary simulations of 3D self-gravitating setup were conducted at a resolution of $N^3 = 256^3$, which is the same effective resolution as the 2D simulations with $N = 1024^2$ of Baehr & Klahr (2015). When initialized with $Q_0 = 1$, these simulations show no gravitational instability, remaining static for tens to hundreds of orbital timescales. As can be seen in the top two panels of Figure 3.2, the density and temperature structure remains largely undisturbed, suggesting that self-gravity is not strong enough to excite unstable modes. This was tested at a higher resolution $N^3 = 512^3$ and found that $Q_0 = 1$ still shows no evidence for gravitational instabilities, suggesting that this is not a resolution dependent effect. For this reason I use a new value of $Q_0 = 0.676$, a value close to what is typically assumed for disks of finite thickness (Wang et al., 2010).

Accordingly, simulations with this new Q_0 at three resolutions and at 3 cooling timescales, summarized in Table 3.1. The 3 different cooling timescales are selected assuming a critical fragmentation value of $\beta = 3$. For $\beta = 2$, one expects fragmentation in the typical scenario of 2D disk fragmentation (Gammie, 2001) and for $\beta = 5, 10$, one expect a stable gravitoturbulent disk, but should also be aware of the possibility of non-convergent effects, where fragmentation is possible for longer cooling times.

Fragmentation was observed with a critical cooling timescale of between $\beta = 2$ and $\beta = 5$ for the higher resolution simulations, but not in the case of $N = 256$, where efficient cooling did not result in fragmentation. Lack of fragmentation at this resolution for short cooling times is also observed in identical 2D simulations suggesting simulations at this domain and resolution are poorly resolved rather than an issue with the setup of the simulation (See Figure 3.12.

These results coincide with the calculations of α stresses, which show $Q_0 = 1$ disks generate weak turbulence (for a self-gravitating disk) at or below $\alpha \sim 0.02$. With more mass and stronger self-gravity, turbulence becomes significantly stronger as in Figure 3.6, with fragmentation coinciding with significant spikes of $\alpha > 1$.

Figures 3.3 and 3.4 show the maximum density and average temperature for the $N^3 = 512^3$ and $N^3 = 1024^3$ simulations, respectively, and demonstrate the anticipated fragmentation behavior of my 3D self-gravitating setups. At the medium resolution $N^3 = 512^3$ fragmentation occurs in the case where the cooling timescale is short enough to remove pressure support and the clump collapses before it can be torn apart by tidal shear. For longer cooling timescales, the fragment cannot collapse fast enough to remain bound and

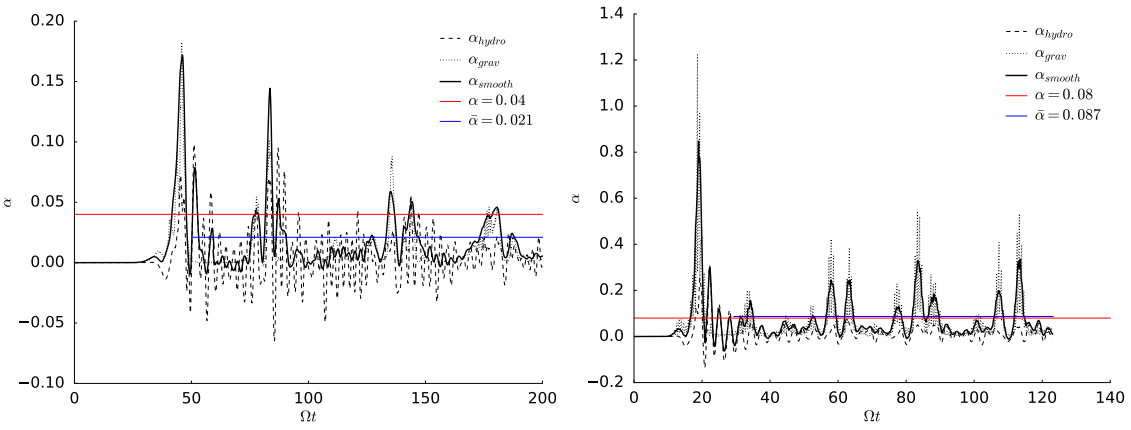


Figure 3.6: The gravitational and hydrodynamics components of α for the simulation labeled in Table 3.1 as *G256t10* (left) and *G512t5* (right). In the gravitoturbulent steady state the stress is dominated by the hydrodynamic component and significant deviations are a results of fragmentation and the dominance of gravitational stresses. The red line indicates the analytic expectation from Equation (3.13). The blue represents the average value of the total stress over the span of the line which begins when the simulation has settled.

is soon disrupted; the difference between the two states is illustrated in the two panels of Figure 3.5. Also shown in Figures 3.3 and 3.4 are the box-averaged temperatures over time, showing that the fragmenting simulations are able to withstand the rising temperatures as the fragment collapses and becomes denser due to the efficient cooling.

Using a this lower value of $Q_0 = 0.676$ immediately resulted in the unstable behavior expected from local disk instability simulations. THe bottom panels of Figure 3.2 show the average vertical density and temperature profiles over time in a simulation with $N = 1024$ resolution. I was able to recover the $\beta_{\rm crit} = 3$ condition for the base simulation resolution $N^3 = 512^3$. From the lower left panel of Figure 3.2 one can see that the vertical density profile makes a slight adjustment to reach hydrostatic equilibrium and mass concentrates in the midplane with time as non-linear gravitational instabilities grow and the fragment collapses. The adjustment to hydrostatic equilibrium results in non-negligible heat generation in the layers surrounding the midplane. Once a simulation settles, the thermodynamics are dominated by the dissipation of shocks, evident in the regions just above and below the midplane of the temperature profiles of Figure 3.2.

These medium resolution simulations were then mirrored to higher resolutions, only making slight adjustments to the numerical dissipation terms in the process. While simulations with shorter cooling timescales show a slightly greater propensity to fragment, the fragmentation behavior varied much greater with resolution. At the highest resolution ($N^3 = 1024^3$) fragmentation is quick and clear for short cooling timescales, where fragmentation is expected but also for long cooling timescales, where thermal pressure support is expected to suppress disk fragmentation.

At lower resolutions ($N = 256^3$), fragmentation appears stochastic, with fragments occasionally forming, but often promptly torn apart. THis suggests that a resolution limit has been encountered and $256/12.7 \approx 20\,H$ grid cells per scale height are insufficient to resolve fragmentation. This is 5 times more resolution than necessary to avoid spurious fragmentation (Truelove et al., 1997; Nelson, 2006), but not as high as the resolution requirements to resolve the magnetic dynamos which can impede the collapse of stellar cores (Turk et al., 2012).

3.4 Turbulence

An important disk parameter to be measured by the shearing sheet model is the α-stress, first posited by Shakura & Sunyaev (1973). This formalism relies again on the assumptions that the disk is thin and that angular momentum is transported locally through a dimensionally defined viscosity (Pringle, 1981)

$$\nu = \alpha c_s H. \tag{3.12}$$

This makes α a useful value with which to compare analytic expressions of a disk to the turbulence that arises in simulations. A derivation of the standard two-dimensional α-stress can be found in Appendix Section A.3. For now, the value in these three-dimensional simulations should be scrutinized against not just the 2D analytic expression, but other studies on gravitoturbulence as well, to ensure energy is well conserved by the code.

The analytic expression for α viscosity in a gravitoturbulent disk, given by Gammie (2001), is a constant dependent on the cooling time t_c, orbital frequency Ω and adiabatic index γ

$$\alpha = \frac{4}{9}\frac{1}{\gamma(\gamma-1)t_c\Omega}. \tag{3.13}$$

This gives a prediction for the stress in a disk in terms of the cooling timescale of a gravitoturbulent viscously heated disk. A relation between α and the cooling time also means that one can also define the fragmentation boundary in terms of an limit in the stress that a disk can sustain, around $\alpha \simeq 0.1$ (Rice et al., 2005). This relation can be seen in Figure 4 of (Paardekooper, 2012) which uses 2D shearing sheet simulations with simple cooling. They find an agreement to within 5% between data and the model as long as the disk does not fragment.

The study on gravitoturbulence by Shi & Chiang (2014) used a density-weighted sound speed to determine the strength of the α parameter, so that velocity perturbations in the denser midplane layer factor into the value more than more less dense regions away from the midplane which may have high mach numbers which do not represent the gravitational or Reynold's stresses of interest. Due to this, they found that the three dimensional α stresses were more appropriately represented by a value

$$\alpha_{3D} = \frac{3\gamma}{2}\alpha_{2D}, \tag{3.14}$$

for adiabatic index γ. The stresses in the 3D simulations of this thesis are calculated according to Equation (A.28) and shown in Figure 3.6. As Shi & Chiang (2014) also found, these stresses indicate that the conventional 2D value of α was still on average, a closer fit. More critical was the bursty, volatile behavior of the viscous stresses, as shown in Figure 3.6. While the average value was roughly comparable to the analytic expression, wild swings of particularly strong stresses balance out longer periods of weak stress. This was studied further in (Booth & Clarke, 2019), showing that small box sizes lack the full range of unstable modes which leads to particular modes dominating the stress.

Threshold

Gravitational instability does not necessarily result in fragmentation; if $1 \leq Q \leq 1.4$ the disk will be unstable to nonaxisymmetric perturbations, or spiral arms, that will transport angular momentum outwards, but not collapse into fragments (Bodenheimer & Lin, 2002; Armitage, 2011). Additionally, clumps may reach an overdensity only to fall apart and

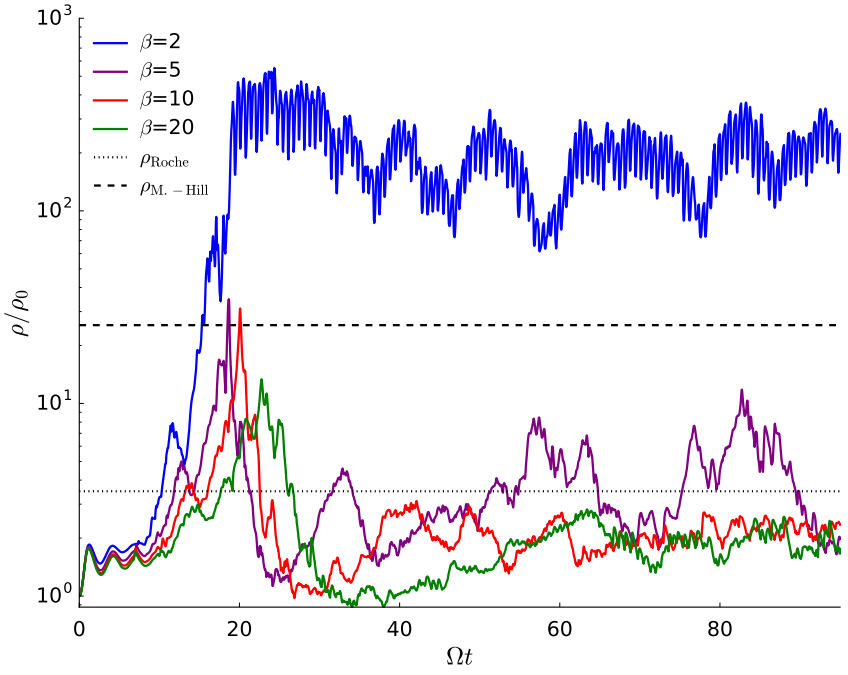


Figure 3.7: The medium resolution simulations in Figure 3.3, but displayed for longer times and with the two fragmentation thresholds. The dotted line is the Roche density from Baehr & Klahr (2015), which worked for 2D simulations, but clearly not for the 3D ones presented here. The pressure-modified Hill threshold (dashed line) is more appropriate at separating between fragmentation and non-fragmentation (Baehr et al., 2017).

the disk returns to a gravitoturbulent state. Therefore, fragments need to both be able to form and survive shear forces (Kratter et al., 2010). A fragment has survived when it has reached Roche density, defined by the stability of a rotating Darwin ellipsoid (Darwin, 1906; Chandrasekhar, 1964)

$$\rho_{\mathrm{Roche}} = 3.48\frac{M_*}{R^3}, \tag{3.15}$$

where M_* is the mass of the central star and R is the radial distance between the fragment and star. A Darwin object is one that as a fluid, such as the gaseous clumps formed here, deforms into an ellipsoid under the tidal forces of its host star.

Since all presented simulations are local and take no consideration of any absolute central mass besides a vertical gravitational potential consistent with an object, one can use the Keplerian frequency $\Omega = \sqrt{GM/R^3}$ and the sound speed $c_s = H\Omega$ leads to an expression of the surface density in terms of simulation scale and dimensionless quantities

$$\Sigma_{\mathrm{Roche}} = 3.48\frac{c_s^2}{HG}, \tag{3.16}$$

where $\Sigma_{\mathrm{Roche}} = \rho_{\mathrm{Roche}}H$. Here fragments have formed when the clump density is greater than this Roche surface density for more than a few cooling timescales t_c.

In the 2D case, fragmentation can be easily established with a Roche surface density threshold (See section 2.4 of Baehr & Klahr (2015)), but this threshold does not translate to 3D simulations and a volume density threshold, as seen in Figure 3.7. Instead it is the pressure-modified Hill condition of Kratter & Murray-Clay (2011) which results in a suitable threshold density. This threshold considers that pressure support stabilizes a

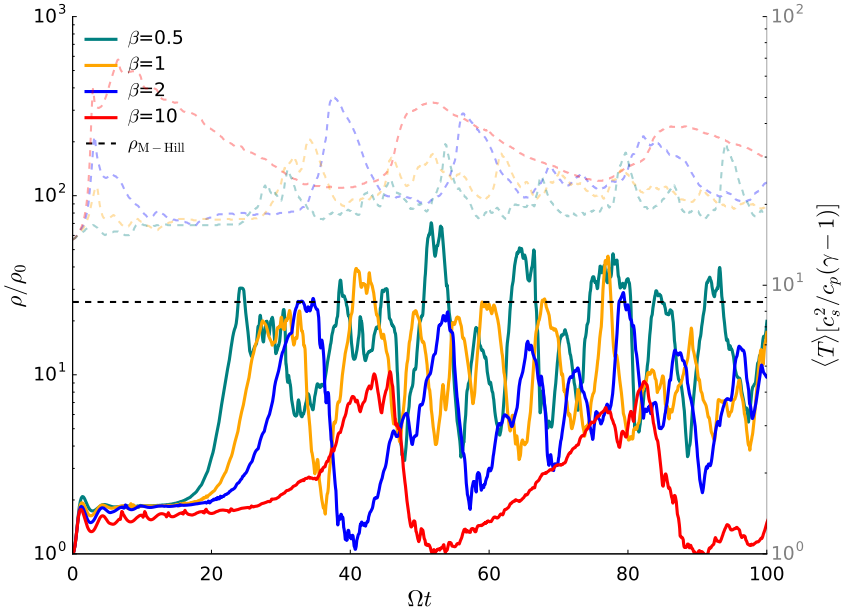


Figure 3.8: The maximum surface densities of four simulations carried out at resolution $N = 256^3$ and $Q_0 = 0.676$. Fragmentation outcomes are based on the cooling timescale, consistent with a $\beta_{\rm crit} \sim 3$. Dashed lines are the respective fragment density thresholds calculated according to Equation (3.20).

fragment against collapse and leaves it more susceptible to tidal shear. Starting with the condition that a fragment has a radius that is a fraction of it's Hill radius

$$r_{\rm frag} = \left(\frac{1}{f}\right)^{1/3} \left(\frac{M_{\rm frag}}{3M_*}\right)^{1/3} R, \tag{3.17}$$

where M_* is the mass of the central star, R is the location away from the center and f is factor determined by the choice of specific heat ratio γ. Defining the fragment mass as the integration of density over radial shells

$$M_{\rm frag} = 4\pi \int_{h_{50}}^{r_{\rm frag}} \rho_{\rm h} \left(\frac{r_{\rm h}}{r}\right)^k r^2 dr, \tag{3.18}$$

with the density at the center of the fragment $\rho_{\rm h}$ at a size one fiftieth the scale height $r_{\rm h}$, a fragment outer radius of $r_{\rm frag}$ and a power law index of the fragment density distribution k. Then inserting the fragment mass into Equation (3.17) and assuming $r_{\rm h} \ll r_{\rm frag}$, one then solves for the central density $\rho_{\rm h}$

$$\rho_{\rm h} = \frac{3f(3-k)M_*}{4\pi R^3} \left(\frac{r_{\rm frag}}{r_{\rm h}}\right)^k. \tag{3.19}$$

Assuming $\gamma = 5/3$ makes $f = 4$, whereas a softer choice of $\gamma = 7/5$ would mean $f = 16$, resulting in a slightly higher central density. Equation (3.19) will serve as my threshold density for these simulations, provided that the maximum density is above this threshold for several timescales, thus separating fragments from overdensities which have not yet collapsed enough to survive.

Fragmentation Criterion

Finally, I apply the fragment threshold, Equation (3.19), to the simulations, from which one needs to extract the density profile of a fragment. In this case, a fragment scales with $k = 0.7$ to a fragment radius of $r_{\rm frag} = 2/3H$. With this, I find that for an overdensity to withstand tidal disruption, as is shown in Figure (3.7), it must satisfy

$$\rho_{\rm h} \geq 25.57 \frac{M_*}{R^3}, \tag{3.20}$$

which, for a protoplanetary disk around a solar-mass star, results in a densities of 1.2×10^{-10}g/cm^3 at 50 au and 1.5×10^{-11}g/cm^3 at 100 au. Comparing this to the simulations and to the Roche density in Figure 3.7 one sees that this density threshold more consistently delineates between fragments and non-fragments.

Applying the density threshold to the low resolution simulations which failed to form sustained fragments, as one can see in Figure 3.8, simulations with very short cooling times ($\beta = 1, 0.5$) surpass this threshold. If the maximum density of a fragment can surpass this critical density for more than an orbit, it can be considered a fragment, even if it is subsequently disrupted. Thus, this threshold may provide a useful sink cell formation criterion for simulations with resolution too low to fragment.

Mass Estimation

To estimate the mass to which this threshold corresponds, one needs to specify a radial location in the disk for the previously scale-free simulations to inhabit. Using the mass of a fragment as defined by Equation (3.18) and assuming that the fragment radius is much larger than the length of a single grid cell ($r_{\rm frag} \gg r_{\rm h}$) and is spherically uniform and symmetric

$$M_{\rm frag} \approx \frac{4\pi\rho_{\rm h} r_{\rm h}^k}{3-k}\left[r_{\rm frag}^{3-k} - \cancel{r_{\rm h}^{3-k}}\right] = \frac{4\pi\rho_{\rm h} r_{\rm frag}^3}{3-k}\left(\frac{r_{\rm h}}{r_{\rm frag}}\right)^k. \tag{3.21}$$

A disk with the temperatures as in these simulations will have an aspect ratio $H/R \approx 0.11$ at 50 au and $H/R \approx 0.16$ at 100 au. With these disk parameters specified one can calculate the densities quoted above which are then inserted into Equation 3.21 with $r_{\rm frag} = 2/3H$, $r_{\rm h} = H/50$ and $k = 0.7$, I approximate minimum fragment masses of $5M_{\rm J}$ and $15M_{\rm J}$ at 50 and 100 au, respectively. These minimum mass estimates are consistent with the observations of directly observed gas giants and brown dwarfs (Rameau et al., 2013; Bonnefoy et al., 2014, 2018; Maire et al., 2018; Cheetham et al., 2018), suggesting there are some observed planets and brown dwarfs which could possibly have been formed by disk fragmentation, but much more information about these objects would need to be known before making any determination.

3.5 Parametric Instabilities

In showing that convergence requires a smoothing length on the order of H, Young & Clarke (2015) noted that fragmentation does not depend strongly on collapse at lengths smaller than this scale (see also Gammie (2001); Shi & Chiang (2014)), otherwise fragmentation would be affected by the introduction of a smoothing length. For this reason, they argued it is unnecessary to model fragmentation with 3D models as one could get improved resolution with a similar 2D simulation without sacrificing any of the necessary unstable

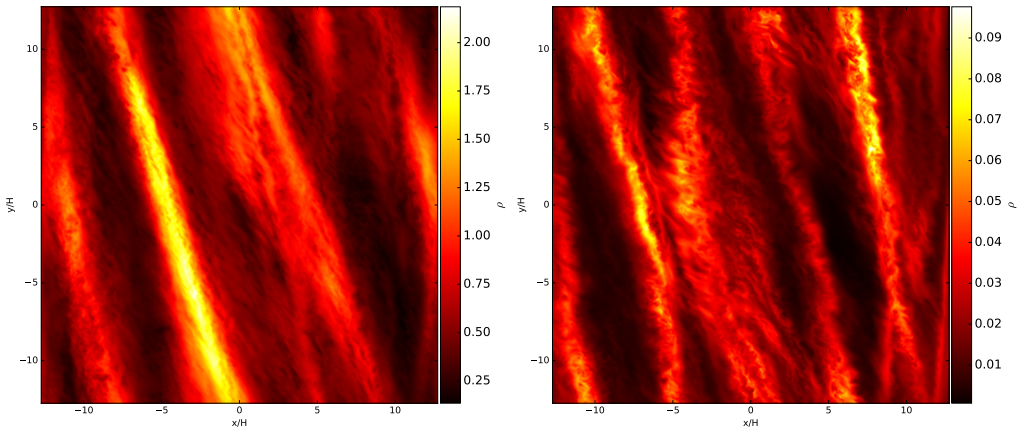


Figure 3.9: Top left: Slice of the midplane density of a simulation with gravitoturbulence, evident from the large non-axisymmetric features. Top right: Density at one scale height above the midplane from the same simulation and time as left. Color scale and linear density scale used for direct comparison to Figure 3 of Riols & Latter (2017).

modes for fragmentation. Others have seen sparse small scale structure when studying gravitoturbulence (Shi & Chiang, 2014; Hirose & Shi, 2017), although the relatively low resolutions may have limited the visibility of small scale structure.

However, (Riols & Latter, 2017) found evidence of a small scale parametric instability (Goodman, 1993) in 3D gravitoturbulent simulations which could necessitate the future use of 3D simulations to understand the effect of this instability on fragmentation, particle sedimentation, and gravitoturbulence. Parametric instabilities occur when inertial waves within the disk couple to a large scale axisymmetric oscillation, which can grow into an instability when (Fromang & Papaloizou, 2007; Riols & Latter, 2017)

$$\omega_F = \omega_{i,1} + \omega_{i,2} \tag{3.22}$$

where ω_F is the frequency of the axisymmetric oscillation and $\omega_{i,1}$ and $\omega_{i,2}$ are the frequencies of the pair of inertial waves. The axisymmtric oscillation in a self-gravitating disk is a consequence of the unstable $m = 0$ mode when $Q < 1$, which is not necessarily excited in a gravitoturbulent simulation, but can be locally excited provided the density is high enough. Inertial waves of varying frequencies develop from the rotation of the disk, restored by the Coriolis force. Riols & Latter (2017) find axisymmetric modes with $\omega_F \approx 1.6\Omega$ and inertial modes at $\omega_i = \omega_F/2$ and $\omega_i = \omega_F/4$, supporting the idea of inertial mode couplings driving the parametric instability.

The simulations presented here offer a very close comparison of results of Riols & Latter (2017) considering that I also study self-gravitating disks with 3D shearing boxes of a similar size. My attempts to recreate their results are shown in Figure 3.9. The top row recreates Figure 3 from Riols et al., showing that the midplane density structures appear to have fewer small scale structures and perturbations, compared to the regions at $z = 1H$ which appear more turbulent. This can also be seen in the plots of the bottom row, which show a line integral convolution of the velocity field superimposed on top of the density distribution. In these plots it is perhaps more evident that there is more laminar flow at the midplane layer, particularly within the non-axisymmetric density structures, in the plot on the left.

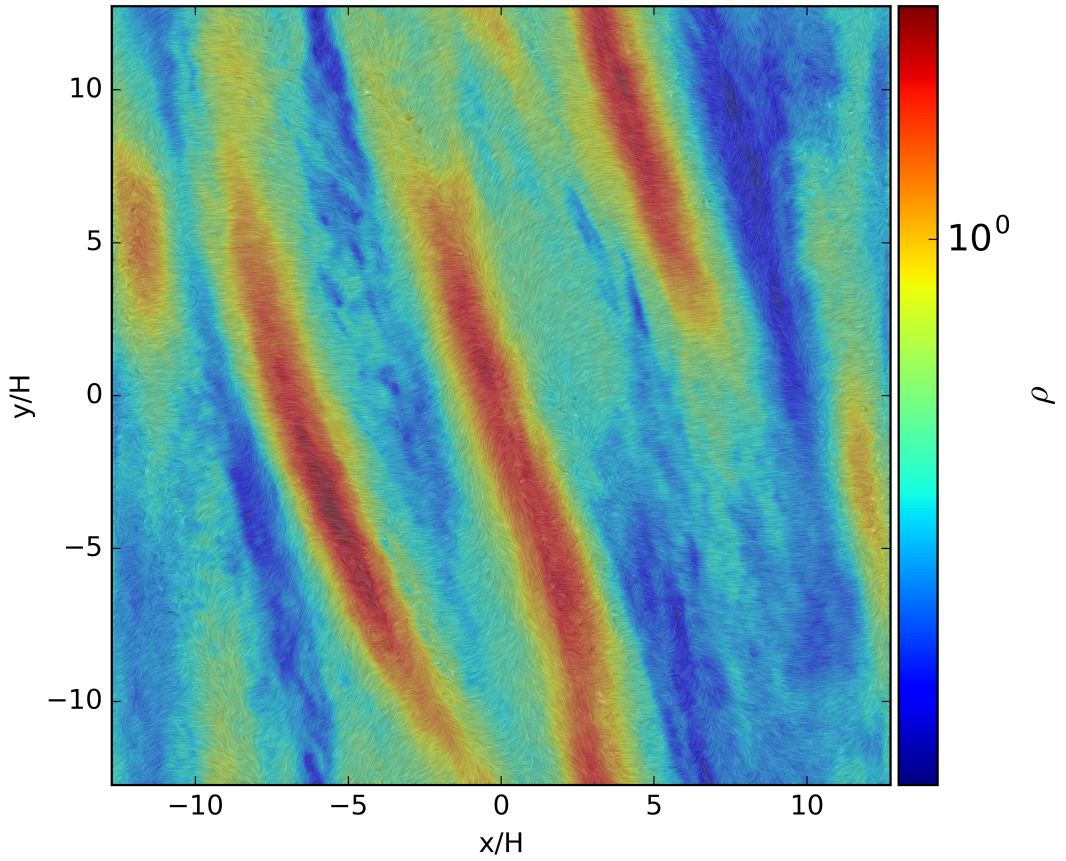


Figure 3.10: Midplane density with a line integral convolution visualization of the velocity field (Cabral & Leedom, 1993). The laminar flow in the midplane is evident particularly near the dense structures.

3.6 Discussion

Convergence

Following the results of Young & Clarke (2015), which demonstrated the need for a smoothing length to achieve convergent disk fragmentation in thin disk simulations, one would expect 3D simulations to converge since a smoothing length is no longer required for a disk with finite thickness. There might be additional effects which affect fragmentation at different resolutions, including but not limited to cooling, viscosity and numerical effects, but investigating the extension of local self-gravitating disks to include a vertical component is the most natural extension of that work and these results support their findings in this regard.

Recent work has focused on numerical effects and their influence on non-convergence. Klee et al. (2017) found that numerical oversteepening could be responsible for the overestimation of overdense peaks resulting in fragmentation. Through my use of sixth-order hyperdiffusion, a strong shock-capturing method, I do not expect oversteepening to be a significant factor. Deng et al. (2017) find that dissipation of angular momentum in SPH methods removes shear support from regions with high flow velocities and as numerical viscosities depend on resolution, fragmentation becomes easier with increasing resolution. This means that one must take care in how they prescribe a numerical viscosity to avoid

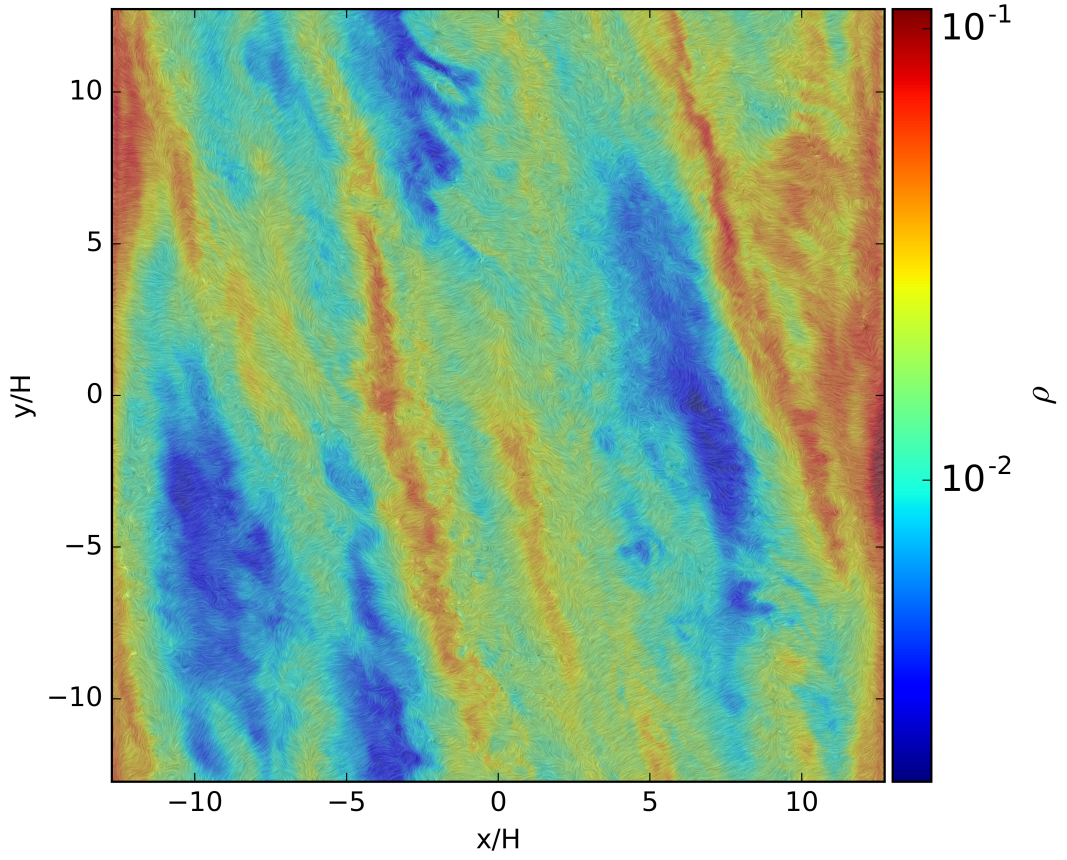


Figure 3.11: Same as Figure 3.10, but the velocity field is taken from one scale height above the midplane, showing significantly more chaotic flow even within overdense structures.

promoting fragmentation.

While these results strongly suggest convergence, at the lowest resolution fragmentation is suppressed due to insufficient resolution for both 2D and 3D models, similar to the results of Turk et al. (2012), which found that turbulent velocity fluctuations suppressed the formation of small scale magnetic dynamos when the Jeans length is not resolved by at least 64 grid cells. At higher resolutions the results follow the standard fragmentation scenario, with clear fragmentation at $\beta = 2$ and a lack of fragmentation evident at $\beta = 5, 10, 20$. The β cooling prescription is not a true substitute for realistic cooling, only a simple parametrization of the underlying physics, which makes for useful estimates that can be easily translated to aa disk fragmentation criterion. The use of a fixed background term may have an effect on convergence (Rice et al., 2012), thus one must be careful that convergence is not lost if this term is removed (Lin & Kratter, 2016). However, since simulations using β cooling and background irradiation converge, it supports our basic understanding of the physics.

Reconciling with the Thin Disk Approximation

A large part of this investigation has been about the differences between 2D and 3D simulations of self-gravitating disks, but it is also necessary to discuss how they are similar. Vertical structure and the accompanying gravitational dilution mean disk instability and

Table 3.1: 3D Simulations at various resolutions N^3, initial Toomre criteria Q_0, and cooling times β.

Simulation	Grid Cells (N^3)	Q_0	β	Fragmentation
Q256t2	256^3	1	2	No
Q256t10	256^3	1	10	No
Q512t2	512^3	1	2	No
Q512t10	512^3	1	10	No
G256tp5	256^3	0.676	0.5	No
G256t1	256^3	0.676	1	No
G256t2	256^3	0.676	2	No
G256t10	256^3	0.676	10	No
G512t2	512^3	0.676	2	Yes
G512t5	512^3	0.676	5	No
G512t10	512^3	0.676	10	No
G512t20	512^3	0.676	20	No
G1024t2	1024^3	0.676	2	Yes
G1024t5	1024^3	0.676	5	No
G1024t10	1024^3	0.676	10	No
G1024t20	1024^3	0.676	20	No

fragmentation require even higher gas surface densities to collect locally before GI can set in. Once instabilities develop, the results of 2D simulations with smoothed self-gravity and 3D simulations show similar structures and features on similar scales (Figure (3.13)). Gravitoturbulent simulations in 3D show non-axisymmetric structures of the same scale as in 2D, but do not show as strong a contrast as 2D simulations due to the weaker self-gravity on scales smaller than H.

Ultimately, both 2D and 3D simulations converge to the same critical cooling timescale but 3D simulations take significantly more resources. While 3D simulations are convergent and can capture effects at small scales, there is growing evidence that scales smaller than $\sim H$ contribute little to the growth of gravitationally unstable modes (Young & Clarke, 2015; Riols & Latter, 2016). Thus, for the modeling of self-gravitating disks, 2D simulations may be sufficient so long as an appropriate smoothing length is used to account for the vertical scale of the disk.

Fragmentation and Planet Formation

Planet formation by gravitational instability is hindered by the large disk masses required. With the requisite Q_0 for fragmentation decreasing in these simulations, it is be even more difficult to form planets by GI by requiring either more mass or a colder disk for fragmentation. Larger relative disk sizes required for fragmentation also suggest that the resulting fragments will draw from a larger mass reservoir to overcome stabilizing effects and form larger objects. Thus, it is more likely fragments already have enough mass to be considered brown dwarfs or low-mass stars, without taking into account post formation accretion and growth.

The density threshold of Equation (3.20) and the subsequent fragment mass suggest objects formed in these simulations are formed with at least $4.9 M_J$ at both 50AU and 100AU around a solar-like star. Since companions will generally form larger than these

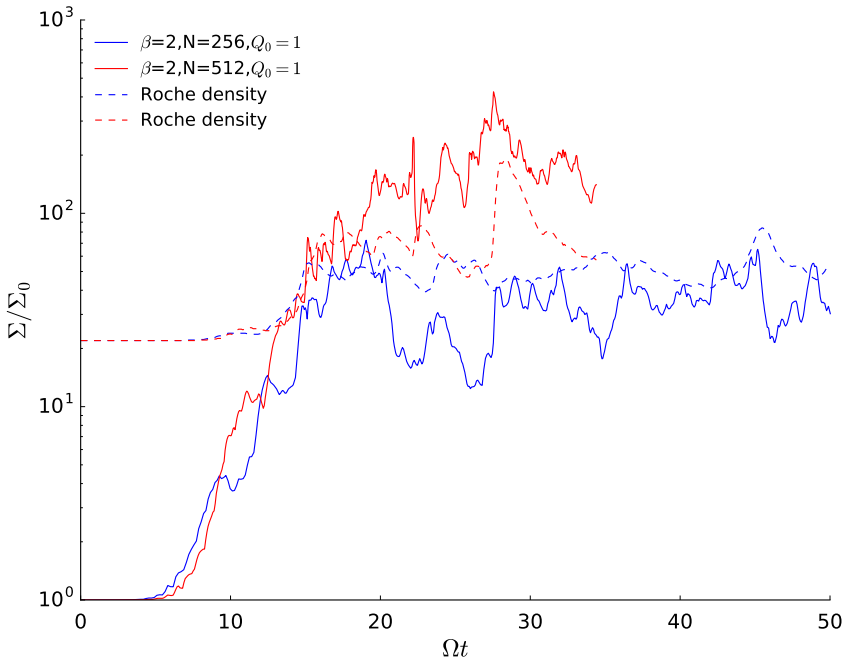


Figure 3.12: A pair of 2D simulations with the same resolution in x and y directions as the 3D simulations in Figures 3.3 and 3.4. These simulations show the same inability to fragment at $N = 256$ as the 3D simulations, suggesting there is not an issue with the inclusion of a vertical dimension, some strong dissipation affecting the fragmentation at lower resolutions.

thresholds and subsequently accrete more mass, these fragments will likely form planets many Jupiter masses in size and brown dwarfs as well. Recent observations show fragments could indeed be quite large, as is the case of L1448 IRS3B (Tobin et al., 2016a) which is consistent with theoretical expectations (Stamatellos & Whitworth, 2009).

Further Considerations

Global simulations can model the growth, migration, and potential destruction of fragments formed in a disk, but with these shearing box simulations one can resolve the small scales of GI structure and development. For a resolution and convergence study this makes local simulations ideal for the relatively small patches of the disk where fragmentation occurs.

The Fourier solver in Pencil is limited to domains with the same number of grid cells in the vertical direction as in the x-direction, meaning that a simulation with a flattened box would have significantly higher resolution in z, resulting in different values of hyperviscosity in each direction, possibly leading to unphysical effects. For that reason, I carried out my simulations spanning roughly 12.7 scale heights in the vertical direction, much of which was not crucial to the simulation of fragmentation. Accommodating these largely empty spaces require carefully establishing a minimum density low enough to prevent undesired additional mass from settling onto the disk and affecting fragmentation. A more efficient approach would be to simulate the vertical direction in 4 to 6 scale heights while still retaining the radial and azimuthal extents. Additionally, the periodic vertical boundary conditions for the self-gravitational potential rather than typical outflow conditions and may contribute to the somewhat erratic α stresses calculated in Figure 3.6.

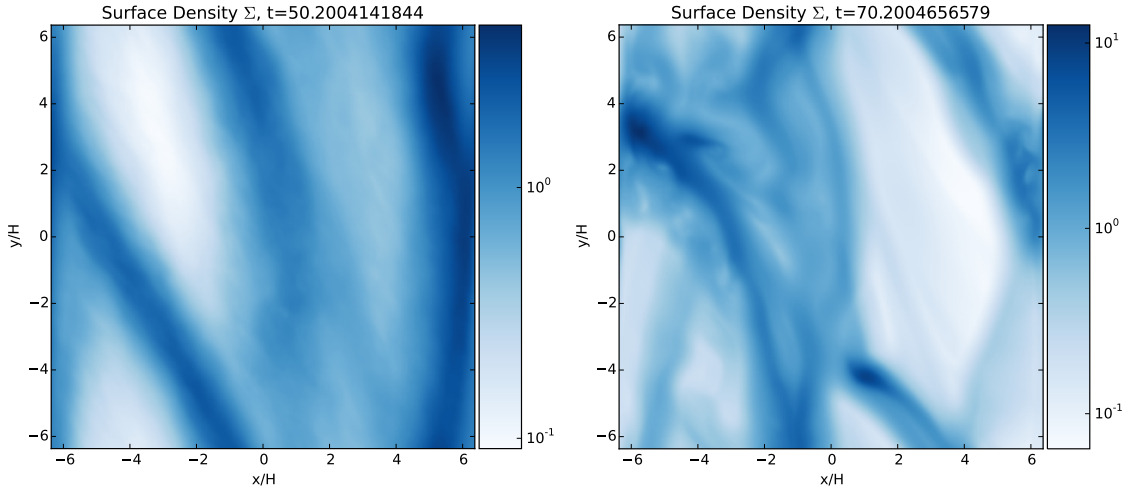


Figure 3.13: Left: Map of projected surface density from a low resolution 3D run $N^3 = 256^3$ with efficient cooling $\beta = 2$. While not as clumpy as its 2D counterpart to the right, similar sheared self-gravitating filaments are clear. Right: Map of surface density from a 2D simulation with the same effective resolution as the 3D model to the left. This simulation is the same as the models from Baehr & Klahr (2015) and Chapter 2, but scaled down to mimic the size and resolution of the 3D runs presented in this chapter.

While the simple cooling with irradiation used here might result in convergence of the cooling criteria, a better description of collapse and fragmentation will ultimately come from using radiative transfer calculations, whether in the form of approximations such as flux-limited diffusion (Mayer et al., 2007; Boley et al., 2006) or more sophisticated but intensive methods like ray tracing. These methods take into account the gas opacities and optical depths to more appropriately model variations of cooling from one region of the disk to the next.

One feature of protoplanetary disks which is typically ignored in self-gravitating contexts but might have an effect on fragmentation is the disk magnetic field. While ionization is expected to be low in the cold outer regions of circumstellar disks, the effect of magnetic fields on fragmentation is uncertain. Only a pair of studies have looked into the affects of magnetic fields on the fragmentation criterion (Riols & Latter, 2016; Forgan et al., 2016), and came to opposite conclusions about the importance of magnetic fields on disk fragmentation. Gravitational instabilities are expected to dominate at the larger scales, but whether magnetic fields are a comparable stabilizing influence to the pressure support is still a subject of further investigation.

One peculiar behavior encountered in these simulations was a tidal distortion of the fragments with a regular frequency on the order of Ω, apparent in the wiggles of the maximum mass of Figures 3.3 and 3.4. The effect on the fragment is clear in Figure 3.14, shearing the fragment in both the y-direction and the lightly in the vertical direction. Initially thought be the result of the equation of state or the parametric instabilities mentioned earlier, this now appears to be due to the choice of simulation box size. This will be addressed further in the following chapter, as the work of (Booth & Clarke, 2019) informed the decision to use a larger x-y simulation domain for those simulations.

In this chapter I have conducted a convergence study of 3D shearing box self-gravitating disks with simple cooling and background irradiation. Starting with a marginally unstable disk, simulations were run at various cooling timescales through the initial burst phase where they have either formed a fragment or settled to gravitoturbulence. These simula-

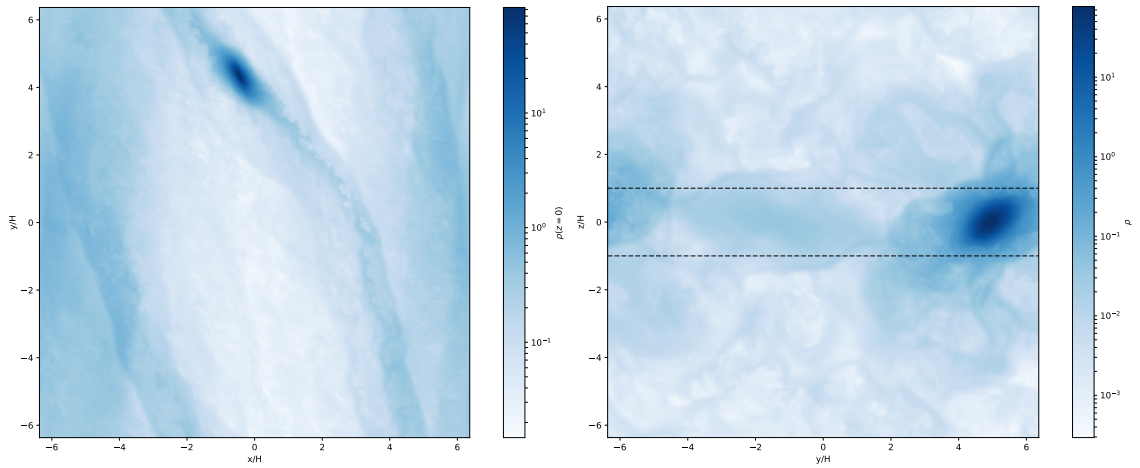


Figure 3.14: Left: Tidal distortion in the radial-azimuthal midplane of a fragment shown using the logarithmic mass density. Right: In the vertical-azimuthal plane. Dashed lines indicate the vertical region $-H$ to H.

tions were run at several high resolutions to investigate fragmentation conditions and the convergence of the solution with respect to the cooling timescale and my results are as follows:

1. Consistent with analytic results, I find that 3D vertically stratified disk simulations fragment for two high resolution simulations with a cooling timescale boundary of roughly $\beta_{\rm crit} = 3$. Fragmentation only occurs when a scale height is resolved by at least 40 grid cells.

2. Convergence indicates that there is a firm criterion for fragmentation in a disk: where disks cool rapidly with $\beta < 3$. This limits fragmentation to the distant regions around a star where temperatures are relatively low and dominated by stellar irradiation.

3. By achieving convergence, it had been demonstrated that the standard β cooling model with a fixed floor temperature may have no inherent flaws numerically when used in 3D models. It is nevertheless desirable to use a more realistic model for disk cooling in future studies.

4. Using a pressure-supported Hill criterion, I find a suitable minimum density consistent with the formation of brown dwarfs and low-mass stars. This serves as a useful criterion for the inclusion of sink cells to aid in the efficient computation of fragmentation as it can be applied to simulations with low resolution simulations which are unable to fragment.

4 Solid Particles in a Fragmenting Disk

The following chapter is based on the forthcoming publication Baehr & Klahr (2019) titled "The Concentration and Growth of Solids in Fragmenting Circumstellar Disks".

Gas giant planets are predominantly gaseous objects, but likely with solid cores and overall metallicities closer to stars than terrestrial planets (Helled & Bodenheimer, 2010; Kreidberg et al., 2014). Recent observations of the gravitational moments of Jupiter by the Juno mission indicate that it has a substantial solid or partially dissolved core of $> 10 M_\oplus$ (Bolton et al., 2017). Such a massive core is a strong indicator of formation by core accretion, as the accretion of a significant gas envelope requires a planetary core of around $10 M_\oplus$ (Pollack et al., 1996). Formation through gravitational instabilities (GI), on the other hand, involves the direct collapse of dense, gaseous regions with low dust-to-gas ratios into fragments. These massive, cool disks are most likely when a disk is young and devoid of considerable particle accumulations, neither a core or metallicity over solar is expected through gravitational instabilities, however recent simulations suggest that fragments may have atmospheres which are distinct from the disk and envelope from which they form (Ilee et al., 2017).

Recent observations of giant exoplanets and their atmospheres have led to the discovery of a number of trends in atmospheric chemical compositions and the development of theories which link these compositions to planet formation. Carbon to oxygen ratios (C/O) are expected to be an indicator of formation location or formation mechanism (Öberg et al., 2011) and can be observed through spectroscopic measurements of transiting planets (Charbonneau et al., 2002) and directly imaged planets (Lee et al., 2013). While transiting planets are typically located at orbits too close to be formed *in situ* by GI, they still provide a useful constraints on expected metallicity values of exoplanet atmospheres (Espinoza et al., 2017; Thorngren et al., 2016). In the case of the HR8799 system, the elemental abundances abundances are most consistent with superstellar C/O ratios in favor of core accretion, but the discrepancy between inner and outer planets suggests a different accretion histories and formation pathways are possible (Lavie et al., 2017).

Gravitationally unstable disks may still be capable of producing large gaseous objects with significant solid cores by considering that even in low dust-to-gas environments, dust can collect in fragments (Helled et al., 2006). While spiral arms produced by GI are well documented locations of particle concentration (Gibbons et al., 2012; Boss, 2015), it is still unknown just how enriched a fragment formed therein may become enriched with solids during formation and immediately after, when temperatures are not high enough to evaporate solids. Boley & Durisen (2010) investigated particle concentration in self-gravitating disks with global simulations that included two particle sizes, and found considerable solid concentrations (i.e. cores and embryos), interior and exterior to fragments. With high-resolution shearing box simulations of fragmentation, I aim to explore this process further.

A number of studies have looked into the behavior of solid particles in marginally gravi-

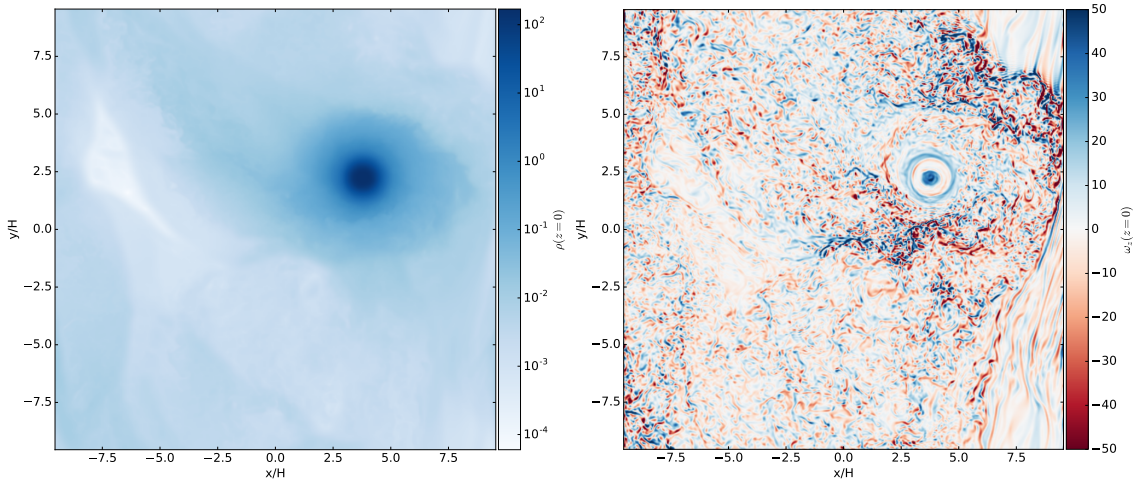


Figure 4.1: A slice of the gas density (top) and vertical component of the midplane vorticity (bottom) of a fragmenting 3D simulation without particles. Unlike in 2D simulations, vorticity is not particularly strong or coherent at the location of the fragment, although non-axisymmetric structure is prominently seen in the vorticity.

tationally unstable disks, but have mostly focused on particle evolution in gravitoturbulent disks. These range from looking into the collection of particles in voritices (Gibbons et al., 2015) and fragments (Boley & Durisen, 2010), shock fronts (Gibbons et al., 2012), and other relations with the GI turbulence (Shi et al., 2016). These find that the particles of typically moderate sizes will collect and form high solid concentrations through aerodynamic drag and gas-dust coupling. Shi & Chiang (2013) consider a concentration of particles well above the typical dust-to-gas ratios for gas disks so that the particles will naturally fragment even if the gas disk is not gravitationally unstable (i.e. $Q_{\rm gas} \geq 10$). Of interest is the regime where the gas disk is gravitationally unstable ($Q_{\rm gas} \leq 1$), but the dust is not, therefore the gas will fragment but initial dust overdensities are not formed through the direct gravitational collapse of the dust, but by the hydrodynamic interactions between the gas and the dust.

For the study of fragment formation including particles, Boley & Durisen (2010) used global radiation hydrodynamic simulations, investigating the large-scale structure of gravitationally unstable disks with two particle sizes. In these simulations, particles are first concentrated into the spiral arms, reaching surface densities which start contributing to the self-gravity in these regions, warranting the inclusion of particle self-gravity. At this point, the particles and gas then collapse together, and the particles rapidly concentrate at the center and potentially constitute a new mode of planet formation (Nayakshin et al., 2014). Boley & Durisen (2010) ultimately find that there are significant concentrations of small grains at the center of fragments, however, their resolution is insufficient to capture the effects of small scale turbulence caused by parametric instabilities near the midplane (Riols & Latter, 2017).

In this thesis, I include figures which show the vorticity of the gas in the simulation. The vorticity of the gas is a measure of the local rotation of a fluid and is a simple diagnostic of turbulent motion, defined as the curl of the velocity field $\boldsymbol{u}$: $\omega = \nabla \times \boldsymbol{u}$. In the case of the 2D midplane figures, the value of the vertical component is plotted, which also serves

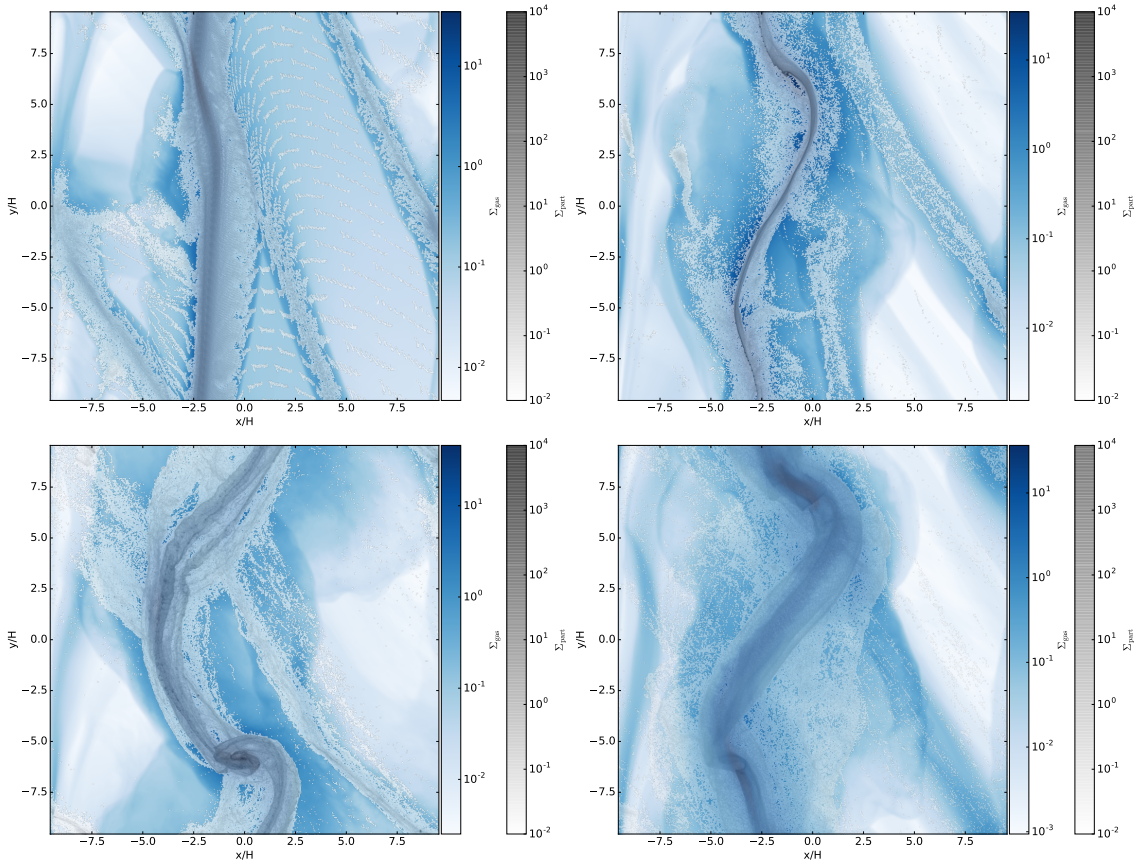


Figure 4.2: Map of the gas surface density (blue) with particle surface density superimposed at roughly the time the simulation transitions from forming the linear axisymmetric perturbation to the non-linear collapse of the fragment. It is at this stage that particles are being concentrated into densities which begin affecting the total self-gravitational potential of the system. Each simulation has a million particles of a single species: St = 0.01 (top left), St = 0.1 (top right), St = 1.0 (bottom left) and St = 10 (bottom right). Particle sizes St = 0.1 and St = 1.0 will concentrate enough to considerable solid cores immediately, while St = 0.01 and St = 10 will have marginal solid concentrations.

to show the rotation in and around a fragment

$$\omega_z = \left(\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y} \right) \hat{\boldsymbol{z}}. \tag{4.1}$$

4.1 Particle Aerodynamics

For this study I conducted 3D hydrodynamic shearing box simulations of a self-gravitating disk with Lagrangian 'swarm' particles embedded in the Eulerian mesh. Both gas and particles are treated as self-gravitating and drag and back-reaction is calculated on the particles and gas, respectively, by assuming particles are mapped to the grid with triangular-shaped clouds. Local simulations simulations allow the Toomre wavelength $\sim H$ to be well-resolved and keep the boundary conditions periodic. The shearing box equations are the same as those used in the previous chapter (see Section 3.2), except with the addition of particle dynamics.

Self-gravity is solved in Fourier space by transforming the surface density to find the potential at wavenumber k and transforming the solution back into real space. The solution to the Poisson equation in Fourier space at wavenumber k is

$$\phi(\boldsymbol{k}, t) = -\frac{2\pi G\rho(\boldsymbol{k}, t)}{\boldsymbol{k}^2}, \tag{4.2}$$

where $\phi = \phi_{\rm g} + \phi_{\rm d}$ and $\rho = \rho_{\rm g} + \rho_{\rm d}$ are the potential and density of the gas plus dust particles combined.

Drag and Backreaction

Particles in these simulations are not pure tracer particles which merely follow the flow, but interact with the gas through their aerodynamic properties, which depend on their mass and

Particles are added such that the initial distribution maintains a physically motivated dust-to-gas ratio of 1:100, roughly that of the interstellar medium (ISM), although this may be an overestimate of the gas content of the disk (Ansdell et al., 2016; Miotello et al., 2017). Furthermore, it is assume that particles have already sedimented to the midplane and so are initially only distributed in this plane. Particles evolve as a collection of solids, or as a superparticle i, with position $\boldsymbol{x}^{(i)}$ and velocity $\boldsymbol{w}^{(i)}$ as in Youdin & Johansen (2007)

$$\begin{aligned}\frac{d\boldsymbol{w}^{(i)}}{dt} &= 2\Omega w_y^{(i)}\hat{\boldsymbol{x}} - \frac{1}{2}\Omega w_x^{(i)}\hat{\boldsymbol{y}} - \nabla\phi \\ &\quad + \frac{1}{\tau_s}(\boldsymbol{w}^{(i)} - \boldsymbol{u}(\boldsymbol{x}^{(i)})) \end{aligned} \tag{4.3}$$

$$\frac{d\boldsymbol{x}^{(i)}}{dt} = \boldsymbol{w}^{(i)} - \frac{3}{2}\Omega x^{(i)}\hat{\boldsymbol{y}}, \tag{4.4}$$

Drag is applied to particle assuming that their mass is distributed over several cells using a weighting function W, such that the particle velocities at position $\boldsymbol{x}^{(i)}$ in Equation (4.3) are

$$\boldsymbol{u}(\boldsymbol{x}^{(i)}) = \sum_j W\left(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\right)\boldsymbol{u}^{(j)}, \tag{4.5}$$

for cell-centered gas positions and velocities $\boldsymbol{x}^{(j)}$ and $\boldsymbol{u}^{(j)}$, respectively. This spreads particle mass primarily in one cell, but also in the neighboring 26 cells. This produces a smoother overall particle distribution which is less likely to introduce self-acceleration onto a particle. This particle mass weighting scheme, called triangular-shaped cloud (TSC), is defined for distances δx between particle position and grid centers and assuming grid cells of width Δ (Youdin & Johansen, 2007).

$$W(\delta x) = \begin{cases} \dfrac{3}{4} - \dfrac{\delta x^2}{\Delta^2}, & |\delta x| < \Delta/2 \\ \dfrac{1}{2}\left(\dfrac{3}{2} - \dfrac{\delta x}{\Delta}\right)^2, & \Delta/2 < |\delta x| < 3\Delta/2 \\ 0, & |\delta x| > 3\Delta/2 \end{cases} \tag{4.6}$$

Lower order methods for mapping particles to the grid are nearest grid point (NGP, 0^{th} order) take the drag force from the center of the nearest cell, which can be a very inaccurate estimation. The first-order interpolation of particle position, called particle-in-cell (PIC)

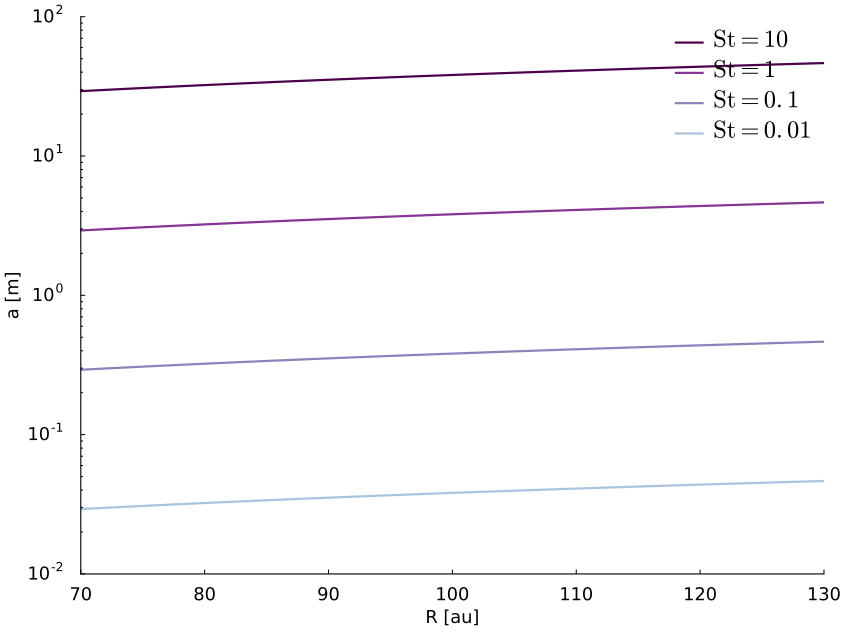


Figure 4.3: The particle sizes to which the Stokes numbers used in this paper correspond, for a range of disk radii around 100 au. Each line assumes a $R^{-1/2}$ dependence of the gas density with radius in Equation (4.8).

or cloud-in-cell (CIC) distributes the particle mass evenly over a space of one grid cell center at any position which may overlap with up to 8 cells (Hockney & Eastwood, 1981). The assignment of the back reaction force of the particles to the gas is applied using the same weighting function W.

Size

The friction time τ_f (also referred to as the Stokes number) is the particle stopping time τ_s normalized by the dynamical time Ω^{-1}

$$\mathrm{St} = \tau_f = \Omega \tau_\mathrm{s}. \tag{4.7}$$

This gives a hydrodynamic sense of the particle size in the Epstein regime, where the particle size is smaller than the mean free path of the gas. The stopping time is defined as (Weidenschilling, 1977)

$$\tau_\mathrm{s} = \frac{a\rho_\cdot}{c_\mathrm{s}\rho_\mathrm{g}}, \tag{4.8}$$

where a is the particle diameter and $\rho_\cdot$ is the density of a superparticle. Larger particles are less coupled to small scale gas motions, retain their initial perturbations for longer and thus have higher Stokes numbers. As a corollary, smaller particles are well-coupled and will quickly take the form of the gas motions in its vicinity and thus have low Stokes numbers. Particle mass is calculated through the dust-to-gas ratio $\epsilon = \rho_\mathrm{d}/\rho_\mathrm{g}$.

No radial pressure gradient is assumed so the radial migration of particles is not included. Because particles are added at the ISM dust-to-gas ratio $\epsilon = 10^{-2}$ the contribution of particles to the gravitational potential is negligible and thus the potential is dominated by the gas distribution.

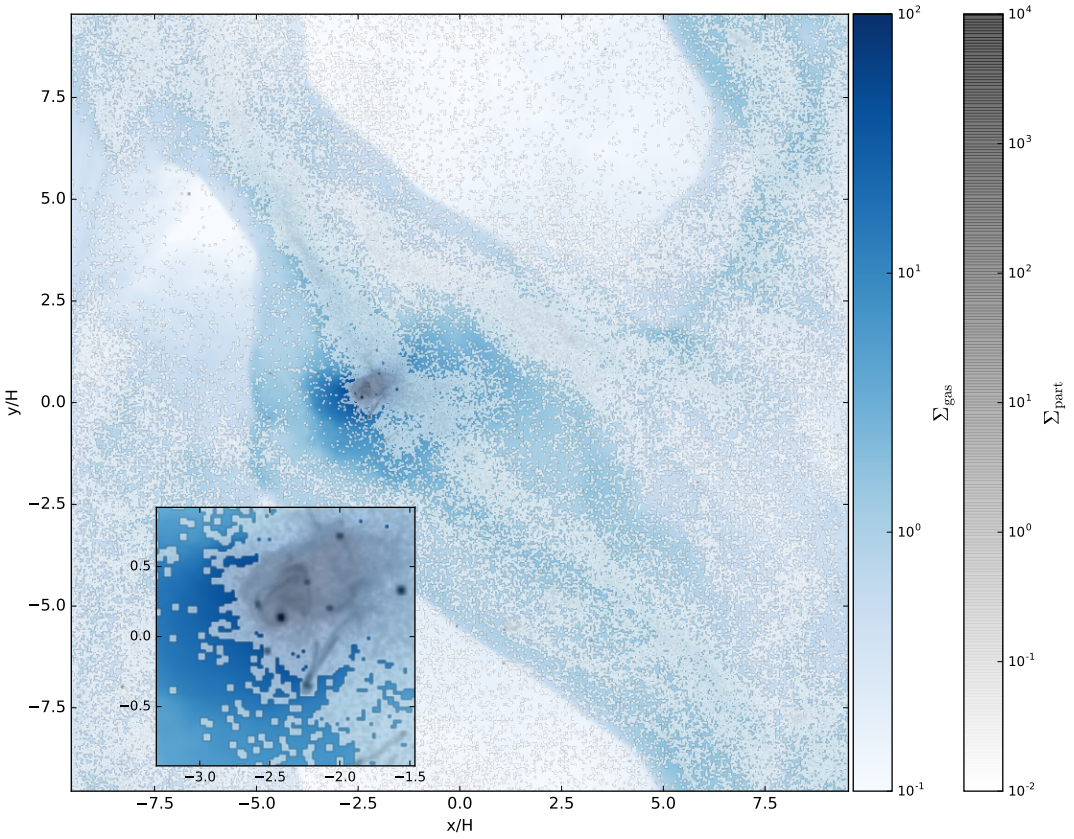


Figure 4.4: A map of the vertically integrated gas and particle densities of a fragmenting 3D simulation. The inset of the left figure shows a more detailed view of the fragment center and the various particle clusters in its vicinity.

4.2 Fragmentation

All simulations were run using the high-order finite-difference code Pencil (Brandenburg, 2003) well-suited for turbulent flows. The code is stabilized through a six-order hyper-dissipation scheme (Yang & Krumholz, 2012) ensuring power is preserved at large scales, but numerical noise damped at small scales. The 3D gas-only simulation shown in Figure 4.1 and included here for reference, is based on those in Baehr et al. (2017) but altered to match the dimensions of the other runs presented here; more detailed descriptions of the numerical methods and initial conditions can be found therein.

Implementation in Pencil

Particles are included in the Pencil code using Lagrangian 'swarm' particles which represent collections of similar-sized particles all moving together (Johansen & Youdin, 2007; Youdin & Johansen, 2007; Schreiber & Klahr, 2018). These particles feel the effect of self-gravity and feel drag forces from the gas but there is no feedback of the particles onto the gas and particles do not collide. Included are particles of various sizes, i.e. with a range of Stokes numbers $\mathrm{St} = [0.01, 0.1, 1, 10]$ that also quantify how quickly the particle couple to the gas, to evaluate the ability of these various sizes to be captured by fragments. Each simulation only contains one particle size, or species, at a time.

The evolution of the particles is governed by equations (4.3) and (4.4), and are mapped

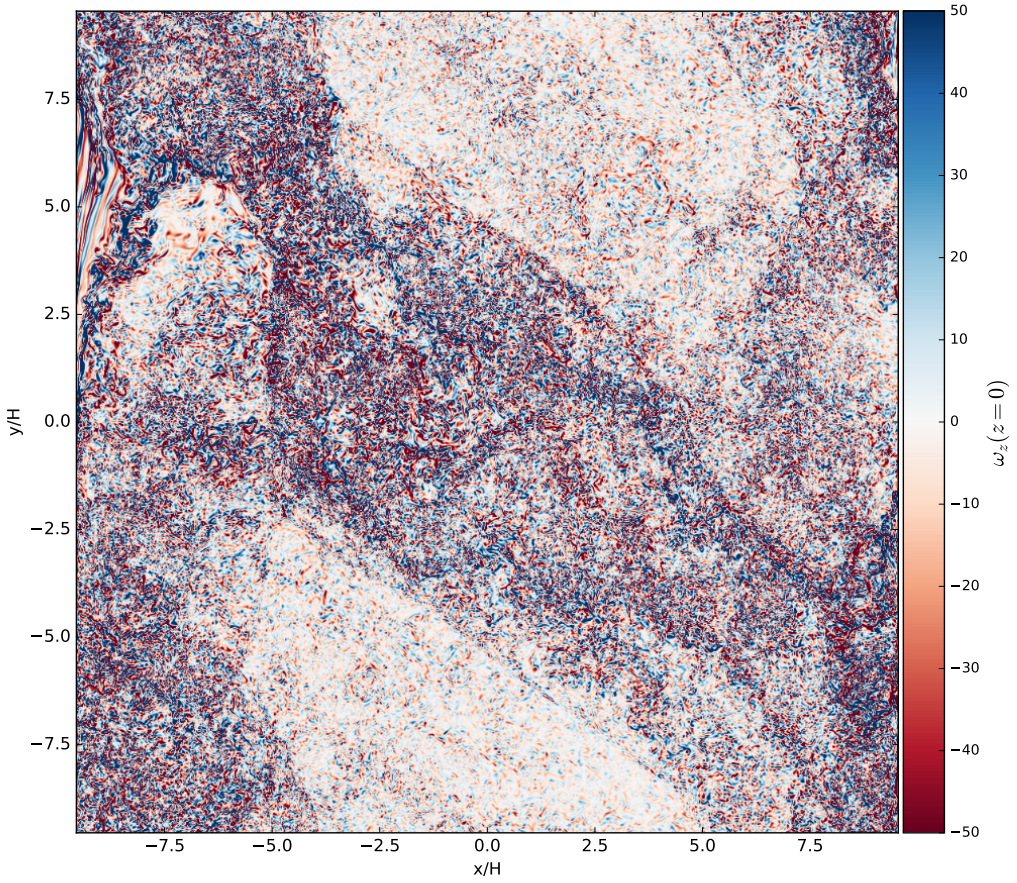


Figure 4.5: A slice at the midplane of the vertical component of the vorticity of a fragmenting 3D simulation, corresponding to the density map in Figure 4.4. Compared to the Figure 4.1, where a vortex indicates rotation of the fragment and the absence of a vortex with particles included.

to the grid using a triangle-shaped-cloud method, to keep the particles from becoming too discretely distributed and for the correct calculation of the back-reaction of the gas onto the dust (Yang & Johansen, 2016).

Initial Conditions

These simulations cover $18H$ in radial and azimuthal directions and $L_z = 2H$ in the vertical, with a resolution of either 30 or 60 grid points per scale height in all spatial dimensions. This fulfills the resolution requirements to avoid truncation errors which can become unstable (Truelove et al., 1997; Nelson, 2006) as well as the requirements for individual particles, which require that the grid spacing $\Delta x < c_s \tau_s$ (Bai & Stone, 2010). Additionally, the number of particles must be enough such that at the midplane layer there is 1 particle per cell, which is marginally met in the high grid resolution simulations for a 1024×1024 midplane with 10^6 total particles. Each simulation contains a single particle species of constant size defined by its Stokes number. All simulations include a million particles which are initially uniformly distributed throughout the domain, but rapidly settle to the disk midplane and concentrate with the gas overdensities. Because the gas is stratified and the particles are not, the initial dust-to-gas values near the vertical boundaries will be higher than the overall metallicity of 1/100 and the midplane values

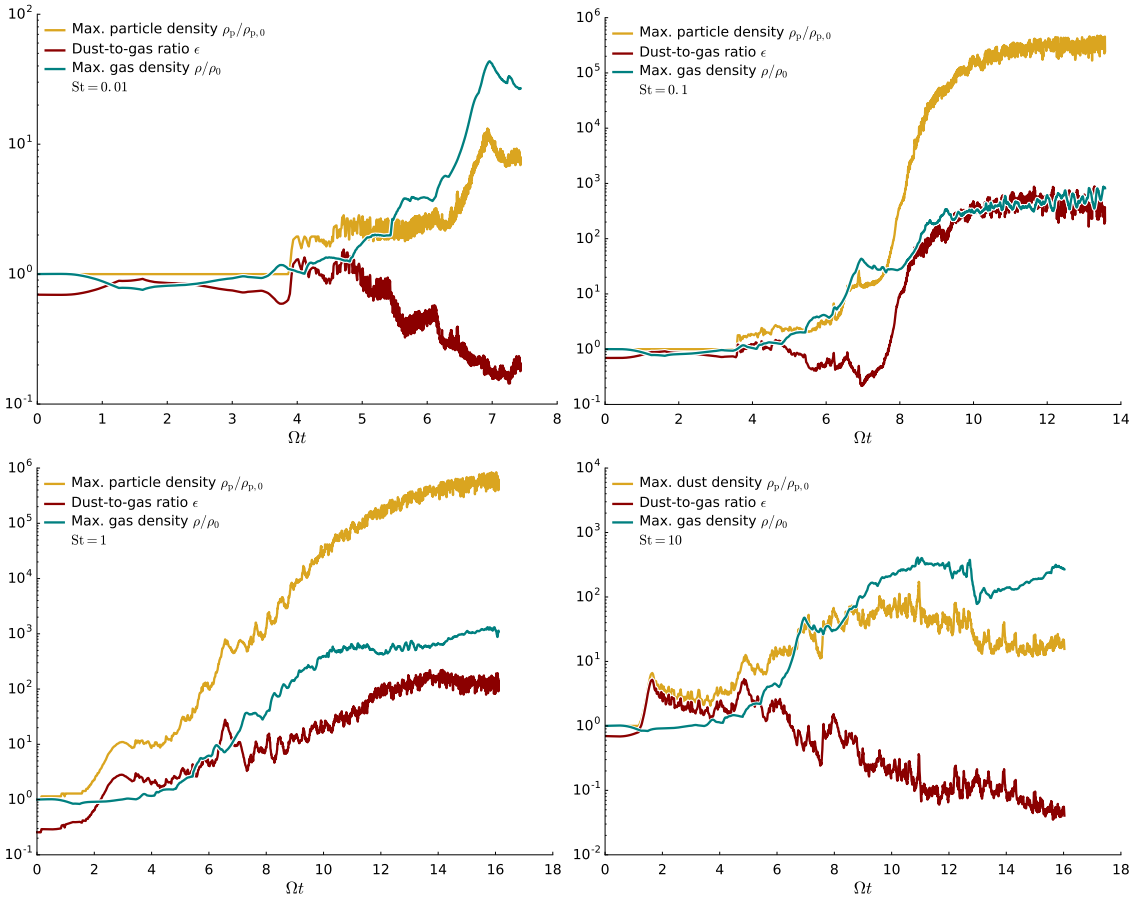


Figure 4.6: Time evolution of the maximum particle density (yellow lines), maximum gas density (teal lines) and maximum dust-to-gas ratio ϵ (red lines) for the particle stokes numbers in these 3D simulations: St = 0.1 (top), St = 1 (middle), and St = 10 (bottom). The intermediate sizes grow to the highest densities, but St = 1 are able to maintain a high particle density.

will be lower although the overall metallicity is still at the desired value.

Figure 4.1 show the particle-less fiducial simulation which demonstrates the fragmentation of the gas which is expected in all simulations. Particularly interesting is the presence of a strong coherent anticyclonic vortex at the fragment center as well as a second rotating region separated by a largely stationary gap. If particles concentrate at the center of fragments the existence of a vortex could aid the process.

4.3 Concentration within Fragments

Three-dimensional models of self-gravitating disks were run to consider the effects of vertical mixing and gravitoturbulence on particle concentrations within gravitationally fragmenting disks. Vortices are normally invoked to explain the trapping of particles in non-axisymmetric structures (Gibbons et al., 2015) but vortices may not persist long enough in vertically-stratified simulations (Lin & Pierens, 2018), affecting the ability of particles to form planetesimals or cores.

Initiated with $Q_0 = 1$, all simulations quickly proceed to collapse, with particles settling to the midplane at different rates depending on their Stokes number. Smaller particles, i.e.

Table 4.1: Performed simulations and their various resolutions, cooling times β, particle sizes St, metallicity Z, and fragment/core masses (M_f and M_c, respectively)

Simulation	Grid Cells	Stokes #	Z_0	Z/Z_0	$M_f/M_{\rm Jup}$	$M_c/M_\oplus$
G512t2	$512^2 \times 64$	–	–	–	17.2	–
P512t2pss	$512^2 \times 64$	0.01	10^{-2}	3.4	5.24	57.9
P512t2ps	$512^2 \times 64$	0.1	10^{-2}	14.6	1.32	62.3
P512t2p	$512^2 \times 64$	1	10^{-2}	1.8	16.2	89.1
P512t2pl	$512^2 \times 64$	10	10^{-2}	1.7	2.3	–
P1024t2pss	$1024^2 \times 128$	0.01	10^{-2}	1.7	–	–
P1024t2ps	$1024^2 \times 128$	0.1	10^{-2}	5.1	4.8	69.0
P1024t2p	$1024^2 \times 128$	1	10^{-2}	3.8	5.7	70.7
P1024t2pl	$1024^2 \times 128$	10	10^{-2}	1.5	7.19	–
P1024t2psl	$1024^2 \times 128$	0.1	8×10^{-3}	5.0	3.83	49.3
P1024t2psm	$1024^2 \times 128$	0.1	1.33×10^{-2}	4.8	8.01	147.6

St = 0.01 will take longer as minute gas disturbances will deflect their motion towards the midplane, while larger particles, St > 10, will be able to fall largely unimpeded. Figure 4.6 shows the evolution of the maximum gas and particle densities with time for particle sizes (from top to bottom) St = $[0.1, 1, 10]$. One can see a clear trend towards higher particle concentrations with smaller particle size.

The rapid concentration of particles within collapsing axisymmetric and non-axisymmetric structures can be seen in the different panels of Figure 4.2. Each panel shows the collapse of the gas and particles for a different particle size, with background color showing the gas surface density with particle surface density overplotted in semi-transparent greyscale. As one might expect, the particles which are most coupled to the gas show the most concentration in the gas overdensities. In the case of the largest particle size (St = 10), the inability to couple strongly with the gas results in a steadily decreasing dust-to-gas ratio, unlike in the case of the smaller particles. This is similar to the simulations with 1 km sized rocks from Boley & Durisen (2010), which showed the delayed response of the larger particles to the gas drag resulted in wide particle arms which were in different positions than the gas spiral arms. This could lead to scenarios where solids can concentrate outside of the gas overdensities, enriching very different regions of the disk.

Small to intermediate particles eventually collect the best within the collapsing fragment and eventually result in at least moderate cores. This is apparent in the radial profiles of the gas fragments which form in these simulations and are shown in Figure 4.7. The gas (teal line) shows a roughly Bonnor-Ebert density profile, with constant density within a radius of $\sim 0.1H$ and a power law decline thereafter. This is consistent with the isothermal temperature structure assumed for Bonnor-Ebert spheres, here shown with the dark blue line.

Regardless of the particle size, particles will concentrate to some degree compared to the background distribution, coinciding with the location of the gas fragment. This is apparent in the top left and bottom right panels of Figure 4.7, which correspond to the largest and smallest particle species, and are least likely to concentrate strongly. While no core is formed, there is still increasing particle density towards the center of the fragment.

In the case of St = 0.1, the core is especially pronounced from the surrounding protoplanetary envelope, able to form a distinct particle clump within the fragment itself. This constitutes the most significant core formed in these simulations and as detailed in the following section, is tens of Earth masses in size.

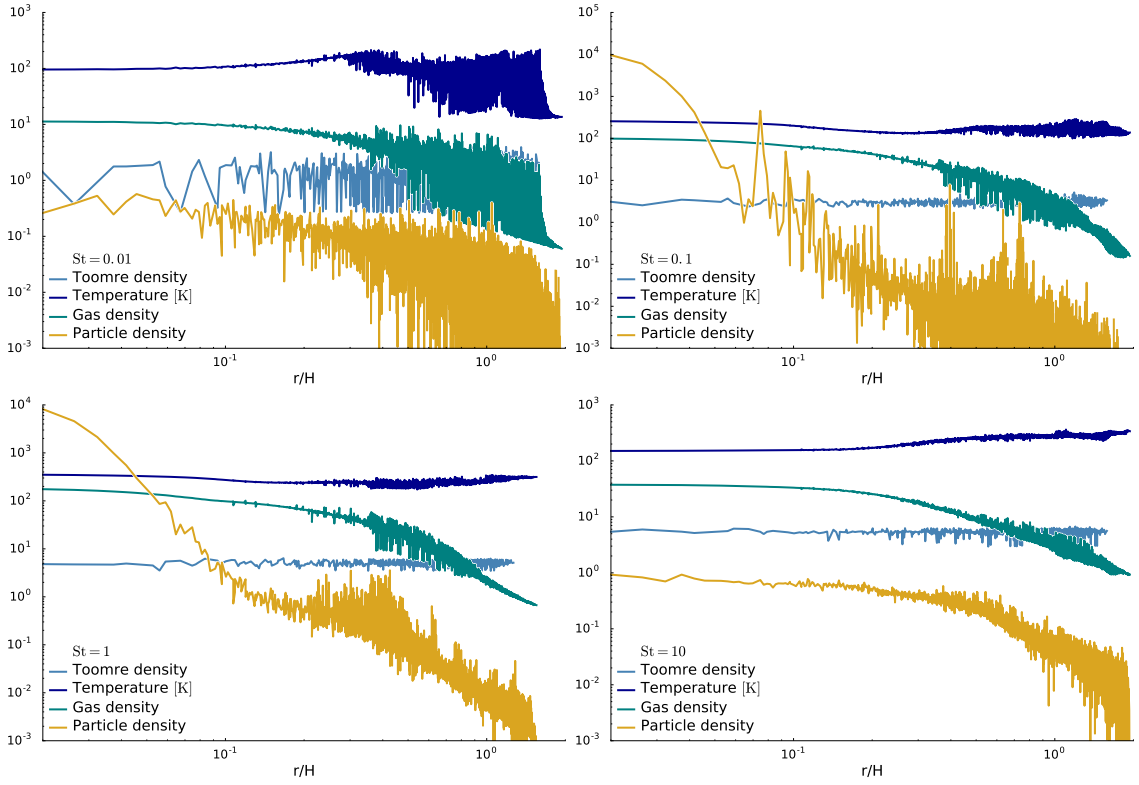


Figure 4.7: Radial profile of a fragment by gas density, particle density (both in code units) and temperature (in) for simulations containing a single particle species of size St = 0.01 (top left), St = 0.1 (top right), St = 1.0 (bottom left) and St = 10 (bottom right). Values are determined by first locating the cell with the highest gas volume density and then averaging the separately the particle and gas densities for each radial distance away from that cell.

Masses

While solid clumps are small in size and dozens form in the simulation, their peak densities can be far larger than that of the gas. While the dust and gas have similar Toomre stability criteria (Shi & Chiang, 2013), the dust has a scale height smaller by a factor of a few, which explains why the unstable wavelength is much lower and thus many more dust fragments appear.

Even though solids do not necessarily concentrate at the center of fragments, they do tend to collect at moderate gas overdensities, either in the form of clumps or in the gravitoturbulent shear waves. Particles with St = 0.1 couple efficiently to the motions of the gas as expected and collect more efficiently than St = 1 particles.

Importantly, from these simulations it is desirable to determine the size and mass of the solid clumps and how enriched are the atmospheres of gas giants/brown dwarfs that form. First one needs to define the initial parameters in physical units consistent with a shearing box located 100 au from a 1 $M_\odot$ star. The orbital frequency at 100 au is 1.98×10^{-10} s^{-1} and the sound speed of a 10 Kelvin monatomic gas (with $\gamma = 5/3$) is 262 ms^{-1}. The Toomre surface density Σ_T of the entire simulation box in this case is

$$\Sigma_{T,0} = \frac{c_{s,0}\Omega}{\pi G Q_0} = 37 \mathrm{gcm}^{-2} \tag{4.9}$$

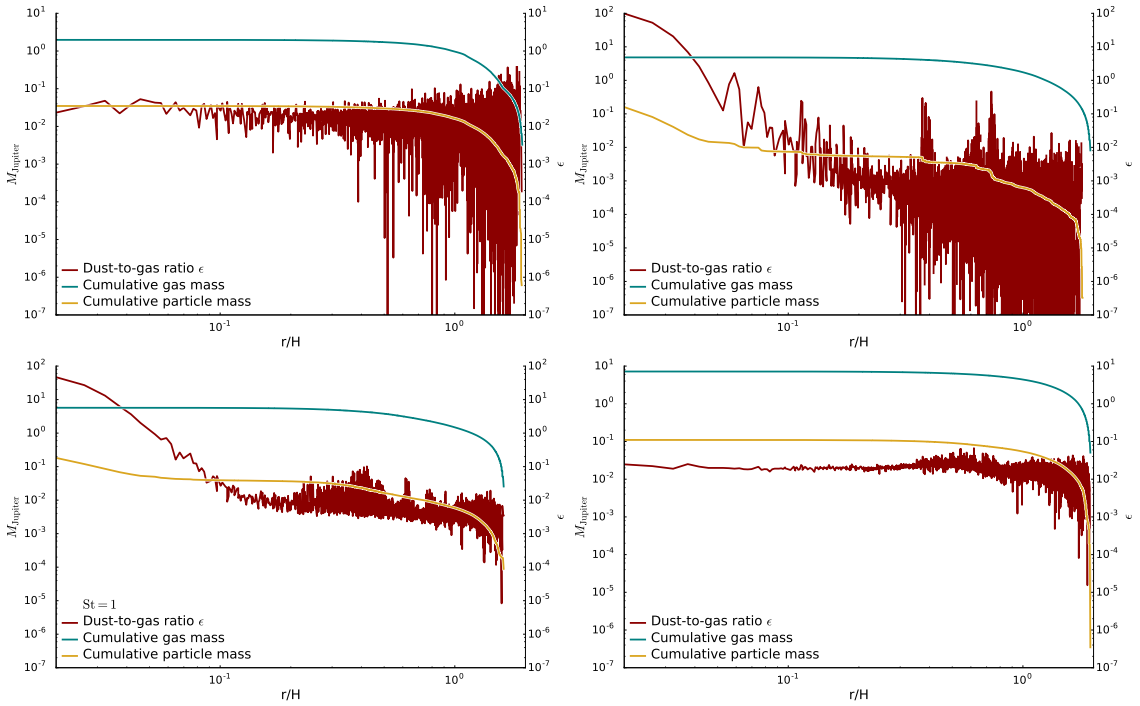


Figure 4.8: The cumulative gas and solid masses (teal and yellow lines, respectively), starting from the outside and going inwards, of the fragments whose radial mass distributions are shown in Figure 4.7. The red line shows the average dust-to-gas ratio for each radius All masses are those of objects expected to be formed at an orbital distance of 100 au from a 1 $M_{\odot}$ star.

which is roughly an order of magnitude greater surface density than the minimum mass solar nebula density at this radius.

From Σ_{T}, I calculate the unit temperature from the Gaussian profile of the initial vertical gas mass distribution

$$\Sigma_{\mathrm{T}} \equiv \int_{-H}^{H} \rho_0 e^{-z^2/2H^2} dz \tag{4.10}$$

where $\rho_0 = 0.29\hat{\rho}$. Solving for the unit volume density $\hat{\rho}$ gives a value of $5 \times 10^{-12}\mathrm{gcm}^{-3}$, with a disk aspect ratio of $H/R = 0.1$.

The results of these calculations are shown in table 4.1, along with the solid mass ratio

$$Z_{\mathrm{solid}} = \frac{M_{\mathrm{solid}}}{M_{\mathrm{total}}} \tag{4.11}$$

which is measure of the amount of solids to gas with in roughly one scale height of the peak density. Solid mass ratios calculated here are all at least 0.01 and in the case of St ≤ 1, often more comparable to the ratios observed in Jupiter or Saturn. Figure 4.8 shows the total mass in gas (teal) and solids (yellow) as a function of radius for the fragment profiles in Figure 4.7. From this one can see that gas fragments are consistently formed with a few Jupiter masses and varying degrees of particle concentration. Particularly in the case of St $= 0.01$ and St $= 10$, one can see that although there is several Earth masses worth of material in the entire fragment, it is distributed in a similar fashion to the gas (i.e.: ϵ remains constant throughout the inner regions of the fragment), and thus no significant core is formed. In the cases of St $= 0.1$ and St $= 1$, ϵ keeps rising as one approaches the

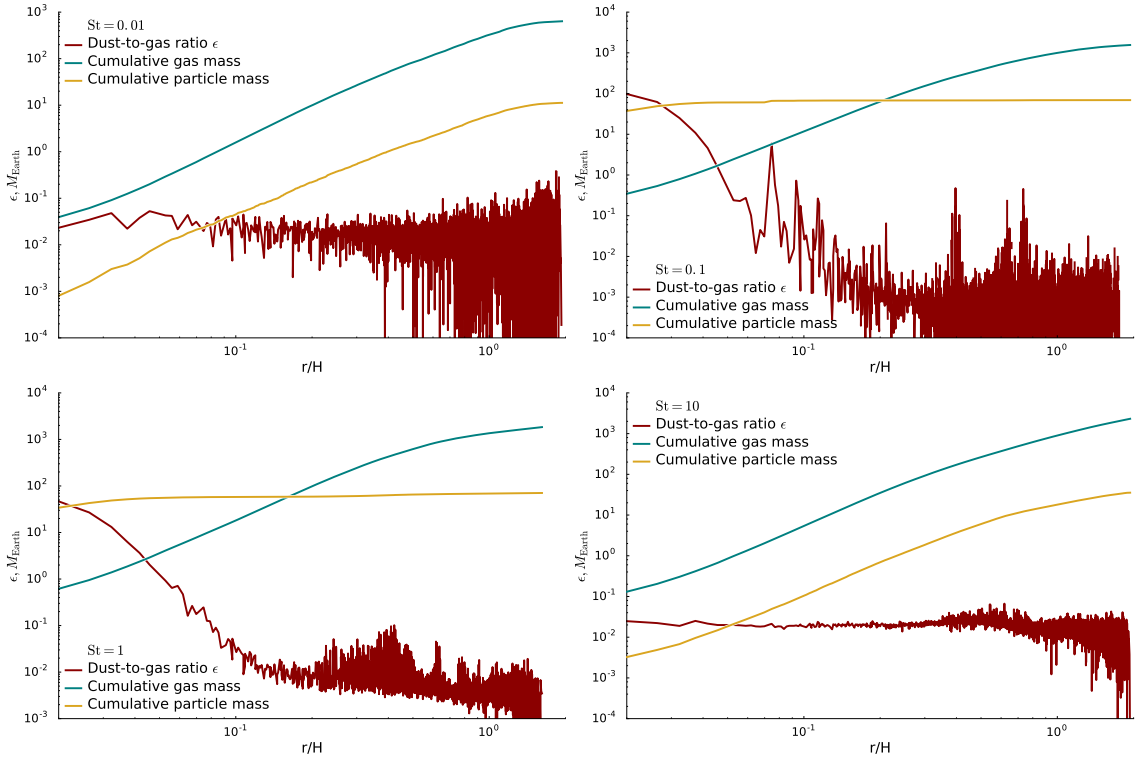


Figure 4.9: The cumulative gas and solid masses (teal and yellow lines, respectively), starting from the inside and going outwards, of the fragments whose radial mass distributions are shown in Figure 4.7. The red line shows the average dust-to-gas ratio for each radius All masses are those of objects expected to be formed at an orbital distance of 100 au from a 1 $M_\odot$ star.

fragment center and the formation of a core begins to take shape.

In the case of all particle sizes, fragment envelopes show dust-to-gas ratios occasionally larger than their initial $\epsilon_0 = 0.01$ level, although the scatter is significant. Only the simulation with larger St = 10 particles show ratios consistently above the initial level.

Metallicity

In simulations which do not form a core, the fragment envelopes show radial dust-to-gas ratios consistently larger than their initial $\epsilon_0 = 0.01$ level, although the scatter is sometimes significant, particularly for values below. Clumps which do form cores have atmospheres depleted in heavier grains, but may have significant peaks in the local dust-to-gas ratio which are the locations of smaller secondary solid bodies which have not accreted to the fragment center. Since later accretion is expected to further alter the C/O ratio by introducing material from chemically distinct regions of the disk, the amount of solid material available at formation is important to understand the observations of gas giant atmospheres (Espinoza et al., 2017). Whether or not a core forms could have a significant impact on the atmospheric metallicity and C/O ratio.

It is speculated that the satellites are due to the transfer of angular momentum from the spinning fragment without interaction with particles of Figure 4.1 to the secondary and tertiary bodies which form in the close proximity of a fragment in Figure 4.4. These bodies have masses on the order of an Earth mass and remain in an orbit around the primary for the duration of the short simulation runtime.

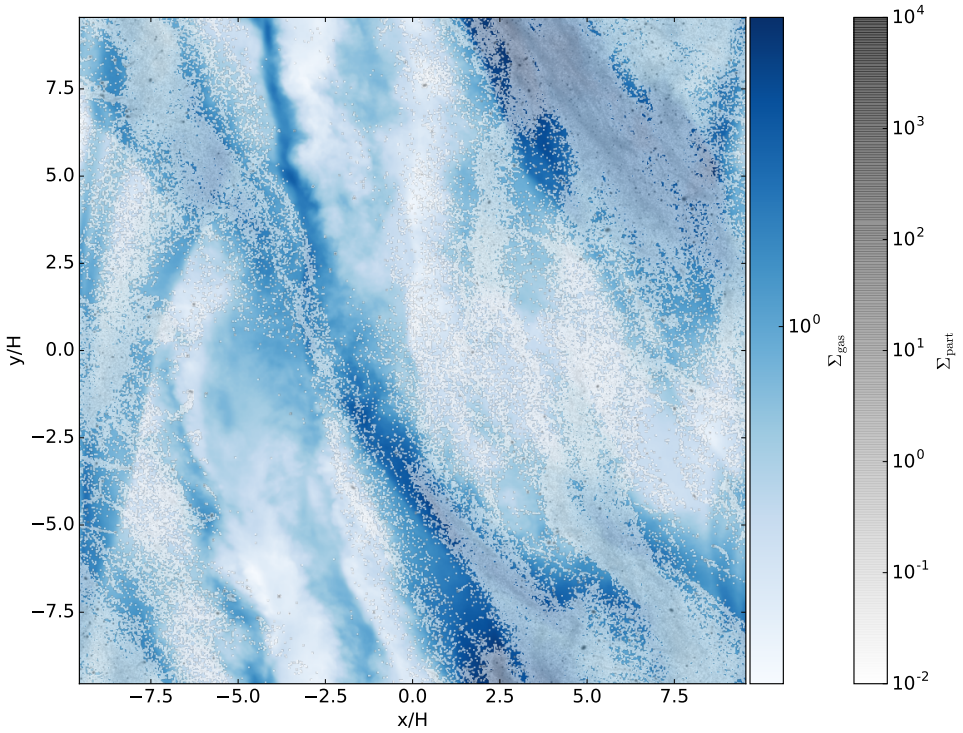


Figure 4.10: Surface density distribution of gas (blue) and solids of St = 1 (greyscale) in a 3D gravitoturbulent simulation. The gas is initiated in stable gravitoturbulence and the particles are also initially stable to fragmentation, with an overall dust-to-gas ratio of 1/100.

There is a noted correlation of giant planet occurrence with the stellar metallicity, with observed planets more common around metal-rich stars, a trend which typically supports core accretion planet formation (Fischer & Valenti, 2005). A pair of simulations were added which deviated from the initial dust abundance to $Z = 0.008$ and $Z = 0.013$ for the simulations which formed the most prominent core, those with particles of size St = 0.1. As seen in Figure 4.11, the result is a fragment with a core regardless if the initial metallicity is (from top to bottom) $Z_0 = [1/125, 1/100, 1/75]$. The difference lies in the fragment and core masses, which are significantly higher in both gas and dust components when more solid material is included. Still, the overall metallicity enhancement remains constant for the clumps in all three simulations, with an increase by a factor of around 5.

4.4 Planetesimal Formation

While fragmentation is the principle interest of these simulations, when fragmentation does not occur, the solid material will still be prone to concentrating within the non-axisymmetric gas overdensities and collapsing into numerous solid clumps. The concentration of solid material in gravitationally unstable disks has been studied before (Gibbons et al., 2012; Shi & Chiang, 2013; Shi et al., 2016) and found that through aerodynamics and self-gravity, particles can concentrate in shock fronts and gas overdensities, even in regions where fragmentation does not occur.

Even when fragmentation does occur, a number of smaller minor solid bodies form elsewhere in the simulation, as seen in Figure 4.4. However, preliminary simulations were carried out which would never fragment doe to a stable Q_0 and a long cooling timescale

$\beta = 10$, with an example result shown in Figure 4.10. Clearly dust collects in a fashion similar to the gas density but also forms a multitude of smaller solid bodies throughout the and not only in the non-axisymmetric features. This however, does not consider the particle relative velocities and fragmentation will likely hinder any significant particle growth outside of spiral arm overdensities (Booth & Clarke, 2016).

This has significant consequences for early planet formation, which may be necessary to explain the rings seen in young disks (Andrews & Birnstiel, 2018; Dullemond et al., 2018) which could possibly be explained by embedded planets. The formation of planets at the location of these rings at such an early time remains tremendously difficult for core accretion and even for pebble accretion scenarios. The formation of planetesimals within the first 0.5 to 1 million years could provide the early material necessary to accelerate the planet formation process.

Future work on the formation of planetesimals, will focus on the mass distribution of these solid clumps and what their fate will be with the inclusion of additional physics, including but not limited to, radiative transfer, collisions and radial drift.

4.5 Discussion and Implications

The concentration of solids in non-fragmenting disks is well-documented (Gibbons et al., 2012; Shi et al., 2016), and the fragmenting simulations here show many similar features. For example, intermediate particles sizes $\mathrm{St} = 0.1, 1$ concentrate particularly well while smaller and larger particles are either too well coupled to the gas' smallest gas motions or barely influenced by the gas at all, respectively. This is apparent in the non-axisymmetric features of both 2D and 3D simulations, as both cases show most particles concentrating in these structures (Boley et al., 2010; Humphries & Nayakshin, 2018).

Vortices are known as potential particle traps which could lead to the high dust-to-gas ratios which enable the formation of planetesimals (Raettig et al., 2015). Even in the early, gravitationally unstable disk, vortices have shown the potential to efficiently collect material (Gibbons et al., 2015) however vortices may not persist long enough in vertically-stratified simulations (Lin & Pierens, 2018), affecting the ability of particles to form planetesimals or cores. In Figure 4.1, one can see that the particle-less fiducial simulations fragment and a strong cyclonic vortex spins at it's center. However, cyclonic vortices are not good dust traps because since they rotate in the same direction as the disk spins, they will increase the angular momentum of solid material which enter into the vortex, preventing settling.

Theoretical and observational work is coming to the agreement that planetary masses are increasingly rare if formed by gravitational instabilities. Observations have found candidates to be rare occurrences (Vigan et al., 2017) and simulations and analytic work have frequently had difficulty forming planetary masses at the right positions and masses and keeping them from migrating away (Baruteau et al., 2011; Kratter & Murray-Clay, 2011). These simulations show that initial fragments are already quite large and assuming that they will further accrete material, their masses could very likely increase into the brown dwarf regime. If they do not acquire much further mass the masses here are mostly within the range of many directly observed planets of a few Jupiter masses.

The analysis of Schlaufman (2018) into the relation between giant planets masses and the metallicity of their host stars finds that companions below a mass of $4\,M_{\mathrm{Jup}}$ are predominantly found around metal-rich stars while those larger than $10\,M_{\mathrm{Jup}}$ do not preferentially form around stars of any metallicity. Thus they find that there are two distinct populations of objects which are separated into planets and brown dwarfs/stars at roughly

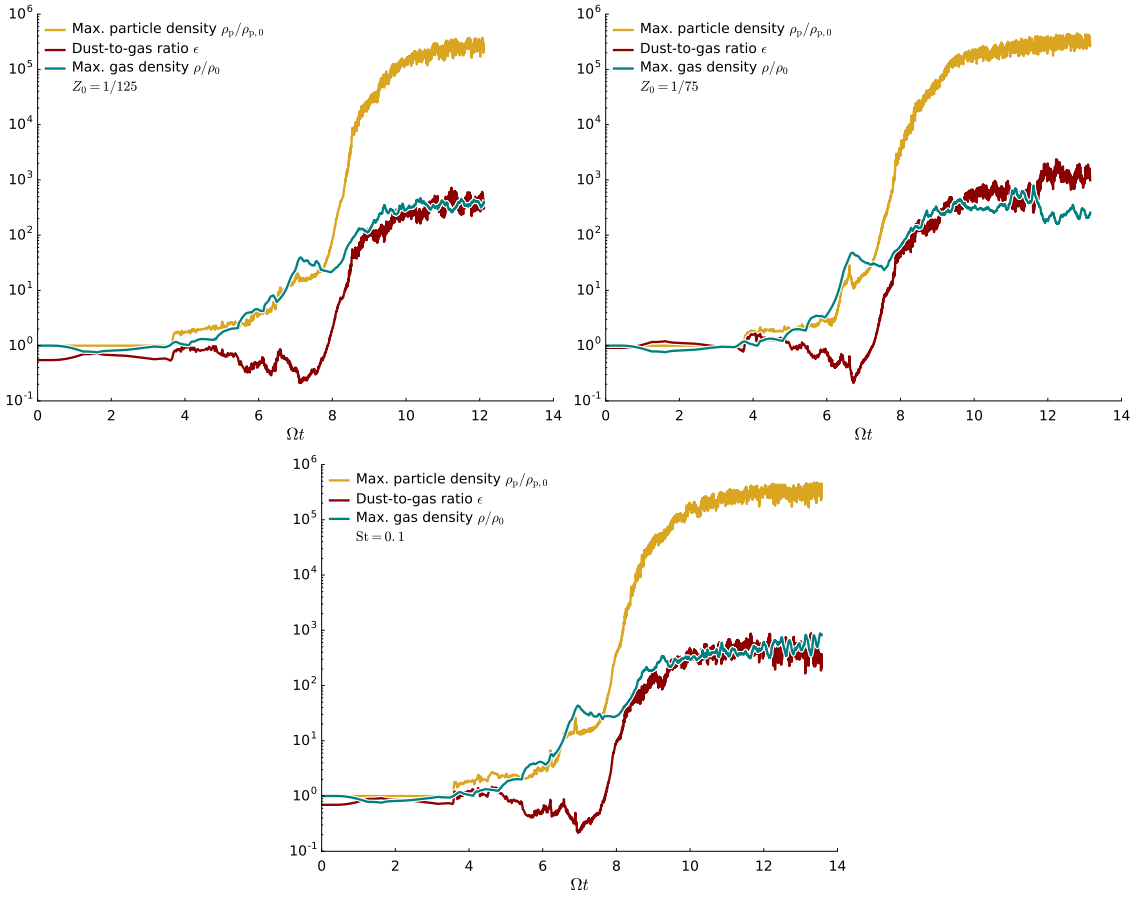


Figure 4.11: Same as Figure 4.6, but with one particle size (St = 0.1) and three different initial dust abundances: $Z_0 = 1/125$ (top), $Z_0 = 1/100$ (middle), $Z_0 = 1/75$ (bottom). Higher values Z_0 show faster particle concentration during the non-linear collapse of the gas leading to earlier formation of a massive core, but has little effect on whether or not a core actually forms.

$10\,M_{\rm Jup}$. While metallicity does not affect the ability of the simulations to fragment, it does appear to have an effect on the total mass with higher initial metallicities producing fragments with more massive gaseous components. In this respect no evidence is found of a separation between two populations from the initial fragment properties, but that excludes all later dynamics and evolution of the fragment.

In addition to the formation of a core and some satellites, I also notice the formation of a number of smaller solid bodies, particularly in the spiral arms, where particles tend to concentrate when not within the fragment. These objects become much more numerous in the case fragmentation does not occur but instead gravitoturbulence prevails and fragmentation of the dense particle layer only results in many smaller fragmentation events rather than a single blob. Planetesimals such as these may be the seed for a another fragment, contribute to the eventual formation of a terrestrial planet or may be quickly accreted into the fragment and will be the subject of subsequent investigations.

Limitations

Many gas-dust interactions have been neglected or simplified in these simulations, including grain heating and cooling, collisional destruction and aggregation, among other effects

(Nayakshin, 2010). Collision speeds are expected to be high and may prevent the growth of larger solids, but may be low enough within the spiral arms to allow for aggregation to occur. Additionally, grain growth alone may contribute to the overall Toomre instability of the disk and aid in it's eventual fragmentation (Sengupta et al., 2019).

Since the particle size is defined though a Stokes number, the effective size of a particle changes with the properties of the surrounding fluid, as illustrated in Figure 4.3. A particles' size in the Epstein regime scales linearly with the gas density (see Equation (4.8)) so as a particle moves into a high gas density region, its size will increase which may not necessarily reflect the true nature of the system. This would mean that particles which are originally on the order of centimeters in size would instead have an effective size of meters once inside a gas overdensity one hundred times the initial value. However, as they move through fluid regions of varying density, particles of a particular Stokes number will tend to grow to a mass which maintains their aerodynamic properties (Birnstiel et al., 2011).

Since the further collapse of a fragment is not modeled, which may significantly increase central temperatures over 1400 K and evaporate the material which constitutes the core (Helled & Schubert, 2008). Additionally, the grains within the clump would begin to dominate the opacity and likely reduce the effectiveness significantly if not completely, rendering the β-cooling used here lacking. One possible solution which remains computationally inexpensive is to used a modified cooling prescription which drastically reduces cooling efficiency in overdense regions (Baehr & Klahr, 2015). While an already-present core would survive the high temperatures, further growth by dust and pebbles would likely be halted. The long term fate of a core in these simulations will someday benefit from more extensive radiative transfer models which take into account the grain opacities and heating at the fragment centers which could lead to grain evaporation. It should be explicitly noted that the cooling prescription includes no metallicity dependence, which is particularly important due to the dominance of dust opacities. For this reason, results might change significantly for a treatment of cooling which takes this into account, especially for the particle dense central region.

These simulations also assume that the there is a uniform distribution of particle sizes, all with a dust-to-gas ratio of 1:100. This is most certainly not the case, and one might expect the smaller particles to be more common since there has been little opportunity for collisional growth and aggregation (Mathis et al., 1977). There is however evidence of some growth in the molecular cloud phase, but with a limit at micron-sized particles (Steinacker et al., 2015), but early gravitoturbulence and particle growth could lead to larger grain sizes (Sengupta et al., 2019). While larger $\mathrm{St} = 10$ particles do not collect particularly well within fragments, even $\mathrm{St} = 1$ particles may be far less common than smaller species, resulting in less solid accumulation within a fragment.

No radial pressure gradient is included, which would result in migration velocities, particularly for $\mathrm{St} = 1$ particles, and affect their ability to concentrate at any one location. While these simulations may appropriately describe the local concentration of particles, the absence of a pressure gradient and thus particle drift means that particle concentrations are likely overestimated as particles around $\mathrm{St} = 1$ would migrate quickly away from the gravitationally unstable regions of interest. The spiral arms induced by gravitational collapse are however significant pressure maxima which decrease the particle velocity dispersions and facilitate growth (Booth & Clarke, 2016). Shearing boxes of this size are large enough to have gradients corresponding to global disk properties, but since these simulations are not intended to explore the longer term nature of the fragments that form within, the shearing box approximation is suitable to model the local collapse and initial conditions of a fragment.

As fragments migrate they may accrete more material, whether it be gaseous or solid, after formation which cannot be captured in such a local model and will affect the overall metallicity (Mercer & Stamatellos, 2017; Ilee et al., 2017). A process similar in principle to pebble accretion (Ormel & Klahr, 2010) could result in later metallicity enhancements (Humphries & Nayakshin, 2018). Fragments are also prone to rapid migration (Baruteau & Masset, 2008) which may take them to new regions of the disk which may be gas or dust rich and accretion will alter their composition which cannot be captured in this study. What is presented in this paper represents a look at the possible initial conditions of gas giant planets formed by GI and their later evolution is a subject of further study.

I have performed 3D hydrodynamic simulations of gravitationally unstable gaseous disks using high-resolution finite-difference simulations. Using Lagrangian superparticles within the Eulerian mesh, I modeled the evolution of solids of various sizes in simulations which fragment into dense gas clumps. These fragments may be gravitational and hydrodynamic traps for particles which can then concentrate into planetesimals or enrich the metallicity of gas giant planets. I summarize my findings in the following points:

1. Fragmentation of the gas disk concentrates particles of intermediate sizes initially through aerodynamics, but once concentrated enough they may be comparable gravitationally to the gas fragments themselves. Thus, even though the dust is not initially gravitationally unstable there is still potential for significant solid accumulations to form in fragmenting gravitationally unstable disks, potentially resulting in the growth of solid bodies.

2. Once collected within gas overdensities through aerodynamic drag, solids can concentrate to the point where their contribution to the total self-gravitational potential is non-negligible and thus form considerable core within proto-gas giants. Similar to the results of Boley & Durisen (2010), I find that for gas fragments up to several Jupiter masses, with central solid cores typically of a few tens of $M_{\oplus}$.

3. I find that intermediate particle sizes $\mathrm{St} = 0.1 - 1$, fragments form early with already sizable cores before they have core temperatures in excess of $\sim$1400 K which would sublimate solid material. This would mean that sedimentation after formation may not be necessary to explain core in planets formed by gravitational instability.

4. I find fragments in total to have solid fractions are around $1.5 - 5\%$, similar to the amounts suggested by core accretion and by observations of Jupiter and Saturn. This suggests that core formation and enhanced metallicity are not unique to core accretion and more information is required of a planet to determine formation mechanism.

5. Fragments have atmospheres with metallicities above the initial value $Z_0 = 1/100$ in the case when a core does not form, but below initial metallicities when a core is formed. Whether or not a core forms may have an impact on the atmospheric metallicity of the resulting gas giant planet.

5 High Resolution Gravitationally Unstable Full Disk Simulations

Astrophysical problems can often span many orders of magnitude in spatial and temporal scales, making an accurate and efficient simulation challenging to implement. Grid codes in particular suffer from the necessity of approximating the fluid variables at all points in the domain, whether they are pertinent to the problem or not. Adaptive mesh refinement (AMR) methods allow for a grid code to overcome this disadvantage to some degree, focusing most of the fine resolution to regions which require it and leaving other regions more sparsely resolved (Berger & Oliger, 1984; Berger & Colella, 1989). With this approach, grid codes can compare well with smoothed particle hydrodynamic (SPH) codes which are generally better suited for simulations with significant density disparities (Kim et al., 2016).

AMR strategies start with a coarse grid over the entire domain and at each time step, selects regions with high gradients, high values or high errors to regrid with higher resolution. This is often desired because regions with these features often require higher resolution to adequately resolve the steepness of the gradients. This can be done recursively, with refined grids on top of refined grids giving improved spatial and temporal resolution of especially dynamic regions. This allows a code to focus computational power on the regions which most affect the solution and less effort goes into the areas of the simulation which do not change much over long spatial and temporal scales. Assuming a cell is refined by splitting it down the middle in each direction, the finest level n is advanced in time first, with a timestep $\Delta t_n = \Delta t_0 2^{-n}$ before the next finest level is advanced. This means that only a fraction of the total domain is advanced at high spatial and temporal resolution and the rest of the simulation is advanced with the default time increment. In this way only the regions with high density or velocity gradients that need better spatial and temporal resolution receive it.

Self-gravitational systems are common in astrophysical contexts, but can often be ignored when another potential dominates on the necessary scales, such as the central potential of a star in a circumstellar disk, or when the gravitational potential is roughly constant with time. There are still a large number of situations where calculating the effects of self-gravity are not only useful but necessary. For example, the collapse of a molecular cloud into a star requires one to model a rapidly changing gravitational potential as the center becomes more dense, eventually forming a protostar with surrounding disk. The disk itself will likely be considered gravitationally unstable, so even when the protostar provides a constant potential, computing the self-gravity of the system is still necessary.

An elliptic equation is a specific form of the second-order partial differential equation where one or more spatial second-order partial derivatives are non-zero, but all other

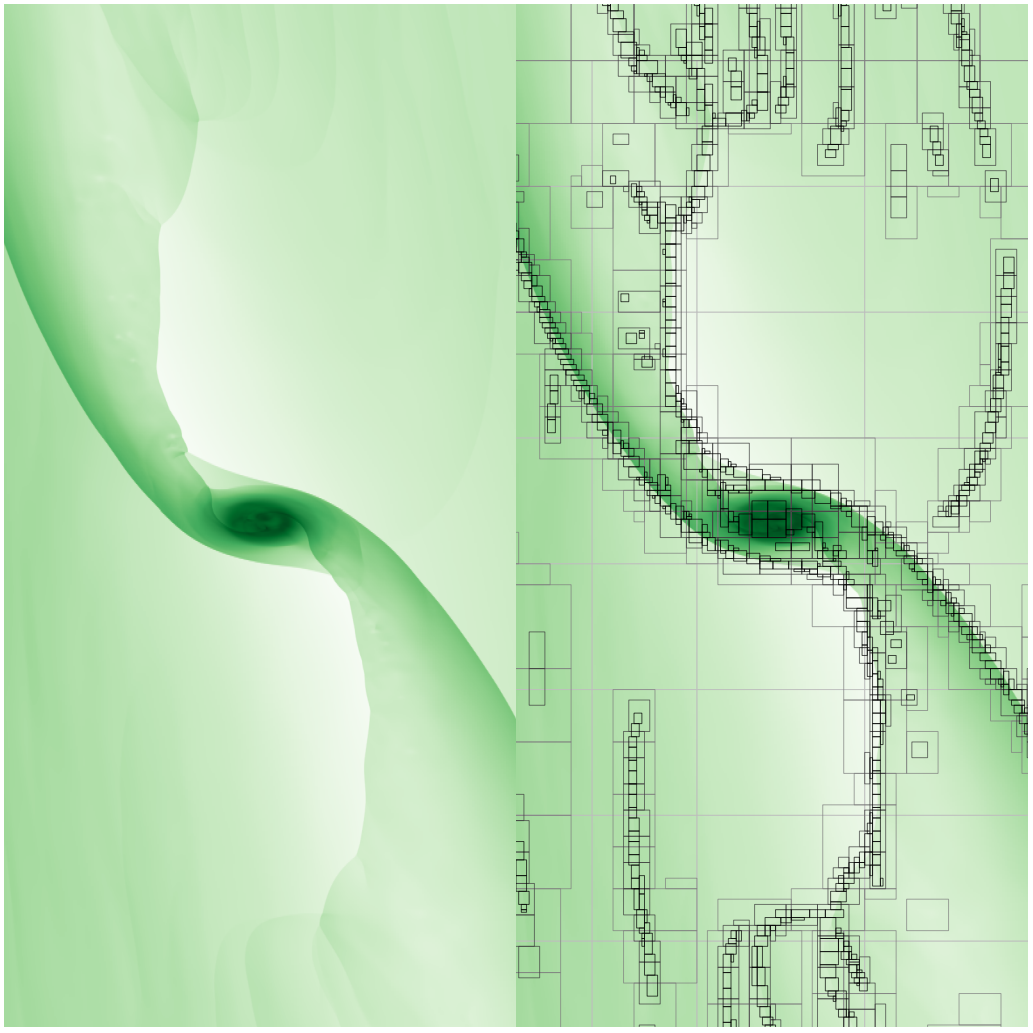


Figure 5.1: Visualization of an example problem without self-gravity using adaptive mesh refinement. Depicted is the vertically integrated mass density of a planet potential embedded in disk and the spiral wakes which are generated by the presence of the planet. Both figures are zoomed in at the position of the planet to show the locations of structure and associated grid refinement.

derivatives including time derivatives vanish resulting in an equation of the form

$$(\alpha\mathbb{I} + \beta\nabla^2)\phi = \rho, \tag{5.1}$$

for some scalar field ϕ, identity matrix $\mathbb{I}$ and arbitrary constants α and β.

Elliptic problems in astrophysics come in a variety of forms, but the most common is the self-gravitational potential of a system. For a non-relativistic problem this is defined by the Poisson equation

$$\nabla^2\phi = 4\pi G\rho, \tag{5.2}$$

where ρ is the underlying density distribution, G is the gravitational constant, and ϕ is the desired self-gravitational potential. For any non-trivial density distribution $\rho(x, y, z)$ the solution can only be found using a numerical method. However, boundary conditions typically do not cause significant issues, as only constant Dirichlet boundaries are necessary for a smooth solution of ϕ.

Once a solution for the potential is found, the equation of motion is modified to include the gravitational acceleration due to the self-gravitational potential

$$g = -\nabla\phi. \tag{5.3}$$

There are several ways to approach an elliptic problem numerically, from slow but effective relaxation methods to Fourier transforms to more complex methods dependent on the problem and method at hand (Schmitz & Ying, 2014). For AMR grids, elliptical solvers generally come in two forms, tree methods (Barnes & Hut, 1986) and multigrid methods (Brandt, 1973).

Multigrid methods take advantage of the fact that relaxation methods converge at different rates for different scales. At a fixed, uniform resolution the the solution to a relaxation problem with a Jacobi-Seidel method will converge slowly for the moderate scales while the smallest will converge faster. This is because the smoothing by a Gauss-Seidel method only considers adjacent points when averaging, which quickly damps out small perturbations, but takes much longer for large scale perturbations to smooth out. A multigrid method therefore maps or restricts the solution on a fine grid to a coarser grid and solves another relaxation problem. This solution will suffer the same converge limitations as the finer grid, but since the resolution has changed, convergence of the solution will be attained at different scales. This process is repeated for finer and finer grids until the solution is within a prescribed threshold over the entire domain (LeVeque, 2007).

The method employed here further uses a red-black ordering scheme when smoothing with Gauss-Seidel. This figuratively assigns all points a color, red or black, in a checkerboard pattern, and smooths all the red points first and before continuing with the black points. This is useful for speeding the convergence rate of the solution, particularly when done in parallel, since an update of all red points does not require any information from any of the other red points, but rather the black points, which are unchanged during the update. The same is true in reverse for the black points.

The solution on the coarsest level is then interpolated or prolongated up to the highest up to the higher resolutions and corrected again at each level by a relaxation method until the solution is reconstructed for all resolutions. The process of cycling through the resolution levels can be managed in many ways, depending on the nature of the grids and the desired efficiency to reach a small residual. The simplest goes down to the coarsest grid and all the way back up and is therefore called a V-cycle. Additional sub-cycles within the V-cycle gives rise to more complicated W-cycle and F-cycle strategies, but unless otherwise stated, it can be assumed in this thesis that I am referring to multigrid methods that V-cycle only.

Since AMR methods already make use of grids of increasing resolution to model the problem, a multigrid method can make use of the same refinement hierarchy to solve the elliptic problem, saving the overhead of defining and creating the grids necessary for the multigrid solver. Ideally however, the solver will be flexible enough to allow the user to define the grids for the solver independent of the grids defined by the PLUTO code.

An OctTree method for solving the self-gravity potential is often used for a system of particles which is defined on a grid recursively divided into smaller and smaller constituent grids. To calculate the potential, one combines two independent approaches into an algorithm which reduces the N^2 scaling of a system of N objects interacting with each other to $N\log(N)$ which drastically speeds up calculations of tens to hundreds of millions of particles. This combines multipole expansion, which means to reduce multiple particle-particle interactions to a single interaction, and hierarchical subdivision of the domain,

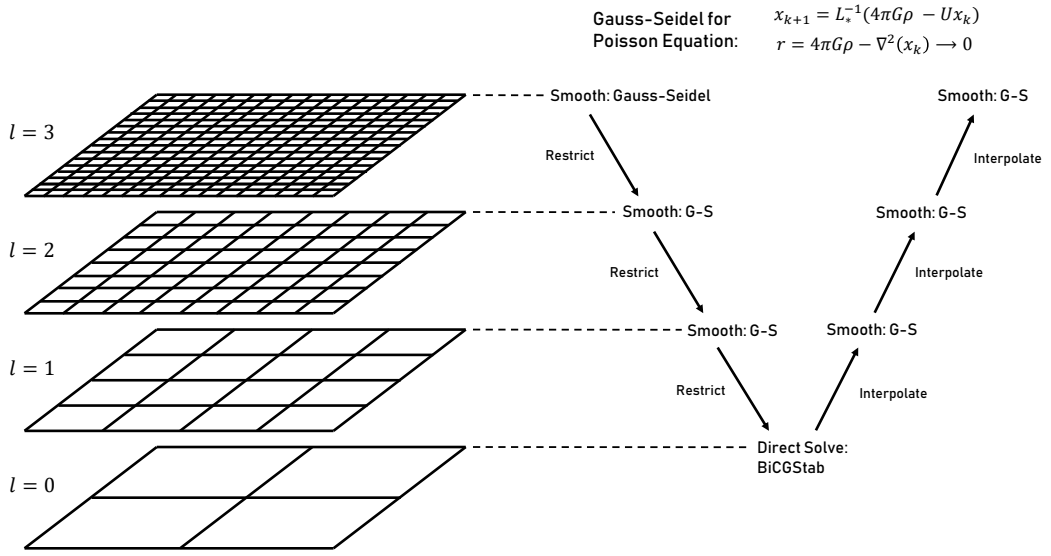


Figure 5.2: An illustration of the multigrid V-cycle. Starting from the finest grid at the top left, the solution is smoothed with a Gauss-Seidel relaxer, before being restricted to a coarser grid and smoothed again. This process is repeated until reaching the bottom level, where an exact solution is solved for with a Bi-Conjugate Gradient Stabilized method. This coarse solution is then interpolated up to the finer levels where it is smoothed again before reaching the finest grid and the end of the V-cycle. A more technical description can be found in the report of Martin & Cartwright (1996)

.

which provides the structure by which the code goes through the particles and calculates distances and centers of mass (Barnes & Hut, 1986). This method is ideal for domains with irregular or embedded boundaries where the adaptive refinement is necessary to conform to the peculiarities of the system.

This paper will detail my implementation of the multigrid elliptic solver in PLUTO, starting with Section 5.1 where I describe the code and how the Chombo elliptic solver fits into that framework. In Section 5.2, I detail the specific implementation of the multigrid solver and I continue with preliminary results and concluding remarks in Section 5.4.

5.1 The PLUTO code

The PLUTO code[1] is a Godunov code written in C for solving the conservative (magneto)hydrodynamic equations with a finite volume scheme. PLUTO is highly modular, suitable for astrophysical problems in Cartesian, polar and spherical geometries as well as with varied physical processes and parallelized for high resolution and performance. For further details on the method and implementation see the Appendix Section B.3 and Mignone et al. (2007). In the static grid implementation, PLUTO supports multiple reconstruction schemes and Riemann solvers, which are all available with AMR.

Making use of the Chombo libraries[2], PLUTO has block-structured adaptive mesh capabilities, including thermal conduction and limited MHD (Mignone et al., 2012; Adams

[1] http://plutocode.ph.unito.it/
[2] https://commons.lbl.gov/display/chombo/

et al., 2015). Chombo uses FORTRAN array boxes (FABs) to handle arithmetic operations on the data arrays within the grids or patches on each level and uses C++ to manage all higher level operations and functions. An overview and comparison of the different available block-structured adaptive mesh refinement frameworks can be found in Dubey et al. (2014). Chombo comes with a library of useful classes for solving generic elliptic problem in an adaptive mesh scheme and I drew upon the available materials in the Chombo documentation and example problems to implement a multigrid elliptic solver that is compatible with PLUTO.

Currently the AMR scheme in PLUTO is compatible with non-constrained transport MHD and thermal conduction

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0 \tag{5.4}$$

$$\frac{\partial (\rho \boldsymbol{v})}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}\boldsymbol{v} - \boldsymbol{B}\boldsymbol{B}) + \nabla P_t = -\rho \nabla \phi \tag{5.5}$$

$$\frac{\partial \epsilon}{\partial t} + \nabla \cdot ((\epsilon + P)\mathbf{v} - (\boldsymbol{v} \cdot \boldsymbol{B})\boldsymbol{B}) = \nabla \cdot \Pi_\epsilon - \Lambda + \rho \boldsymbol{v} \cdot \nabla \phi, \tag{5.6}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} - \nabla \times (\boldsymbol{v} \times \boldsymbol{B}) = -\nabla \times (\eta \boldsymbol{J}) \tag{5.7}$$

with a number of explicit integrators but additional work remains to implement more advanced features, such as super time-stepping (i.e. Meyer et al. (2012)) and radiative transport. More details about the PLUTO, Chombo and the methods therein can be found in Appendix Section B.3.

5.2 A Multigrid Elliptic Solver

As described above, the V-cycle of the multigrid solver starts at the finest level and the solution is smoothed to the solution four times with a Gauss-Seidel relaxation according to the algorithm detailed in Martin & Cartwright (1996). For a system described by $A\boldsymbol{x} = \boldsymbol{b}$, Gauss-Seidel iteratively solves the equation

$$L_* \boldsymbol{x}_{k+1} = \boldsymbol{b} - U \boldsymbol{x}_k \tag{5.8}$$

for $\boldsymbol{x}_{k+1}$ starting with an initial guess $\boldsymbol{x}_0$. The matrix A has been decomposed into a lower triangular component L_* and strictly upper triangular component U. This approximate solution is compared to the exact solution through the residual

$$r \equiv 4\pi G\rho - \mathfrak{L}(\boldsymbol{x}_{k+1}) \tag{5.9}$$

for Laplacian operator $\mathfrak{L}$ and considered converged when r is less than an acceptable threshold ϵ. This fine solution is then mapped to the next coarsest level in a process called restriction, but is in practice just averages the value over multiple cells into a single cell. At each level of the V-cycle excluding the bottom/coarsest level, the residual is calculated using Gauss-Seidel relaxation according to the algorithm mentioned above and detailed in Martin & Cartwright (1996).

For the relaxation problem on the coarsest grid, a Bi-Conjugate Gradient Stabilized (BiCGStab) method is used to get the exact solution. BiCGStab is a direct solver for systems of linear equations, suitable for sparse, non-symmetric matrices A. An adaptation of the more general, but limited Conjugate Gradient (CG) and Bi-Conjugate Gradient

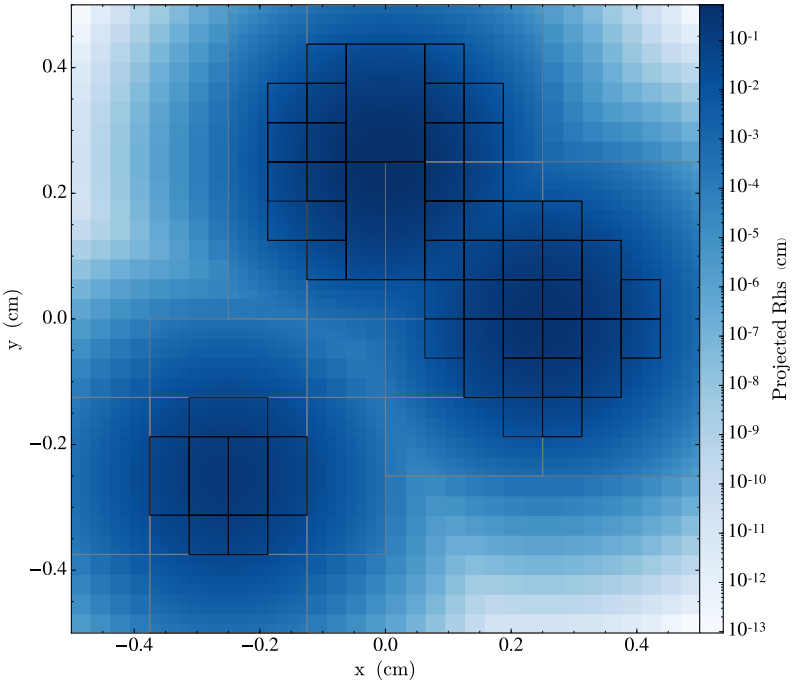


Figure 5.3: Test of the example elliptic solver available in Chombo showing the first and second refinement levels in gray and black boxes respectively. The domain is cubic, with aribitary units and the purely demonstrative right-hand side distribution has been integrated in the line-of-sight direction. The refinement criterion is based off the derivative of the quantity, and so the centers of the Gaussian blobs plotted here may not be fully refined, since the number of cells per level is limited, even if the neighboring slopes are well-refined.

(BCG) methods, BiCGStab uses a residual vector

$$\boldsymbol{r}_{k+1} = (I - \omega_k A)(\boldsymbol{r}_k - \alpha_k A \boldsymbol{p}_k) \tag{5.10}$$

and searching direction vector

$$\boldsymbol{p}_{k+1} = \boldsymbol{r}_{k+1} + \beta_k (1 - \omega_k A)\boldsymbol{p}_k \tag{5.11}$$

to calculate the solution, where $\alpha_k = \boldsymbol{r_k} \cdot \boldsymbol{r}'_0 / (A\boldsymbol{p_k}) \cdot \boldsymbol{r}'_0$, $\beta_k = \alpha_k \bar{\rho}_{k+1} / \omega_k \bar{\rho}_k$ and $\bar{\rho}_k = \boldsymbol{r}_k \cdot \boldsymbol{r}'_0$. The initial choice $\boldsymbol{r}'_0$ need only be chosen such that $\bar{\rho}_1 \neq 0$. The stabilizing coefficient ω_k, which keeps the solution bounded, is

$$\omega_k = \frac{(A\boldsymbol{s}_k) \cdot \boldsymbol{s}_k}{(A\boldsymbol{s}_k) \cdot (A\boldsymbol{s}_k)} \tag{5.12}$$

for $\boldsymbol{s}_k = \boldsymbol{r}_k - \alpha_k A \boldsymbol{p}_k$ to minimize the residual (5.10) in an L_2-norm. One could also include the conjugate vectors in terms of A^T, the transpose of A, however, by the formulation of the formulation of the coefficients α_k and β_k, it is only necessary for one residual and searching direction to be calculated by the algorithm.

With an exact solution on the entirety of the coarse domain through the BiCGStab solver, this solution is then prolongated or interpolated to the finer levels and smoothed again with Gauss-Seidel up to the finest level. At this point the residual is compared to the solution tolerance, typically on the order of 10^{-8} to 10^{-10}. If the residual is less than

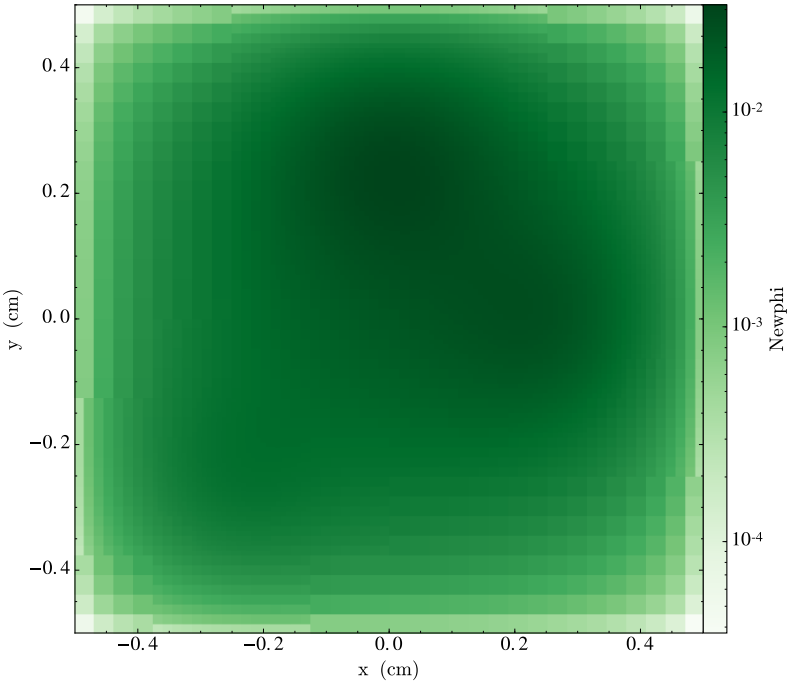


Figure 5.4: A slice of the potential generated by the multigrid elliptic solver. Due to the boundary conditions the solution does not converge at the coarse-fine interface, but only on the boundary grid cells.

the tolerance at all points, the multigrid solver stops and returns the solution array. If the residual is anywhere greater than the tolerance, the multigrid solver repeats the process with another V-cycle.

This has the advantage of converging in much fewer iterations than other relaxation methods.

A similar implementation of a multigrid self-gravity solver in the RAMSES code is detailed in Guillet & Teyssier (2011). An OctTree method for FLASH can be found in Wünsch et al. (2018).

Boundary Conditions

Boundary conditions for elliptic problems are typically straightforward, and as long as they are constant there will be a smooth solution over the domain Ω. This implementation currently supports constant Dirichlet boundary conditions, i.e.

$$\nabla^2\phi(\partial\Omega) = k, \tag{5.13}$$

for constant k on the boundary $\partial\Omega$, and periodic boundary conditions, such that the ghost zones around the boundary are filled like

$$\nabla^2\phi(x) = \nabla^2\phi(x + L) \tag{5.14}$$

for domain length L in any given coordinate direction. More sophisticated and useful boundary conditions will be added in the future. Naturally, the boundary conditions of the solver can be set independently of the hydrodynamic boundary conditions established

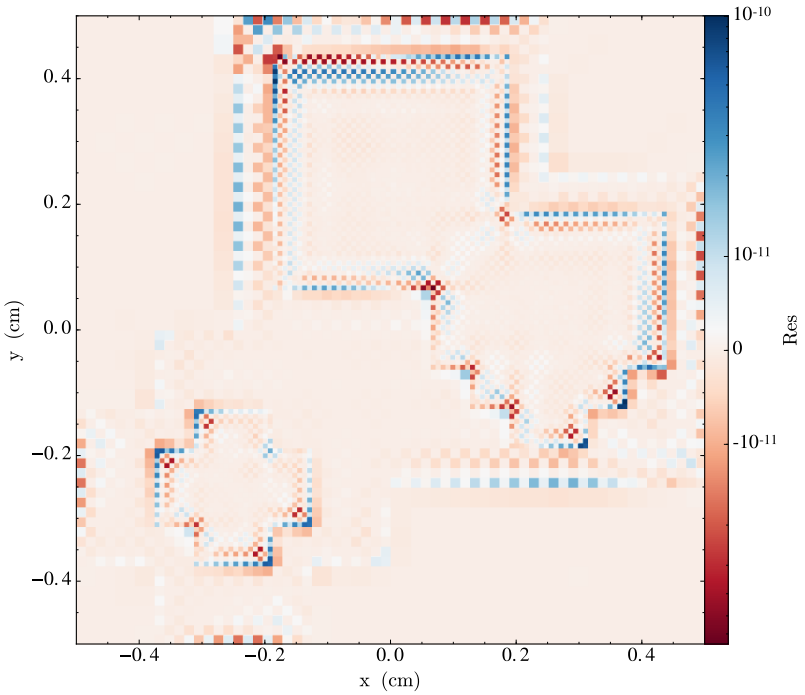


Figure 5.5: Slice of the residual of the solution ϕ. While the solution is well within the 10^{-10} threshold over the entire domain of the problem, the pattern of the residual is interesting. Clearly there is a larger discrepancy between the solution and right-hand side distribution near the coarse-fine interfaces, particularly when the interface is a right angle. The apparent checkerboard pattern is likely an artifact of the redblack smoothing scheme as part of the Gauss-Seidel relaxation.

on the coarse domain.

Input Grids

The simulation of complicated domains may require complicated unions of nested grids and by default the grids used by the multigrid solver are the refined grids of the computation. While the grids are complex when viewed in their entirety, thanks to the way the hydrodynamics is updated by the PLUTO code, the evaluation of the potential is simplified dramatically. The multigrid solution is calculated all at once, but the potential is applied to the equations of motion one patch at a time, starting from the patches at the finest level and cycling through all the patches on all levels. On each patch, a new uniform grid problem is solved and at this point the gradient of the potential is evaluated. Thus there are no coarse-fine interfaces to worry about when adding the self-gravitational acceleration to the hydrodynamics.

5.3 Solver Implementation

The Chombo Elliptic Solver

Below I include the important portions of code related to the Chombo multigrid ellitpic solver. The majority of the code in this section is modified from the pre-existing sample

elliptic solver provided in the Chombo example problems. The code in it's entirety can be found at my Github page or in a future release of PLUTO.

First one must look at the solver itself. Within the class file `AMRPoissonPluto.cpp`, the first function defines the input parameters which are necessary to set the grid and level hierarchies which the solver needs. Quantities which are defined as `Vector<>` templates are those which are multilevel parameters and the index of the Vector yields the quantity or class at that level. Of the classes which are contained with the Vectors, of special note is `LevelData`, which contains/stores multiple attributes pertaining to the particular level as they relate to the collection of `FArrayBox` objects on that level. Thus, each `FArrayBox` corresponds to a patch on that level, including the data, stored on a number of intervals equal to the number of variables, the location and size of the patch. Then, all of the patches on a level are within the `LevelData` and all the levels are put into their proper location in the `Vector<>`.

All parameters are predetermined by the existing PLUTO code and need only be temporarily stored for later use by the solver. The number of adaptive mesh levels and thus the number of grids over which the multigrid method cycles is determined by `a_numLevels` and the ratio of refinement, or how many subdivisions of a single cell are made from one level to the next, is given by `a_ref_ratio`. The physical grid spacing and total domain of each level are input through `a_dx` and `a_domain`, respectively. The array of `m_grids` defines the layout of the patches or blocks within each level, important for defining the level space upon which data must be written. Finally, the density, the quantity from which the multigrid solver will calculate the potential is stored in `a_rhs`.

```
// Define this object and the boundary condition object
void AMRPoissonPluto::define(Vector<LevelData<FArrayBox>* >   a_rhs,
                             Vector<DisjointBoxLayout>        a_grids,
                             Vector<ProblemDomain>            a_domain,
                             Vector<Real>                     a_dx,
                             Vector<int>                      a_ref_ratio,
                             int                              a_numLevels)
{
  // Store the level data to be used later
  m_rhs       = a_rhs;
  m_grids     = a_grids;
  m_domain    = a_domain;
  m_dx        = a_dx;
  m_numLevels = a_numLevels;
  m_ref_ratio = a_ref_ratio;
  m_isDefined = true;
}
```

Listing 5.1: The parameters necessary to define the solver

Following some minor functions which handle the periodic or Dirichlet boundary conditions, comes the creation of the multigrid solver object `a_amrPoissonSolver`. It is created as a Poisson solver (as opposed to a generic elliptic solver) through the constants `alpha` and `beta` (where $\beta = 1/4\pi$ is the right hand side constant for the Poisson equation assuming $G = 1$) and defined starting on the coarsest level denoted by index 0. `CH_TIME` is a macro which times the function for diagnostic purposes.

The multigrid solver parameters regulate the error on each level, where `numSmooth` is the number of Gauss-Seidel relaxation iterations performed on each level and `numMG` determines the type of multigrid cycle, with `1` yields a V-cycle (default) and `2` performs a W-cycle. While a W-cycle spends more time on the coarse grids where smoothing is naturally more efficient and may aid convergence of a solution, it is typically unnecessary. The solver will run for a maximum of 100 iterations or until it with the tolerance threshold set by `eps` or if the norm of the solution does not change by more than `hang` from one iteration to the

next.

```
void AMRPoissonPluto::setupSolver(AMRMultiGrid<LevelData<FArrayBox> > *
    a_amrPoissonSolver,
                                  LinearSolver<LevelData<FArrayBox> >&
    a_bottomSolver,
                                  Vector<DisjointBoxLayout>&          a_grids,
                                  Vector<ProblemDomain>&              a_domain,
                                  Vector<int>&                        a_ref_ratio,
                                  Vector<Real>&                       a_dx,
                                  int                                 a_level)
{
   CH_TIME("setupSolver");

   AMRPoissonOpFactory opFactory;

   // solving poisson problem here
   Real alpha = 0.0;
   Real beta = 1.0/(4*3.14159265);

   opFactory.define(a_domain[0],
                    a_grids,
                    a_ref_ratio,
                    a_dx[0],
                    &ParseBC, alpha, beta);

   AMRLevelOpFactory<LevelData<FArrayBox> >& castFact =
            (AMRLevelOpFactory<LevelData<FArrayBox> >& ) opFactory;

   a_amrPoissonSolver->define(a_domain[0], castFact,
                       &a_bottomSolver, a_level);

   // Multigrid solver parameters
   int numSmooth = 4;
   int numMG = 1;
   int maxIter = 100;
   Real eps = 1.0e-9;
   Real hang = 1.0e-10;

   Real normThresh = 1.0e-30;
   a_amrPoissonSolver->setSolverParameters(numSmooth, numSmooth, numSmooth,
                                numMG, maxIter, eps, hang, normThresh);
}
```

Listing 5.2: The function which sets up the solver object and defines its parameters

After the inputs and the solver have been defined, all that is left is to run the solver. The code which follows shows the creation of the multilevel data handlers `m_phi` to store the final solution and `resid` to store the residual. Then the interlevel multigrid solver `amrPoissonSolver` is initialized along with the `bottomSolver` on the coarsest level and the call to the solver setup function.

```
int AMRPoissonPluto::runSolver()
{
   CH_TIME("runSolver");

   int status = 0, mg_type = 0;

   //int s_verbosity = 4;

   // allocate solution and RHS, initialize RHS
   int numLevels = grids.size();
   Vector<LevelData<FArrayBox>* > phi(numLevels, NULL);
   // this is for convenience
   Vector<LevelData<FArrayBox>* > resid(numLevels, NULL);

   for (int lev=0; lev<=numLevels-1; lev++)
     {
       const DisjointBoxLayout& levelGrids = m_grids[lev];
```

```
18        phi[lev] = new LevelData<FArrayBox>(levelGrids, 1, IntVect::Unit);
19        resid[lev] = new LevelData<FArrayBox>(levelGrids, 1, IntVect::Zero);
20      }
21
22    // initialize solver
23    AMRMultiGrid<LevelData<FArrayBox> > *amrPoissonSolver;
24    if ( mg_type==0 )
25      {
26        amrPoissonSolver = new AMRMultiGrid<LevelData<FArrayBox> >();
27      }
28    else
29      {
30        MayDay::Error("FAS not supported");
31      }
32
33    BiCGStabSolver<LevelData<FArrayBox> > bottomSolver;
34    //bottomSolver.m_verbosity = s_verbosity-2;
35    setupSolver(amrPoissonSolver, bottomSolver, m_grids, m_domain,
36                m_ref_ratio, m_dx, m_numLevels);
37
38    // do solve
39    int iterations = 1;
40
41    for (int iiter = 0; iiter < iterations; iiter++)
42      {
43        bool zeroInitialGuess = true;
44        pout() << "about to go into solve" << endl;
45        amrPoissonSolver->solve(phi, m_rhs, numLevels-1, 0, zeroInitialGuess);
46        pout() << "done solve" << endl;
47      }
48
49 ...
```

Listing 5.3: The function which executes the solver and the multigrid solution to the potential phi

Everything which follows within the function `runSolver` pertains to writing the output to an HDF5 file, which can be activated or deactivated easily.

Solver Interface with PLUTO

The solver introduced in the previous section need to be integrated into the PLUTO code where it can take in the appropriate density and grid input. This was made difficult by the fact that the way the PLUTO code is currently designed, grid, domain and density information for all levels are stored separately even though the solver requires all levels at once in `Vector<>` containers, where the index is the number of the level. This was accomplished through the `getAMRHierarchy` function which retrieves the data for all levels in `AMRLevel` objects, from which the desired attributes of each level can be extracted. The conditional `if (!m_hasCoarser)` is used to execute the solver only on the coarsest level, i.e. the level that does not have a coarser level. The solver is then run, the solution is stored in it's multilevel container and all dynamically allocated memory is deleted.

```
1  #if SELFGRAV
2    if (!m_hasCoarser)
3      {
4        Vector<AMRLevel*> onTheLev = AMRLevel::getAMRLevelHierarchy();
5        int numLevels = onTheLev.size();
6
7        // Create Vector containers for the multilevel input to the solvers
8        m_phi.resize(numLevels, NULL);
9        Vector<LevelData<FArrayBox>* >              rhs(numLevels);
10       Vector<LevelData<FArrayBox>* >              temp_rhs(numLevels);
11       Vector<DisjointBoxLayout>                   grids(numLevels);
12       Vector<ProblemDomain>                       domain(numLevels);
13       Vector<Real>                                dx(numLevels);
```

```
14      Vector<int>                                     ref_ratio(numLevels);
15
16      // Put all the level metadata into the containers at he appropriate level
17      for (int lev=0; lev<numLevels; lev++)
18        {
19          AMRLevelPluto* amrPlutoLevel = dynamic_cast<AMRLevelPluto*>(onTheLev(lev)
    );
20          grids[lev]           = amrPlutoLevel->m_grids;
21          domain[lev]          = amrPlutoLevel->m_problem_domain;
22          dx[lev]              = amrPlutoLevel->m_dx;
23          ref_ratio[lev]       = amrPlutoLevel->m_ref_ratio;
24          temp_rhs[lev]        = *(amrPlutoLevel->m_UNew);
25
26          const Interval densityInterval(0,0);
27          rhs[lev]             = new LevelData<FArrayBox>(grids[lev], 1, IntVect::
    Zero);
28          temp_rhs[lev]->copyTo(densityInterval, *rhs[lev], densityInterval);
29          //delete amrPlutoLevel;
30        }
31      // Set up Poisson solver objects
32      AMRPoissonPluto amdSelfGravSolver;
33      amrSelfGravSolver.define(rhs,
34                               grids,
35                               domain,
36                               dx,
37                               ref_ratio,
38                               numLevels);
39
40      m_phi = amrSelfGravSolver.runSolver();
41
42      // deallocate memory
43      for (int lev=0; lev<=numLevels-1; lev++)
44        {
45          delete rhs[lev];
46        }
47    }
48  #endif
49 }
```

Listing 5.4: The collection of the multilevel data for the solver the creation of the solver object and the solver execution

The data is then passed to each level via the `m_levelPluto.step` function

```
1  #if (TIME_STEPPING == RK2)
2    g_intStage = 2;
3    Real DtCool; // The predictor returns the advective/diffusive timestep
4                 // The corrector returns the cooling timestep
5
6    DtCool = m_levelPluto.step(m_UNew,
7                               flux,
8                               *finerFR,
9                               *coarserFR,
10                              m_split_tags,
11                              *m_phi[m_level],
12                              *coarserDataOld,
13                              tCoarserOld,
14                              *coarserDataNew,
15                              tCoarserNew,
16                              m_time+m_dt,
17                              m_dt,
18                              m_cfl);
19
20   #if (COOLING != NO)
21    newDt = Min(newDt,DtCool);
22   #endif
23 #endif
```

Listing 5.5: Passing the potential to the level integrator

where it is broken down into the `FArrayBox` objects which constitute each patch or block.

```
// Beginning of loop through patches/grids.
for (DataIterator dit = m_grids.dataIterator(); dit.ok(); ++dit){
  CH_START(timeUpdate);

  // The current box
  Box curBox = m_grids.get(dit());

  // The current grid of conserved variables
  FArrayBox& curU = m_U[dit];

  ...

  // Store self-gravity potential to patches
  #if SELFGRAV
    FArrayBox& patchPhi = a_levPhi[dit];
  #else
    FArrayBox patchPhi;
  #endif

  ...

  // Take one step
  m_patchPluto->advanceStep (curU, curUtmp, curdV, split_tags, flags, patchPhi,
  flux,
                             &Dts, curBox, grid);

  ...

  }
```

Listing 5.6: Passing the potential to the patch integrator

Within the `advanceStep` function of the patch integrator, the potential is then added to the main data structure `d`, making sure allocate the memory. The potential on each patch is then passed to PLUTO, where as a simple uniform mesh, the gradient of the potential can be computed through normal pre-existing routines.

```
#if SELFGRAV
    d.Phi = ARRAY_3D(NX3_TOT, NX2_TOT, NX1_TOT, double);
#endif

...

d.Phi = ArrayMap(NX3_TOT, NX2_TOT, NX1_TOT,a_patchPhi.dataPtr(0));
```

Listing 5.7: Allocating memory for the potential and adding it to the main data structure

Adding Self-Gravity

Once within the primary data structure `d` as `d.Phi` or `d->Phi`, the simulation is solved one refined grid at a time. This means that every patch is just a much smaller PLUTO problem with a uniform mesh. The potential is then added to the right-hand side (RHS) of the momentum conservation equation (5.5) via `Src/MHD/rhs.c`.

```
#if SELFGRAV
       rhs[i][MX1]    -= dtdx*vg[RHO]*(d->Phi[k][j][i] - d->Phi[k][j][i-1]);
       IF_ENERGY(phi_e = d->Phi[k][j][i];
                 rhs[i][ENG] -= phi_e*rhs[i][RHO];)
#endif
```

Listing 5.8: Adding the potential as a source term to the RHS of the momentum and energy conservation equations

and energy conservation equation (5.6) via `Src/MHD/rhs_source.c`

```
#if (SELFGRAV & HAVE_ENERGY)
flux[i][ENG] += flux[i][RHO]*d->Phi[k][j][i];
#endif
```

Listing 5.9: Adding the potential as a flux to the energy equation

In both cases only the first coordinate is shown for the sake of brevity and repeated in a similar fashion for the remaining orthogonal coordinates without alteration.

5.4 Discussion

I have implemented a multigrid elliptic solver for the PLUTO code to calculate the self-gravitational potential in astrophysical plasmas. This makes use of the block-structured adaptive mesh framework and solver provided by the Chombo library to produce a solver that is both resource efficient and designed for parallel computation.

Figure 5.6 shows preliminary results of the self-gravity solver used on a disk simulation containing the potential of an embedded planet, visible at the bottom of the surface density plots on the left. Due to unresolved issues with the plotting routines, none of units on the spatial axes are correct. The potential generated by the solver is shown on the right. The fact that the potential is greatest at the center of the radial range is an artifact of the simple boundary conditions. Constant boundary conditions ($\phi = 0$) mean that for a constant density distribution in the radial direction, the solution is a quadratic function with roots at the boundaries and a maximum in the middle. The perturbation in the potential due to the mass flowing onto the planet is small because the inverse Laplacian operator smooths out the three orders of magnitude variation in the density. Compare to the extent with the sample problem in Figure 5.3 is smoothed out to the solution in Figure 5.4, where thirteen orders of magnitude are reduced to three in the solution. This made visualizing the potential induced by the planet difficult, but there clearly is an increase in the potential at the location of the planet and the solution is refined on the necessary nested grids.

The applications of a self-gravity solver within a adaptive mesh code are plentiful. Star formation simulations spanning vast scales in space and density often require the use of AMR grids to capture the large scales of gravitational collapse (Vaytet et al., 2018; Butler et al., 2017; Ibáñez-Mejía et al., 2017; Padoan et al., 2017). Gravitationally unstable disks have already been shown to be well-modeled using an adaptive mesh (Lichtenberg & Schleicher, 2015), but has yet to be exploited for investigations into the long term evolution of fragments and self-gravitating disks.

Future work will make use of the solver for broader elliptic problems such as flux-limited diffusion (FLD) (Levermore & Pomraning, 1981) radiative transport and this solver has been integrated in a way such that it can be applied to any desired elliptic problem as long as the user properly provides the necessary inputs. The inclusion of a one- or two-temperature FLD is a more complicated task than adding self-gravity, requiring one or two additional equations to solve for the gas radiation energy density and the particle radiation energy density.

Disk Fragmentation

Adaptive mesh simulations have been used to model disk fragmentation before, but have been limited mainly to formation and as a proof of concept (Boss, 2017; Lichtenberg & Schleicher, 2015). Adaptive mesh codes are

The Wengen tests, named after the region in Switzerland where the set of standardized self-gravitating simulation results were devised, are useful first benchmark for the solver's ability with self-gravitating problems. In particular I will aim to emulate the Wengen test # 4, which concerns itself with the modeling of a self-gravitating disk with a variety of codes and parameters, including resolution, mass and domains [3].

Additional Features

While the current solver is a working prototype, there are still many important little details which need to be added before it is a usable solver. First and foremost, the current implementation only supports periodic and constant value (Dirichlet) boundary conditions which are very limiting in astrophysical contexts. In Figure 5.4, the self-gravity potential on the right has a clear bias towards the middle of the radial axis which should not exist were it not for the constant boundary conditions enforced on either end which makes for a quadratic solution in the radial direction and a clear ring-like form of the potential. This will be addressed in the future.

Furthermore, no adjustments have been made to the solver or potential due to the geometry of the problem, so at the moment the solver is only reliable on a Cartesian grid. Additional work would be necessary to make the solver compatible with cylindrical and spherical coordinates.

[3] https://users.camk.edu.pl/gawrysz/test4/

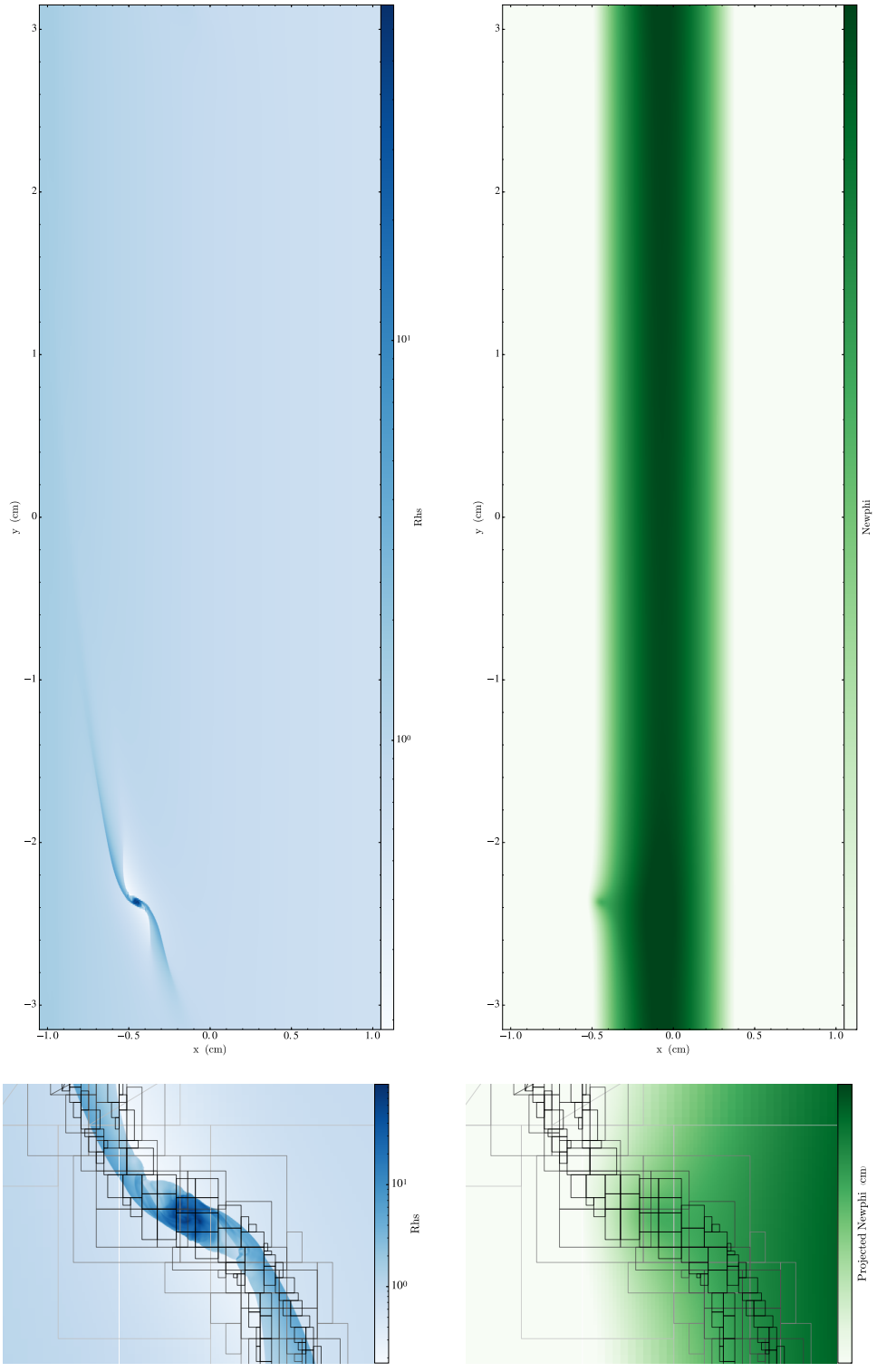


Figure 5.6: A two-dimensional test problem of the multigrid solver showing the density distribution on the left and the potential which results from the multigrid solver on the right. Figures on the bottom are soomed in to the potential where the there is the potential of an embedded planet. The setup is a disk with the potential of a planet near the lower left corner. The potential has not been applied to the equations of motion at this point, so none of the features of the simulation are a result of self-gravity.

6 Final Remarks

Presented in this thesis is a collection of simulations which start from two-dimensional self-gravitating simulations aimed at resolving the inability of simulations to converge with increasing resolution to three-dimensional simulations with either embedded particles or refined meshes. These were conducted at high resolutions to fully resolve the relevant physics of fragmentation and avoid fragmentation due to numerical errors, resolve the dense gas clumps that arise from the fragmentation of gravitationally unstable disks, and resolve the properties of gravitoturbulent flows and it's effect on particle dynamics. I conclude this thesis by briefly summarizing the state of research related to gravitational instabilities, addressing convergence of the cooling timescale criterion, the limitations of the work presented here and how that might shape future work on the topic.

6.1 State of the Field

While the occurrence of fragmentation is likely rare, many more disks may be large enough to have periods of gravitoturbulence which can have a significant effect on the evolution of the disk and the potential to form planets. Self-gravitating disks have been shown to concentrate solid material aerodynamically within the spiral modes, where they may become trapped and grow. This may have important implications for the initial solid distribution which is the basis of planet growth in the core accretion model. If material can grow to planetesimal or embryo size, they could be the seeds of early planet growth and provide an explanation for the observed rings in young disks.

Planet formation by core accretion and pebble accretion is favored as a formation mechanism for the majority of observed planets, but there is a niche regime where disk fragmentation can fill between planet formation and star formation. Provided a disk is massive enough, roughly 10% of the stellar host mass, and cools rapidly enough, on the order of the dynamical timescale, a disk can gravitationally collapse into a clump of dense gas. Core accretion and even pebble accretion are too slow to form such massive planets currently observed beyond 50 au and this is where disk fragmentation can possibly explain the population of gas giants and brown dwarfs observed through direct imaging. Improved disk constraints from observations and simulations suggest that planet formation is less plausible than before, but that is only because the objects formed are closer to brown dwarfs and low-mass stars in size.

As the gravitational collapse of gas is the same process which forms stars, disk fragmentation is often invoked as the origin of binary companions with orbits on the order of 100 au, since the fragmentation of filaments happens on much larger scales. Disk instabilities already form objects with masses and sizes which border on the territory of brown dwarfs and stars, depending largely on the how one separates a planet and a brown dwarf. If one defines a brown dwarf as anything with 8 Jupiter masses or more, it could be that gravitational instabilities are in all but the rarest cases a start formation process.

6.2 Convergence

Non-convergence was clearly shown for both grid and smoothed particle hydrodynamic treatments of self-gravitating disks. These two treatments had two different reasons for showing diverging behavior with increasing resolution, both of which were rooted in the numerical handling of the core physical processes. In the case of SPH, it was shown that numerical viscosity can remove angular momentum and thus rotational support from a collapsing fragment, thus affecting the ability of a simulation to fragment as resolutions and viscosities are adjusted. With an appropriate viscosity, they were able to recover a cooling criterion for fragmentation consistent with earlier work.

In the case of grid codes, the gravitational smoothing length was shown to be the culprit of non-convergence in 2D simulations, since the potential of a 2D overdensity peaks too strong, and stronger for higher and higher resolutions, and collapses despite more and more pressure support from longer cooling times. To this end, it was a natural next step to verify that 3D simulations would converge, given that smoothing was no longer an issue. The fact that no divergence with resolution was found, which has since been confirmed by others indicates that the community is reaching a consensus about the cooling requirements for disk fragmentation. There are some additional factors which may result in fragmentation at longer cooling times, such as equation of state and prompt versus delayed (pressure-supported) fragmentation.

Convergence of the cooling criterion indicates that disk fragmentation is indeed limited to the outer regions of a disk where the disks can cool quickly. Other studies had suggested that the larger critical cooling parameter may allow for fragmentation at smaller and smaller orbital radii, but that does not appear likely. Some recent results which have suggested fragmentation around M dwarfs to occur within 1 au, but the results should be considered with extreme caution due to the use of an isothermal equation of state that does not allow for overdensities to heat up or cool down.

6.3 Limitations

With simplified cooling, one is able to directly compare the physics of the gravitational collapse to dynamic timescales of the disk, which can then be compared to the observations. This is not how realistic cooling works however, and there effects related to the optical depth and external radiation which cannot be captured with the cooling prescription used here. The efficiency with which the fragment cools could have significant effects on the ability of a disk to fragment or not. Radiative transfer should ultimately be the goal, but care must be taken that direct comparisons to observations and disk parameters are possible. Full radiative transfer will also have an impact on the ability of solids grains to sediment to the fragment centers, where the opacity will rise dramatically with the higher particle densities and increase the thermal pressure support.

The scientific results presented in this thesis use local shearing boxes which, by construction, cannot incorporate the large scale dynamics of the disk into the collapse of the fragment. This excludes most discussions of migration and further fragment evolution including growth, mergers and disruption. Future work will aim to maintain high spatial and temporal resolution of the small scales of the fragment collapse, but also include the large scale disk hydrodynamics.

6.4 Prospects for Future Research

Many disks will likely have a self-gravitating phase which will shape the distribution of gas and solids from which future planetary systems depend. If that includes fragmentation, it is important that future research continue studying the formation criteria to understand the location and frequency of companions formed by fragmentation. The evolution of companions after fragmentation also merits further study, as it is highly unlikely that a fragment embedded in a disk will remain at its formation location due to planet-disk interactions. If the disk does not fragment, gravitoturbulence demonstrates the potential to distribute and grow solids to size of planetesimals, which could dramatically affect the planet formation timescale.

The Pencil code is a fantastic code for modeling turbulence and other high-order effects and has many advanced particle modules, making it an extremely useful code for gravitoturbulent simulations of planetesimal formation. Further physics to be used/added would likely include improved particle interactions with radiative processes and each other as well as more advanced equation of state handling, to model the secondary collapse of a fragment.

Current simulations of fragmenting disks are limited by insufficient resolution down to the secondary gravitational collapse of the clump into bound object. Adaptive mesh simulations hold promise to resolve collapse in the full disk while including additional physics such as solids, magnetic fields and chemistry, to name a few, but much work remains to be done. However, AMR codes are generally well-tested for these purposes and should quickly provide reliable results. With PLUTO, with the addition of the elliptic solver, there is the potential to add two-temperature flux-limited diffusion, which would allow for high-resolution studies of self-gravitating disks, including fragment formation and evolution.

Smoothed particle hydrodynamics, which is even better at resolving at wide spatial and temporal scales, has been used to great effect, but is often inadequate for modeling turbulence. Hybrid moving mesh codes such as GIZMO and Arepo have the potential to overcome the disadvantages of both traditional grid codes and traditional SPH codes, hopefully making simulation of gravitational instabilities and fragmentation more efficient in the future.

6.5 Conclusions

The final conclusions of the thesis are summarized by chapter as follows:

Chapter 3 extended models traditionally performed in 2D shearing boxes and added vertical hydrostatic equilibrium initial conditions to study the convergence with resolution of the cooling criterion. It was found that 3D simulations converged at the same effective resolutions at which 2D models failed to converge. This helped to diagnose that it was indeed the lack of a gravitational smoothing length which was promoting fragmentation at higher resolutions, which keeps the formation of companions by disk instability limited to the cool distant regions of the disk. With a more appropriate 3D gravitational potential, collapse was possible without sacrificing small scale gravitational interactions. With these 3D fragmentation simulations, I was able to give better initial minimum mass estimates of the clumps which eventually form planets or brown dwarfs.

In Chapter 4, solid particles introduced to 3D gravitationally unstable and fragmenting local disk simulations concentrated first along the linear gas density perturbations through aerodynamics. Later, in the nonlinear collapse phase, particles collected in the centers

of the fragments through the combined self-gravity of the gas and solids. Fragments consequently form with considerable solid masses at the centers, effectively cores. For typical clump masses of 5 to ten M_{Jup}, cores are 40 to 70 $M_{\oplus}$, as long as the particles are moderately sized ($\mathrm{St} = 0.1 - 1$). The overall metallicity of the fragments is increased to 1.5 to 5% over the initial interstellar medium value of 1%, although the fragment atmospheres showed variable deviations. These initial constraints will allow for a better understanding of which directly observed stellar companions could have been formed based on their mass and solid content.

Chapter 5 details the implementation of a multigrid elliptic solver for the purpose of adding self-gravity to adaptive mesh simulation of protoplanetary disks using the PLUTO code. The multigrid solver already existed in the Chombo libraries, but was integrated with the existing PLUTO-CHOMBO framework to solve for the self-gravitational potential in a number of physical situations and grid geometries. The solvers is flexible and can be extended to accomodate additional physics, such as more sophisticated radiative transfer.

Acknowledgments

A big thanks to all of the kind and supportive colleagues and mentors for their many insights into planet formation, disk hydrodynamics and numerical methods: Kaitlin Kratter, Wlad Lyra, Willy Kley, Chao-Chin Yang, Ken Rice, Cathie Clarke, Richard Booth, Bertram Bitsch, Til Birnstiel, Mario Flock, Kees Dullemond, Ludmila Carone, Colin McNally, Andrea Mignone, Fritz Röpke, Ralf Klessen, and others. I would also like to thank the members of by defense committee, Fritz Röpkea and Arthur Hebecker, for participating in my thesis defense and Henrik Beuther especially for being the second grader of the thesis.

Most plots featured in this thesis were made using Python and the packages NumPy (Hunter, 2007) and Matplotlib (van der Walt et al., 2011) with the iPython extension (Pérez & Granger, 2007) for simple and flexible coding. The visualization of adaptive mesh grids used the package provided by the `yt`-project (Turk et al., 2011). Without their hard work I would still be using IDL.

I would like to thank all my friends and colleagues at the Max-Planck-Institut für Astronomie and the planet and star formation theory group for their constant patience, support, and assistance over the years. In particular, I would like to thank Andreas Schreiber, Adriana Pohl, Asmita Bhandare, Steffi Yen, Matthias Samland, Christian Lenz, Michael Rugel, Mikhail Kovalev and Paul Mollière, not only for their scientific and practical insights, but also for lending a friendly ear during stressful times.

Another special thank you to the residents of office 217, who provided welcome breaks from the tedium of work to make the days more enjoyable: Paolo "Schwanz" Bianchini, Athanasia "Sassa" Tsatsi, Hector "Moreno" Hiss, Tobias "Buli" Buck, Arianna "Andriangela" Picotti, Sven "Poor Redskins Fan" Buder, Martin "Schlegga" Schlecker and the patron saint of weary physicists and astronomers, Papa Francesco.

Vielen dank an Hans-Jürgen und Iris Braum, Lars und Michaela Braum und Familie für ihrer Hilfe und Überstützung die letzten sieben Jahre.

Thanks to my parents, Wolfgang and Jeanne, for their constant encouragement, even from afar, but also for setting the example for the dedicated scientist I strive to become. Being a scientist is clearly hard work, but my parents show it is worth it.

Finally, I would like to thank Hubert Klahr tremendously for taking me in as an aimless, witless master student, giving me a project and thereafter allowing me follow my curiosity wherever it led me.

Appendix A

Modeling Self-Gravitational Disk Fragmentation

Up to this point I have waxed over the intricate details and finer points of hydrodynamic modeling that goes into the current understanding of planet formation research, shifting focus to the astrophysical situation and its consequences. While this suffices in painting picture of a complex system of gas and dust interacting through hydrodynamics radiation, thermodynamics, electromagnetism and gravity, there are many modeling intricacies which go unsaid.

Unfortunately, the assimilation of all these processes into one simulation is impossible to do at this time, so for different problems regarding disks and planet formation one needs to carefully make a number of assumptions to proceed. The material in a circumstellar disk can be fairly accurately modeled as a ideal, inviscid fluid and in the case of cooler outer regions viable for gravitational instability, a thin gaseous disk. Specifically for the disks considered here, total disk masses are on the high end of the expected range, and comparable to its host star $M_{disk} \geq 0.1 M_*$ so that self-gravity of the gas in the disk is significant.

A.1 Hydrodynamic Equations

The hydrodynamic equations can be derived from first principles using the energy-momentum tensor and informed by physical principles such as mass, momentum and energy conservation.

Conservation Equations

Performing high resolution simulations of the disk dynamics associated with GI requires that the task be reduced to an easier problem. For all simulations of this thesis, it is assumed that the disk has negligible magnetization, so the induction equation can be ignored and the conservation equations are reduced to the hydrodynamic (HD) equations. In the case of 2D simulations, the disk is cool and thin enough to warrant equations with only radial and azimuthal directions and all scalar quantities are vertically integrated over a scale height H. Like many astrophysical fluids, the disk is treated with a high Reynolds number so it can be described in a disk around a young stellar object (YSO) using Eulers equation for the conservation of momentum and the typical equations for mass and energy

conservation.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0 \tag{A.1}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\frac{\nabla P}{\Sigma} - \nabla \phi \tag{A.2}$$

$$\frac{\partial \epsilon}{\partial t} + \nabla \cdot (\epsilon \boldsymbol{v}) = -P \nabla \cdot \boldsymbol{v}, \tag{A.3}$$

with mass surface density Σ, velocity $\boldsymbol{v}$, pressure P, gravitational potential ϕ, and energy density ϵ.

Poisson Equation

Since this study is focused self-gravitating disks, the potential due to self-gravity is defined by a potential which is the solution to the Poisson equation for a razor-thin radial-azimuthal disk

$$\nabla^2 \phi = 4\pi G \Sigma \delta(z), \tag{A.4}$$

where δ is the Dirac delta function and ϕ is the self-gravitational potential of the gas. For a 3D simulation, this is simply generalized to

$$\nabla^2 \phi = 4\pi G \rho. \tag{A.5}$$

In the Pencil code, the Poisson equation is most effectively solved by transforming the surface density function into Fourier space via discrete Fourier transform

$$\hat{\rho}(k_x, k_y, k_z) = \sum_{x=0}^{N_x - 1} \left(e^{\frac{-i2\pi}{N_x} k_x \boldsymbol{x}} \sum_{y=0}^{N_y - 1} \left(e^{\frac{-i2\pi}{N_y} k_y \boldsymbol{x}} \sum_{z=0}^{N_z - 1} e^{\frac{-i2\pi}{N_z} k_z \boldsymbol{x}} \rho(x, y, z) \right) \right) \tag{A.6}$$

to find the potential at wavenumber $\boldsymbol{k}$ and transforming the solution back into real space. The solution to the Poisson equation in Fourier space at $\boldsymbol{k}$ is

$$\hat{\phi}(k_x, k_y, k_z, t) = -\frac{2\pi G \hat{\rho}(k_x, k_y, k_z, t)}{\boldsymbol{k}}. \tag{A.7}$$

After the calculation of the potential in Fourier space the process in reversed to get back to real space through a discrete fast Fourier transform. The solution at all wavenumbers is then

$$\phi_{\boldsymbol{x}} = \sum_{\boldsymbol{k}=0}^{N-1} \hat{\phi}_{\boldsymbol{k}} e^{i \boldsymbol{k} \cdot \boldsymbol{x}} \tag{A.8}$$

Since Fourier methods construct a solution using periodic functions, they are ideal for shearing boxes and their periodic domains. The gradient of ϕ is then added to the right-hand side of the momentum conservation equation, thus adding self-gravity to the simulation.

As described in Chapter 2, limiting the the larger wavenumbers is how one adds a smoothing length to a Fourier solver, like that which is used in Pencil. Considering a simulation with a resolution of 2048 grid cells in each direction of length $\sim 50H$, the wavenumbers of the Fourier decomposition range from $k = 1$ to $k = 1024$ and a single scale height spans 41 cells.

Equation of State

Finally, the gas in the disk is considered to be ideal, with mass density ρ, internal energy ϵ, and specific heat ratio γ.

$$P = (\gamma - 1)\rho\epsilon \tag{A.9}$$

The value chosen for the specific heat ratio has an effect on the cooling rate required for fragmentation (Rice et al., 2005). Higher values of γ result in a stiffer equation of state that requires a lower cooling time for fragmentation. In 2D simulations, a ratio of $\gamma = 1.6$ is used here which compared to the value of $\gamma = 2$ by Gammie (2001) might result in a slightly higher fragmentation boundary. Thus, the results here might appear to show a slightly different fragmentation boundary, but this effect is due to the choice of γ. For 3D simulations, $\gamma = 5/3$ the value for a monatomic gas. Even though the cool outer regions of this disk are too cold for atomic hydrogen, it is cold enought that the additional degrees of freedom associated with molecular hydrogen are typically absent.

A.2 Shearing Sheet Approximation

From the conservation equations one shifts to a local description of a small section of the disk using the shearing sheet approximation (Goldreich & Lynden-Bell, 1965). One must first linearize Equations (A.1) - (A.3) around co-moving coordinates in the rotating disk. This proceeds by assuming small perturbations $\boldsymbol{u_0}$ in of the unperturbed velocity

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{u_0}, \tag{A.10}$$

which expands Equation (A.2) to

$$\frac{\partial \boldsymbol{u} + \boldsymbol{u_0}}{\partial t} + ((\boldsymbol{u} + \boldsymbol{u_0}) \cdot \nabla)(\boldsymbol{u} + \boldsymbol{u_0}) = -\frac{\nabla P}{\rho} - \nabla\phi \tag{A.11}$$

yielding four terms

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + (\boldsymbol{u_0} \cdot \nabla)\boldsymbol{u} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u_0} + (\boldsymbol{u_0} \cdot \nabla)\boldsymbol{u_0} = -\frac{\nabla P}{\rho} - \nabla\phi. \tag{A.12}$$

When subtracting the non-perturbed parts and linearizing with respect to $\boldsymbol{u}$, the remaining equations written out in each component in the cylindrical frame r, ϕ, z (radial, azimuthal, and vertical directions respectively) are

$$\frac{\partial u_\mathrm{r}}{\partial t} + (\boldsymbol{u} \cdot \nabla)u_\mathrm{r} - 2\Omega u_\phi = -\frac{\nabla P}{\Sigma} - \nabla\phi \tag{A.13}$$

$$\frac{\partial u_\phi}{\partial t} + (\boldsymbol{u} \cdot \nabla)u_\phi + \frac{u_\mathrm{r} u_0}{r} + u_\mathrm{r}\frac{du_0}{dr} = -\frac{\nabla P}{\Sigma} - \nabla\phi \tag{A.14}$$

$$\frac{\partial u_\mathrm{z}}{\partial t} + (\boldsymbol{u} \cdot \nabla)u_\mathrm{z} = -\frac{\nabla P}{\Sigma} - \nabla\phi. \tag{A.15}$$

The shear perturbed velocity can be rewritten in the local Cartesian frame x, y, z

$$\boldsymbol{u} = \begin{pmatrix} u_\mathrm{r} \\ u_\theta \\ u_\mathrm{z} \end{pmatrix} + \begin{pmatrix} 0 \\ r\Omega(r) \\ 0 \end{pmatrix} = \begin{pmatrix} u_\mathrm{x} \\ u_\mathrm{y} \\ u_\mathrm{z} \end{pmatrix} + \begin{pmatrix} 0 \\ q\Omega_0 x \\ 0 \end{pmatrix}, \tag{A.16}$$

as the shear is expanded around the orbital radius r_0

$$r\Omega(r) = r[\Omega(r) - \Omega(r_0)] = r[\cancel{\Omega(r_0)} + \frac{d\Omega(r_0)}{dr}(r - r_0) - \cancel{\Omega(r_0)}] = q\Omega_0 x, \tag{A.17}$$

using $x \equiv r - r_0$, $\Omega_0 = \Omega(r_0)$ and $(r/\Omega_0)d\Omega_0/dr = q$. Inserting this velocity linearization back into Equations (A.13) - (A.15)

$$\frac{\partial u_x}{\partial t} + q\Omega_0 x\frac{\partial u_x}{\partial y} + \boldsymbol{u}\frac{\partial u_x}{\partial x} - 2\Omega u_y = -\frac{\nabla P}{\Sigma} - \nabla\phi \tag{A.18}$$

$$\frac{\partial u_y}{\partial t} + q\Omega_0 x\frac{\partial u_y}{\partial y} + \boldsymbol{u}\frac{\partial u_y}{\partial y} + 2\Omega u_x = -\frac{\nabla P}{\Sigma} - \nabla\phi \tag{A.19}$$

$$\frac{\partial u_z}{\partial t} + q\Omega_0 x\frac{\partial u_z}{\partial y} + \boldsymbol{u}\frac{\partial u_z}{\partial z} = -\frac{\nabla P}{\Sigma} - \nabla\phi, \tag{A.20}$$

which is more compactly written (without loss of generality to 3D) as

$$\frac{\partial \boldsymbol{u}}{\partial t} - q\Omega x\frac{\partial \boldsymbol{u}}{\partial y} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} = -\frac{\nabla P}{\Sigma} + q\Omega^2 x\hat{\boldsymbol{y}} - 2\Omega\times\boldsymbol{u} - \nabla\phi \tag{A.21}$$

Using these assumptions allows for the modeling of the local properties of the disk while following the evolution of fragments that form. This local approximation ignores other properties of the disk such as accretion, non-local stresses and migration, which are saved for global simulations. As a result, the center of the disk can be considered to be on either side of the simulation depending on how one decides to reconstruct the physical situation from the simulation.

Linearized Co-moving Hydrodynamic Equations

The relevant equations are expressed according to a perturbed velocity with respect to the disk $\boldsymbol{u} = (v_x, v_y + q\Omega x, v_z)^T$ and are similar to the conservation equations as above but with additional terms for the Coriolis effect $2\Omega\times\boldsymbol{u}$ and centripetal force $q\Omega v_x\hat{\boldsymbol{y}}$ as well as heating and cooling terms in the equation for conservation of energy. Additionally, the Pencil code used for these simulations uses entropy s as the thermodynamic variable in the conservation of energy equation. Without loss of generality to the 2D equations, the equations for a 3D shearing box are

$$\frac{\partial\rho}{\partial t} - q\Omega x\frac{\partial\rho}{\partial y} + \nabla\cdot(\rho\boldsymbol{u}) = f_D(\rho) \tag{A.22}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} - q\Omega x\frac{\partial \boldsymbol{u}}{\partial y} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} = -\frac{\nabla P}{\rho} + q\Omega^2 x\hat{\boldsymbol{y}} - 2\Omega\times\boldsymbol{u} - \nabla\phi - \frac{\epsilon}{\tau_f}(\boldsymbol{u} - \boldsymbol{w}) + f_\nu(\boldsymbol{u}) \tag{A.23}$$

$$\frac{\partial s}{\partial t} - q\Omega x\frac{\partial s}{\partial y} + (\boldsymbol{u}\cdot\nabla)s = \frac{1}{\rho T}\left(2\rho\nu\boldsymbol{S}^2 - \Lambda + f_\chi(s)\right) \tag{A.24}$$

The source terms $f_D(\Sigma)$, $f_\nu(\boldsymbol{u})$, $f_\chi(s)$ for hyperdiffusion, hyperviscosity, and hyperconductivity respectively are explicit terms to keep the solution well-behaved when shocks arise and will be expanded upon in Appendix Chapter B.2.

Caveats

The shearing sheet formalism is a power tool to achieve high resolution simulations with simple grid geometries and boundary conditions within a rotating reference frame. How-

ever, this comes with a few drawbacks which merit mentioning, and bear consequence in the presented simulations.

A.3 Viscosity

An important disk parameter to be determined by the shearing sheet model is the α turbulence formalism, first posited by Shakura & Sunyaev (1973). This formalism relies again on the assumptions that the disk is thin, that angular momentum is transported via turbulent eddies locally on length scales less than or equal to the disk pressure scale height H and with speeds less than or equal to the sound speed c_s (Pringle, 1981)

$$\nu = \alpha c_s H. \tag{A.25}$$

Import to note is as part of this assumption, the strength of this α turbulence is evenly distributed between the velocity and length components of ν, and a given value of α does not necessarily endorse a turbulence at extremes of particularly smaller or large turbulent length scales or velocities. This makes α a simple, but limited parameter with which to compare analytic expressions of a disk to the turbulence that arises in simulations.

A thin disk implies that the only non-vanishing term of the vertically integrated stress tensor $\boldsymbol{T}$ is (Balbus & Papaloizou, 1999)

$$T_{r\phi} = -r\Sigma\nu\frac{d\Omega}{dr} \tag{A.26}$$

which, when adding ((A.25)), becomes

$$T_{r\phi} = -\alpha P\frac{d\ln\Omega}{d\ln r}. \tag{A.27}$$

Thus, the total stress in the disk comes only from the pressure and therefore viscosity in this approximation is strictly local.

The turbulent processes from self-gravity, hydrodynamic and magnetohydrodynamic instabilities within a disk combine to produce an effective α viscosity, which can be calculated from the results of a simulation by adding the necessary stresses. In the case of a self-gravitating disk, this means including the average gravitational $\langle G_{r\phi}\rangle$ and Reynolds $\langle H_{r\phi}\rangle$ stresses

$$\alpha = -\left(\frac{d\ln\Omega}{d\ln r}\right)^{-1}\frac{\langle G_{r\phi}\rangle + \langle H_{r\phi}\rangle}{\Sigma c_s^2}. \tag{A.28}$$

This relation is the basis for calculating α where the gravitational and Reynolds stresses are determined by

$$\langle G_{r\phi}\rangle = \int_{-\infty}^{\infty}\frac{g_x g_y}{4\pi G}dz \tag{A.29}$$

$$\langle H_{r\phi}\rangle = \langle\Sigma u_x u_y\rangle \tag{A.30}$$

Figure A.1 shows the total stress derived in the simulations from hydrodynamic stresses (dashed line) and gravitational stresses (dotted line) and their total in the solid dark line.

This α generated should be compared to analytic estimates, which can be derived via

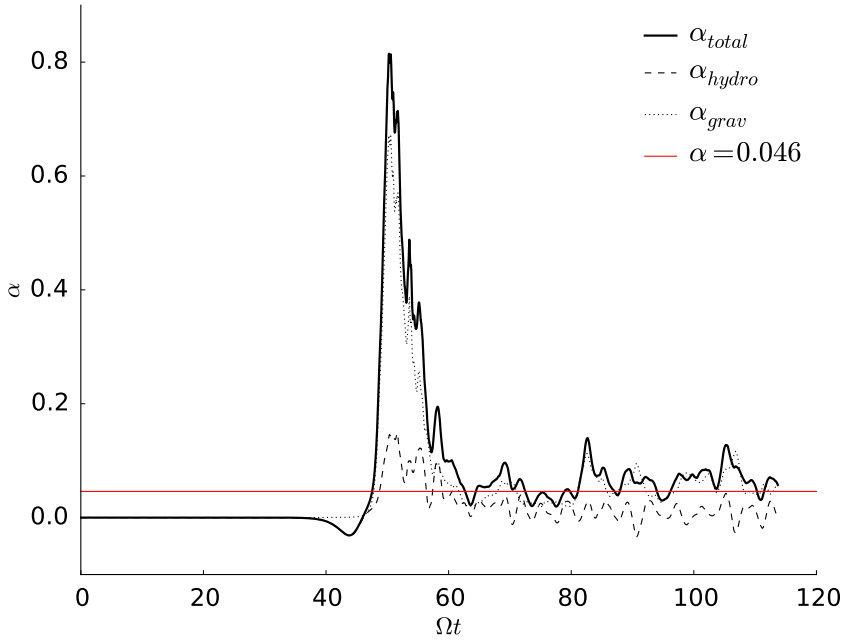


Figure A.1: The calculated (solid black line) versus predicted value (solid red line) of α for a 2D simulation which is a high resolution ($N = 2048$) gravitoturbulent disk with adjusted cooling. Also plotted are the hydrodynamic and gravitational constituents of the α stress. (Baehr & Klahr, 2015)

energy conservation (Gammie, 2001) results in a constant α dependent on the cooling time

$$\alpha = \frac{4}{9}\frac{1}{\gamma(\gamma-1)t_c\Omega}. \tag{A.31}$$

This gives a prediction for the stress in a disk in terms of the cooling timescale of a gravitoturbulent viscously heated disk. A relation between α and the cooling time also means that one can also define the fragmentation boundary in terms of an limit in the stress that a disk can sustain, around $\alpha \simeq 0.1$ (Rice et al., 2005). This is compared to the stress generated by the hydrodynamic motions and gravitational accelerations in Figure A.1 by the solid red line. After the strong gravitational stress induced by the linear collapse phase of the disk, the total stress settles to roughly the value expected from Equation (A.31). It is important to note that this parametrization only holds for stable gravitoturbulent simulations and once fragmentation has occurred, it is no longer reliable.

Though the α-model shows an ability to describe a wide range of influences in a disk, it is still just an approximation that still shows significant variation from model to model. All angular momentum transport processes α values are variable with various disk parameters, including cooling time in the self-gravitating case and even ionization in the case of magnetic turbulence (Armitage, 2011). Additionally, the assumption that viscosity is local in (A.25) is not always consistent with the long range nature of gravitational forces (Balbus & Papaloizou, 1999).

A.4 Thermodynamics

Isothermal hydrodynamics, as in without an energy conservation equation, are often useful for initial studies due to their simplicity, but that fails to capture a lot of important physics. In the case of gravitational instabilities, for a disk to become unstable and fragments to survive, it needs to cool fast enough to overcome the stabilizing forces of shocks and shearing motions. Isothermal simulations do not consider that a fragment heats up as it contracts and thus when any kind of density perturbation starts to grow it inevitably forms a fragment (Backus & Quinn, 2016), which can lead to unrealistic expectations of fragmentation. This means one needs a careful description of how the disk evolves thermodynamically, how heat is generated and released from the disk. Here I cover all of the terms on the right side of the entropy conservation equation (3.8) and their effects on the heating and cooling of the simulation.

Radiative Transfer

Ideally, radiative transfer methods should be employed in the calculation of energy propagation and temperature distribution within optically thin and thick media. This is however computationally expensive in all but the simplest implementations, since the internal energy at a point can be influenced by a countable infinite number of nearby points at any given time. The radiation intensity I_ν at frequency ν at any point along a path s through a medium is given by the equation of radiative transfer, here in differential form (Chandrasekhar, 1960):

$$\frac{dI_\nu(s)}{ds} = j_\nu(s) - \alpha_\nu(s) I_\nu(s). \tag{A.32}$$

The frequency dependent emissivity j_ν adds to the intensity while the extinction α_ν gradually attenuates the radiation along the path. This can be generalized to a time-dependent radiative transfer equation, where a gray approximation is invoked to drop all frequency dependence and ν subscripts

$$\frac{1}{c}\frac{\partial I}{\partial t} + \mathbf{\Omega} \cdot \nabla I + \alpha I = \frac{c}{4\pi}(\alpha_a B - \alpha_s E) \tag{A.33}$$

with the Planck blackbody spectrum B, radiation energy density E defined as the average density of the intensity

$$E = \frac{1}{c}\int I d\Omega \tag{A.34}$$

and total extinction coefficient $\alpha = \alpha_a + \alpha_s$, a the combination of the absorption and scattering coefficients.

The time-dependent radiative transfer equation can be expressed in a conservative form by integrating over the solid angle Ω

$$\frac{\partial E}{\partial t} + \nabla \cdot \boldsymbol{F} = c\alpha_a(B - E), \tag{A.35}$$

for radiative flux $\boldsymbol{F}$. This equation has no closure since the successive moments of radiation intensity add as many unknowns as equations, failing to simplify the problem. Closure can be found in treating radiation as a diffusive process, assuming that the radiative flux is a function of the radiation energy density.

$$\boldsymbol{F} = -\frac{c}{\sigma} D \nabla E \tag{A.36}$$

Crucially, diffusion constant D is a flux-limited term, which keeps the flux from becoming unphysical. With this approximation one can add radiative transfer as another conservation equation in the collection of hydrodynamic equations, where the divergence of flux rewritten as a diffusion of the temperature according to the Stefan-Boltzmann law $E = a_R T^4$ for radiation constant a_R (Kley et al., 2009).

$$\nabla \cdot \boldsymbol{F} = \nabla \cdot \left(-\frac{c}{\sigma} D \nabla E \right) = -\frac{c}{\sigma} (\nabla \cdot D T^3 \nabla T) \tag{A.37}$$

Since the above equation describes an elliptical equation, I intent to use the solver in Chapter 5 to add one- and two-temperature flux-limited diffusion radiative transfer to the PLUTO code. A one-temperature model assumes that the gas and radiation temperatures are coupled or in equilibrium meaning one can use the same equation for both, and is often a good assumption for optically thick media (Kuiper et al., 2010). However, in the case where these temperatures are not necessarily tightly coupled, these temperatures need to be evolved separately, which is known as the two-temperature radiative transport model.

However, for all simulations in this thesis, the β cooling prescription described in Chapter 2 was the most sophisticated radiative cooling technique employed.

Heating Processes

Generally, disks are consistently heated either by irradiation from the central object or by internal processes such as accretion dissipation. Here, heat from accretion is produced by the term $2\Sigma\nu\boldsymbol{S}^2$ with rate-of-strain tensor

$$\boldsymbol{S} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3} \delta_{ij} \nabla \cdot \boldsymbol{u} \right). \tag{A.38}$$

After the initial burst generated by shocks, which produce heat via the adiabatic compression and dissipation of the shock wave, and the simulation settles, viscous heating is the primary source of internal heat.

At larger radii, the dominant effect is irradiation from the star because while flux from the star decreases as $\sim R^{-2}$, while heating by accretion (equation (1.1)) has a steeper dependence $\sim R^{-3}$, making accretion dominant at nearer orbital separations. When under irradiation, the disk will have an effective minimum temperature throughout the vertical extent of the disk (represented by sound speed, in these simulations) which can be related to the Toomre parameter (Rice et al., 2011)

$$Q_{irr} = \frac{c_{s,0}\Omega}{\pi G \Sigma_0}. \tag{A.39}$$

Here the disks background temperature will be maintained at $Q_{irr} = Q_0 = 1$ which prevents the disk from cooling to a 0 background temperature and fragmenting prematurely. A higher background temperature increases the stability of the disk by limiting the effectiveness of cooling and at a certain level, the disk will not be able to fragment (Rice et al., 2011).

A final source of heat which is important is the dissipation of shocks by the final term in equation (3.8), the hyperconductivity $f_\chi(s)$. The simulations here form strong initial shocks before settling and heat generated during this burst is primarily from the dissipation of the shocks. This effect becomes less important when shocks are not as strong in the non-fragmenting steady state of the disk.

Optical Depth

One key feature of disks which goes largely unaddressed in this thesis is the role of the wavelength dependent optical depth

$$\tau_\nu = \int_s^0 \kappa_\nu \rho(s^{'}) ds^{'} \tag{A.40}$$

in the cooling and ultimately the fragmentation of the simulations. While this was attempted in the 2D simulations of Chapter 2, the failure of that cooling prescription to converge fragmentation led to its abandonment. However, it may be worth investigating the fragmentation of 3D simulations which incorporate the optical depth of the gas into the simple cooling prescription.

The traditional β cooling prescription considers that all regions of the simulation cool with the same efficiency which is likely not reasonable. Higher optical depths, like those in the center of a fragment, will trap heat and increase the local cooling timescale which may delay fragmentation or even prevent it altogether. With a background irradiation term which prevents temperatures from dropping too low, it may be that temperatures and cooling times within fragments may never be conducive to gravitational collapse.

A.5 Fragmentation

Disk fragmentation has more aspects which were not fully considered in the course of this thesis which I would like to touch on briefly. The genesis of this thesis was on the conditions for planet formation through gravitational instabilities, but there is growing evidence both observationally and theoretically that the objects that form from gravitational disk collapse are more likely brown dwarfs and low-mass stars than gas giant planets (Rice et al., 2015; Vorobyov, 2013; Kratter et al., 2010).

Star Formation

For starters the disk fragmentation process is not very far removed from the star formation process. Toomre instability is the special case of the Jeans instability, where the rotational support of being in a rotating disk further stabilizes against collapse. This can be seen in the form of the critically unstable Jeans wavelength

$$\lambda_{\mathrm{J}} = \left(\frac{\pi c_{\mathrm{s}}^2}{G\rho_0} \right)^{1/2}, \tag{A.41}$$

which is the same balance of pressure c_{s}^2 and gravity $G\rho_0$ as in the Toomre wavelength of Equation (1.24), but without the rotational support provided by the tidal shear. This shows that disk fragmentation is in principle star formation, and it should not be unexpected that the results of disk fragmentation are very nearly stars. Toomre unstable disks can lead to such massive objects because the critical wavelength is longer for a comparable mass density that is Jeans unstable (Nelson, 2006).

Thus disk fragmentation has been invoked to explain binary star formation in simulations (Tomida et al., 2015; Machida et al., 2011) and also the distinct population of binaries with separations around 100 au (Meyer et al., 2018; Tobin et al., 2016b). Fragmentation around more massive stars, which have more massive, more extended disks (Beltrán et al., 2011; Cesaroni et al., 2017) could potentially result in the fragmentation

at larger scales which could could produce even larger fragments which are initially low- to intermediate-mass stars (Ahmadi et al., 2018; Ilee et al., 2018).

Constraints

One of the biggest factors working against the viability of disk fragmentation is the stringent constraints necessary for it to even occur. The massive disks required for fragmentation are both rare and short-lived. They are rare because not may disks are thought to have the disk mass to star mass ratio ($M_{\rm disk}/M_* \sim 0.1$) necessary for Toomre instability (Kratter & Lodato, 2016) or the short cooling timescales needed for fragmentation (Rafikov, 2005). Even if there is enough mass for collapse to occur the temperatures may be too high due to external heating sources such as nearby high-mass stars or luminous outbursts by the central star. Furthermore any self-gravitating phase of a disk is short-lived due to the high accretion of the spiral arms generated by the instability, which rapidly transports the mass from the disk to the star and weakens the instability.

Even if a fragment does form, it is expected to rapidly migrate (Baruteau et al., 2011; Stamatellos & Inutsuka, 2018) and accrete onto its host star. Even if it does not migrate particularly fast, planet formed via disk fragmentation form so early that the time they have to migrate by type II migration leaves very little chance that they remain in the same form as the when they were formed, likely undergoing tidal disruption as they migrate inwards (Zhu et al., 2012). For all these reasons it seems very unlikely that disk fragmentation should happen or have observable results, but that the fact that some hints towards fragmentation are observed indicates that one or more assumptions could be incorrect.

A.6 Gravitoturbulence

A lot has been said of the Toomre stability and the fragmentation outcome throughout this thesis, but it is also important to understand the nature of the stable outcome of the Toomre analysis, gravitoturbulence. The case of a collapsing fragment is only a very specific outcome when disk cooling is efficient, but it may also be the case that cooling is not strong enough to allow for gravitational collapse. When this is true, the disk settles into a gravitoturbulent state where the balance of viscous heating and radiative cooling leads to a feedback process. A region of the disk might approach the $Q = 1$ Toomre stability limit, but heat generated by the turbulence of gravitational contract brings the region back above the Toomre stability limit. From this repeating process, the disk balances itself near the Toomre instability limit $2 \leq Q \leq 3$ (Rafikov, 2009), producing non-axisymmetric perturbations in the form of spiral arms that can be efficient at transporting angular momentum (Paardekooper, 2012). This means the surface density may fluctuate significantly but will remains stable to the fragmentation of bound objects for a long duration.

General Turbulence

Turbulence in 3D is most generally defined by the cascade of energy from large scales to small scales with scaling with a power law of index $E(k) \propto k^{-5/3}$, illustrated in Figure A.2 with this characteristic slope of the spectral power, known as turbulent cascade. In this way, gravitoturbulence in shearing boxes shows all the characteristics of a 3D turbulent flow, but this is not necessary matched in global simulations, where dissipation occurs

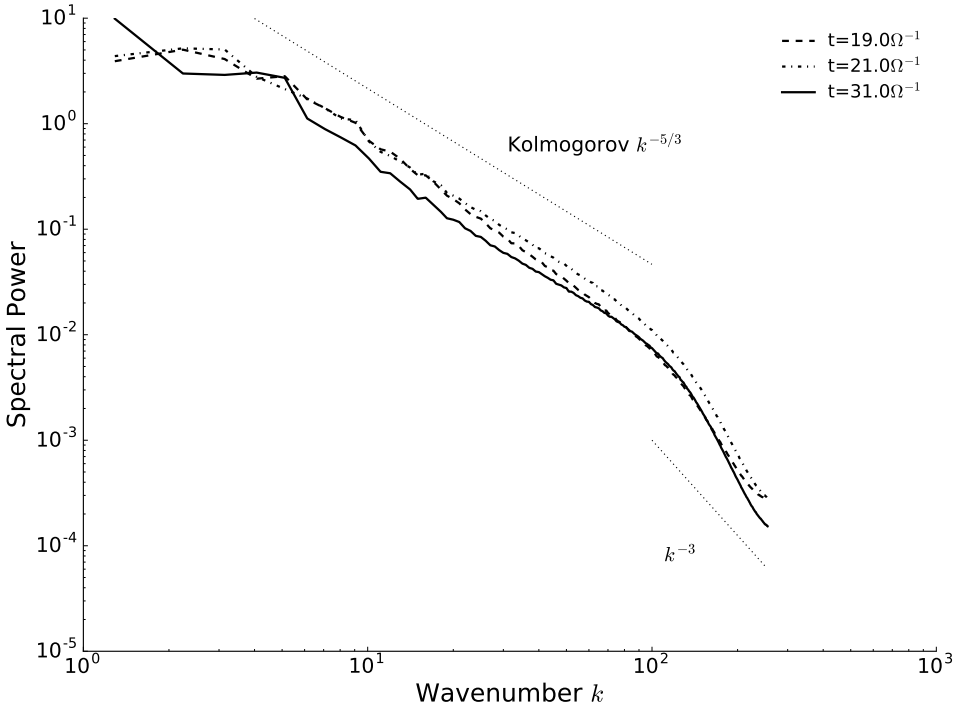


Figure A.2: Power spectrum of the parametric instability run, with clear Kolmogorov cascade from larger scales to smaller scales. The steeper slope at wavenumber $k \approx 100$ is due to the dissipation of

on large scales through shocks (Cossins et al., 2009; Kratter & Lodato, 2016). For 2D turbulence, power is transferred in the opposite direction, to larger scales in an inverse cascade $\propto k^{-5/3}$, while energy is only transferred down in a much steeper cascade $\propto k^{-3}$ (Boffetta & Ecke, 2012). In either case, this can be viewed qualitatively as the transferal of energy from larger eddies to smaller embedded eddies.

Non-axisymmetric Modes

When the Toomre criterion is not met ($Q > 1$) or the cooling timescale is too long, self-gravitating disks display the non-axisymmetric features which are expected of gravitoturbulence. Looking back at the dispersion relation for a self-gravitating disk, one needs to look again at the Lin-Shu dispersion relation

$$(\omega - m\Omega)^2 = \kappa^2 - 2\pi G\Sigma k + c_s^2 k^2. \tag{A.42}$$

Where before only the case where $m = 0$ was considered, which simplified the equation and led to the Toomre parameter, one should now consider the higher order global modes which dominate when growth of the axisymmetric mode is no longer possible. These global modes are not considered in the WKB approximation, but are still excited in gravitationally unstable disks. This is characterized by excitation of numerous radial and azimuthal modes (Cossins et al., 2009) or primarily the $m = 2$ mode (Zhu et al., 2012), depending on the star-to-disk mass ratio and the cooling efficiency (Kratter & Lodato, 2016).

Accretion

The spiral arms generated by gravitoturbulence are very efficient at transferring angular momentum and thus are often linked to high accretion rates through the α parameter. From the discussion about α (see Appendix Section A.3) it is apparent that gravitational instabilites generate unusually large stresses approaching the theoretical limit of $\alpha = 1$. For this reason the fragmentation criterion has been occasionally rewritten in as not a cooling threshold, but an α stress threshold, or the maximum stress the disk can withstand before fragmenting (Rice et al., 2005).

Gravitoturbulent accretion has been used to explain FU Orionis events (Vorobyov & Basu, 2005, 2006), which are sudden, periodic increases in luminosity in young stars which appear tied to accretion (Hartmann & Kenyon, 1996). The strong turbulent viscosities generated by the dissipation of the spiral arms of gravitoturbulence will ultimately drive accretion of large piles of mass onto the protostar, potentially in a periodic fashion which is expected of FU Orionis outbursts.

A.7 Observational Comparisons

Observations of potentially gravitationally unstable disks has always been a challenging endeavor because this phase of disk evolution is short lived and may be obscured by the optically thick envelope which has yet to dissipate. Some recent observations has shown promise towards detecting direct evidence of disk fragmentation. Pérez et al. (2016) observed spiral features in the disk around Elias 2-27, which are potentially a signature of a massive disk, but could also be the result of an unseen planet within the observed disk or outside it (Meru et al., 2017; Forgan et al., 2018; Hall et al., 2018). However, the symmetric $m = 2$ mode of the spirals could still be an indicator of a particularly massive disk and no candidate planet has been discovered.

Shortly thereafter, Tobin et al. (2016a) captured an image of the system L1448 IRS3B which appears to show three co-planar objects in what may be disk fragmentation. Two primary objects are a pair of 0.5 $M_\odot$ binary stars while the large fuzzy object in the foreground is a 0.08 $M_\odot$ clump. It is not guaranteed that the three sources are co-planar and the infrared emission is only an indicator that there is a large clump of dust, not necessarily that it is a gas fragment as well.

Massive stars are expected to have large, extended disks, making them excellent potential locations to search for disk instability. Observations by Ahmadi et al. (2018) and Ilee et al. (2018) have shown that there are good candidates for disk fragmentation around high mass stars, satisfying both the Toomre Q parameter for instability and the cooling criterion. Crucially, Ahmadi et al. (2018) estimate the cooling time in the gravitationally unstable areas around the massive stars of W3(H_2O) and find that the cooling time is easily below unity.

Indirect evidence for gravitationally unstable disks has been inferred from surveys of binary stars in the Perseus cloud, which indicates a bimodal distribution of binary companions by separation. There is a significant number of binary with separations of ~ 10000 au and ~ 100 au and relative deserts elsewhere. This has been interpreted as a sign of two different formation mechanisms for such binary companions. The former are likely formed from the fragmentation of the parent filament, which has a typical separations of around a parsec (= 206265 au) (Chira et al., 2018) while disk fragmentation is most likely within 100 to 1000 au, depending on the size of the host star and disk.

A.8 Numerical Convergence

Finally a word must be said about the convergence of a finite difference approximation. The simulations in this investigation are considered convergent when the behavior is the same regardless of the number of grid cells or particles. The behavior is evaluated by the critical cooling timescale which delineates between fragmentation and gravitoturbulence. When this critical value changes with resolution it means that there is either something wrong with the numerical approach (Lodato & Clarke, 2011) or with the understanding and application of the physics (Paardekooper, 2012).

The change in the planet formation constraints is an issue where non-fragmentation effects the physical interpretation of the simulation. But non-convergence is an issue in and of itself, which may mean the problem is either poorly formulated or affected by numerical schemes. One condition for convergence that must be met is the Courant-Friedrich-Lewy condition, which says that the mathematical domain of dependence should be within the numerical domain of dependence. This is maintained by setting a Courant number

$$C = \frac{u\delta t}{\delta x}, \tag{A.43}$$

such that information cannot travel to other parts of the simulation faster than can be resolved by the spacing of the grid cells δx for a signal speed u. A more general statement of stability is that the Euclidean norm of the solution at a single point is bounded by some real constant K in the domain at all points in time.

$$||(A^h)^{-1}|| < K \qquad \forall\, h < h_0 \tag{A.44}$$

The linear operator A is the matrix which defines the discrete system of equations with unknowns U and grid spacing h, i.e. $A^h U^h = F$. For a linear finite difference model, the Lax-Richtmyer equivalence theorem states that a stable numerical scheme converges if and only if the scheme is also consistent (LeVeque, 2007).

$$stability + consistency \iff convergence \tag{A.45}$$

Showing the first property is straightforward, since the simulation only needs to remain definite, i.e. not crash. Numerical instabilities in the execution of a simulation suggest that there is a significant problem with the physical setup, and this needs to be stabilized for as large a spatial and temporal domain as possible. All simulations in this thesis are bounded and stable for times depicted.

Secondly, showing that the simulations are consistent is difficult, so instead of using consistency to show convergence it is often more convenient to use convergence to infer consistency. For consistency the norm of the local error τ should approach 0 with increasing resolution

$$||\tau^h|| \to 0 \qquad \text{as} \quad h \to 0. \tag{A.46}$$

There have been a few studies into the numerical convergence of the cooling time scale and its effect on fragmentation (Meru & Bate, 2011; Paardekooper, 2012) which fail to show convergence and thus fail to show consistency.

Appendix B

Numerical Tools

To perform the simulations in this thesis in a timely and efficient manner, a number of techniques were used and a lot of parallel computing nodes. The MPIA funded *Theo* and *Isaac* clusters, operated and maintained at the by the Max Planck Computing and Data Facility (MPCDF), handled the majority of the workload. Additional high resolution runs that required computing power beyond the capabilities of these two clusters were run on *Hydra*, also maintained by the MPCDF, and *JUQUEEN* operated by the Jülich Supercomputing Centre (Stephan & Docter, 2015). The following section details some of the basic numerical methods underlying the Pencil and PLUTO codes as well as methods employed to stabilize and streamline code performance.

B.1 Eulerian Methods

An Eulerian description of a flow uses fixed test points in the laboratory frame to indicate a fluid's current state and how it evolves over time. In what are known as grid or mesh codes, the domain is divided systematically into small cell which represent a small part of the fluid at that point. This is often easy to handle, as there is a constant system of coordinates and a uniform mesh from which to extract data and measurements.

Finite Difference Schemes

Since a solution to the equations (A.22)-(A.24) cannot be easily be derived from analytic methods they have to be numerically evaluated on a computer. Analytic solutions can be found for simple linearized systems of equations, but to see the more interesting nonlinear evolution of a set of physical equations, one needs to use numerical tools. A finite difference method accomplishes this by replacing the derivatives with approximations, turning the series of partial differential equations into a large number of algebraic equations which can be solved by a computer.

To proceed one must then define how to approximate derivatives without using a continuous function as a basis. This means dividing the domain of interest into discrete, systematically spaced regions called cells, thus the name 'finite difference' (LeVeque, 2007).

The definition of a derivative requires evaluating the quantities such as density, velocity, internal energy, etc. between some infinitesimally small distance h

$$\frac{df}{dx} \equiv \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}, \tag{B.1}$$

from which one can construct a simple finite difference derivative operator D_+ between

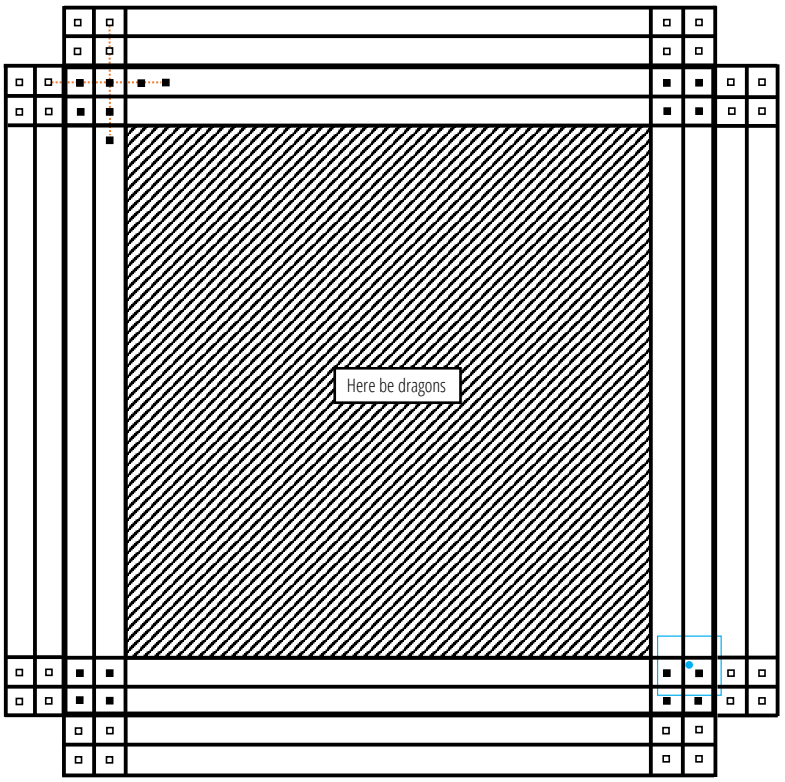


Figure B.1: Illustration of a finite difference simulation domain (area within thicker black border), including the ghost zones (outside of border with open squares) used to evaluate stencils near the borders. Also shown is the five-point stencil used to calculate the fourth-order difference operator at the central cell. A particle positioned at the blue dot has its mass distributed over the cells which intersect with the blue square. Inspired by a similar figure in (Stone & Norman, 1992)

any two points in a simulation domain

$$D_+ f \equiv \frac{f_{i+1} - f_i}{\delta x}, \tag{B.2}$$

which says that the derivative of f at position (i) is approximated as the difference between f evaluated at the current grid cell (i) and the next one over $(i+1)$, divided by the distance between the two grid cells, δx. As this is a simple estimation, the truncation error between the approximation and the true value is significant and will become larger with each iteration of the method, making the approximated solution less and less reliable as the method progresses. This leads to the need for more accurate approximations, like the six-order centered difference expression

$$f_0' = \frac{-f_{-3} + 9f_{-2} - 45f_{-1} + 45f_1 - 9f_2 + f_3}{60\delta x} - \frac{\delta x^6 f^{(7)}}{140} = D^{(cent,6)} + O(\delta x^6). \tag{B.3}$$

where $f^{(7)}$ is the truncation error on the order of $O(\delta x^6)$.

In the Pencil Code grid cells are a representation of the fluid within that region, which

is called an *Eulerian description* of a flow. The quantities such as density, velocity, internal energy, etc., within each cell at the current time are therefore calculated from neighboring cells at the previous tome using what is known as a stencil. Since I used PENCIL's standard sixth-order difference operator, this means that the stencil contains points from three cells in each direction.

PENCIL gets it's name from the way in which it stores derived variable quantities. From the primary physical quantities, a number of secondary quantities can be derived, such as vorticity, entropy, etc., all of which are calculated and updated at specific intervals, forming a bundle of long arrays, like pencils, at each cell. Before the beginning of a simulation, the PENCIL Code checks all included modules for which variables are necessary for the computation and includes them in the 'pencil case'.

The length of the time-step δt is determined by the velocity and sound speed within the simulation, $\boldsymbol{u}$ and c_s, respectively, the grid spacing δx, and the Courant criterion $C_0 = 0.4$, a constant chosen to ensure convergence.

$$\delta t = C_0 min\left(\frac{\delta x}{u_x + c_s}, \frac{\delta y}{u_y + c_s}\right) \tag{B.4}$$

This δt is computed over the entire domain giving a single time-step to advance for the entire simulation uniformly. Other properties such as spatial averages, time averages, and other outputs are calculated based on the pencil quantities taken from the modules at some predetermined simulation time interval. Ultimately, the differencing at each grid cell is calculated using an upwinding scheme to eliminate spurious signals.

Upwinding

Upwinding is a special differencing scheme designed to take into account the direction of the flow at each grid cell for improved code stability. This takes the flow direction into account, biasing the differencing scheme in the direction opposite the direction of the flow. This can be useful in eliminating extra oscillations from central difference scheme and changes the derivative operator in Equation (B.3) to the form which is used in my simulations

$$f_0' = \frac{-2f_{-3} + 15f_{-2} - 60f_{-1} + 20f_0 + 30f_1 - 3f_2}{60\delta x} - \frac{\delta x^5 f^{(6)}}{60} = D^{(up,5)} + O(\delta x^5). \tag{B.5}$$

The use of upwind differencing method will generally improve the stability of a simulation, but it might be at the cost of accuracy, particularly from the downwind direction, where fewer points are used in the update stencil.

Finite Volume Schemes

Eulerian codes come in another common form wherein the quantity U at each position is not determined by the values within the neighboring cells, but the flux at the interface between the two points. Since a flux F is calculated between neighboring cells whose centers are separated by Δx, the area which makes up the interface between the two and their respective cell volumes is explicitly determined.

$$U^{t_{n+1}} = U^{t_n} - \frac{\Delta t}{\Delta x}(F_+ - F_-) \tag{B.6}$$

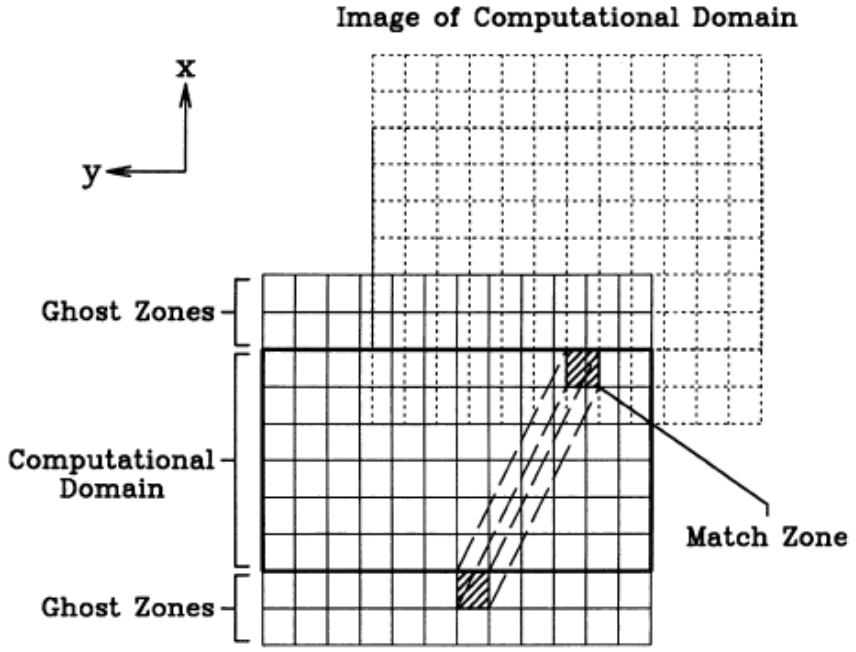


Figure B.2: Illustration of ghost zones used for the calculation of quantities near borders. By periodic boundary conditions, the grid cell just inside the boundary on one side is calculated based on the grid cells on the opposite end of the simulation. Due to the high-order scheme that uses three points on either side of the current position, three rows of ghost zones are used in the Pencil Code compared to the two shown here. Figure modified from Hawley et al. (1995)

A method such as this has the advantage of being strictly conservative, meaning there will be no mass or energy gain/loss from the numerics, although angular momentum may not be properly conserved where Cartesian grids are used for cylindrical or spherical coordinates. This contrasts with the Pencil code which relies on its high order scheme to minimize any numerical losses.

Additionally, rather than smoothing out shocks with artificial viscosities and dissipative and diffusive terms as in a finite difference scheme like Pencil (see Appendix Section B.2 on hyperpissipation), a finite volume scheme uses what is known as a flux limiter or a slope limiter to keep the solution within a cell from having a slope that is too steep. A steep solution could introduce spurious noise to the solution which affects the integrity of the simulation.

Ghost Zones

All grid codes require additional grid cells outside the computational domain, shown in Figure B.2. These zones are necessary for the calculation of physical properties just inside the domain, but also for the implementation of the boundary conditions.

As one can see from the difference schemes (B.3) and (B.5), three grid cells are required on either side of the current one to evaluate the differential operator. Thus, when the current point is within three cells of the boundary, there needs to be a cell outside the domain to calculate the differencing scheme. The ghost zones on one side of the simulation merely contain duplicate information from the computational domain of the opposite side. This is shown in figure B.2, where the ghost zone at the bottom is taken from a corresponding

zone on the other side which is part of the computational domain but shifted doe to the constant shear imposed on the simulation and boundary conditions.

B.2 The Pencil Code

The local simulation of a disk is handled by a finite difference, partial differential equation solver for compressible MHD equations. Used for my investigations is the Pencil Code[1] (Brandenburg, 2003), which is chosen for its high-order numerical scheme and its modularity. The Pencil Code is an open source FORTRAN 95 code designed to efficiently calculate a numeric solution to a series of partial differential equations, in this case the ideal HD equations (A.1)-(A.3). It is parallelized to offer efficient computation on various multi-processor supercomputing clusters. Pencil has some features which are used extensively but not elaborated on in the main body of the thesis. In addition to upwinding already mentioned, Pencil simulations in this thesis make heavy use of hyperdiffusion and shear periodic boundary conditions, which are described in greater detail below.

Numerical Hyperdiffusion

Shocks often present problems in hydrodynamical simulations as discontinuities cannot be well represented by high-order polynomials, leading to additional minimums and maximums which affects the ability of the code to properly take a derivative near the shock. Therefore explicit dissipation terms are added to equations (A.1) - (A.3) above to smooth out the waves and shocks so they do not hinder the performance of the simulation. These terms have two parts, one being a sixth-order hyper dissipation method and the second being a localized shock-capturing method, activated on regions with large negative velocity divergences. The hyperdiffusion term looks like

$$f_D(\Sigma) = \zeta_D(\nabla^6\Sigma), \tag{B.7}$$

where ζ is a constant scaled with resolution and with analogous forms for both hyperviscosity ($f_\nu(\boldsymbol{u})$) and hyperconductivity ($f_\chi(s)$). This has the effect of damping power at high wavenumbers (small scales) but preserving power at small wavenumbers (large scales)(Yang & Krumholz, 2012).

The shock-capturing portions for each dissipative term are

$$f_D(\Sigma) = \zeta_D(\nabla^2\Sigma + \nabla \ln \zeta_D \cdot \nabla\Sigma), \tag{B.8}$$

$$f_\nu(\boldsymbol{u}) = \zeta_\nu(\nabla(\nabla \cdot \boldsymbol{u}) + (\nabla \ln \Sigma + \nabla \ln \zeta_\nu)\nabla \cdot \boldsymbol{u}), \tag{B.9}$$

$$f_\chi(s) = \zeta_\chi(\nabla^2 s + \nabla \ln \zeta_\chi \cdot \nabla s), \tag{B.10}$$

where the ζ term for each is analogous to the following viscous example,

$$\zeta_\nu = \nu_{sh}\langle max_3[(-\nabla \cdot \boldsymbol{u})_+]\rangle[min(\delta x, \delta y, \delta z)]^2, \tag{B.11}$$

where ν_{sh} is a prescribed constant for regions where a shock is present.

Periodic Boundary Conditions

The boundary conditions of a shearing sheet box are periodic along $y = 0$ and $y = L$ which means outflow at one boundary is the inflow at the opposite edge, entering at the

[1] http://pencil-code.nordita.org/

same point where it exited. Being outside the simulation domain, derivatives are not calculated in ghost zones. Rather, the values of density, velocity, entropy, etc. needed for the calculation of derivatives within the simulation are copied from corresponding grid cells from the other side of the domain, as shown in figure B.2.

$$f(x, y, z, t) = f(x, y + L, z, t) \tag{B.12}$$

Periodic boundary conditions are simple to use because one need not be concerned about momentum or mass conservation effects or the reflection/transmission of waves from the boundary surface. Since shearing box simulations are also scale free, periodic boundary conditions are especially useful for directions which otherwise have incongruities from one end to the other. Therefore, the 3D simulations in this thesis also use periodic conditions in the vertical direction.

$$f(x, y, z, t) = f(x, y, z + L, t) \tag{B.13}$$

In the radial direction, one requires a special case of periodic boundaries.

Shear Periodic Boundary Conditions

The boundaries $x = 0$ and $x = L$ require a different boundary condition on account of the shear velocity $u_y = v_y + \frac{3}{2}\Omega x$. The three rows of ghost cells on these boundaries not only move with respect to the simulation, but also with respect to each other. This introduces a complicated situation where the ghost zone that matches with a certain computational point is not fixed in time, meaning the simulation counterpart that matches the ghost zone may straddle two physical grid cells. Dealing with this means interpolating between the two cells to get a more appropriate value (Hawley et al., 1995).

$$f(x, y, z, t) = f(x + L, y - \frac{3}{2}\Omega L t, z, t). \tag{B.14}$$

For example, calculating the potential due to self-gravity over a shear periodic x-boundary requires taking into account the displacement due to the shear by shifting the entire y-direction to make the x-direction periodic before proceeding with the transform in the x-direction. Shear periodic boundaries also produce the odd visual effect apparent in Figure 2.2, where structures are disjointed across the vertical boundary even though they are indeed the same structure.

Initial Conditions

The Pencil Code uses dimensionless scale parameters for all physical values and constants. This helps keep numbers from getting too large or too small for the code to handle. The variables and constants can be scaled back to physical units after the computation is complete. The Keplerian frequency $\Omega^2 = \sqrt{GM_*/R^3}$ is one of the more important parameters because it relates the shearing box to its surroundings, with R the distance to the central massive object and M_* being the mass of the central object. The disk aspect ratio H/R then allows for conversion of scales between pressure scale height and disk radius. In 2D simulations, this value is assumed, but in 3D simulations it must be calculated from the sound speed of an ideal gas $c_s = \sqrt{\gamma kT/m_p}$, where m_p is the mass of a proton.

The initial state of the disk is set so that the Toomre value throughout is on the borderline of stability and instability $Q = 1$. By equation (1.25), this condition is met by setting

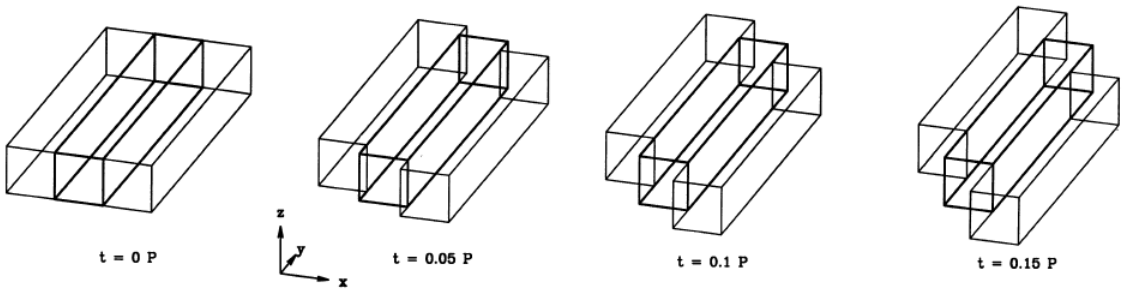


Figure B.3: Schematic of the shear periodic boundary condition. Due to the shear flow, the ghost zones on the x boundary shift in opposite directions and this is accounted for in the boundary condition. Figure modified from Hawley et al. (1995)

the gravitational constant G, the Keplerian orbital frequency Ω, and the uniform surface density distribution $\Sigma_0 = 1$. Furthermore, the sound speed c_s is initially set to π which means the constant background irradiation set by $c_{s,0}^2$ in equation (2.4) is π^2.

The physical length of the simulation needs to be longer than the critical wave length (Toomre, 1964),

$$\lambda_{\rm crit} = \frac{2c_s^2}{G\Sigma_0}, \tag{B.15}$$

however in the case of the 3D simulations of Chapters 3 and 4, the radial extent in particular needs to be larger by a factor of about 2 to contain the full range of unstable waves (Mamatsashvili & Rice, 2010; Booth & Clarke, 2019). The particle-less simulations of this thesis where not large enough, which although it did not affect fragmentation, the result was the erratic *alpha* stresses and gravitoturbulence. In 2D, the ratio of specific heats is set to the 2D adiabatic case $\gamma = 1.6$, which maps to a 3D index of between 1.6 and 1.9 depending on the self-gravitation in the disk (Johnson & Gammie, 2003). For 3D simulations, the value is the typical value for a cold astrophysical gas $\gamma = 5/3$.

B.3 PLUTO Code

For full disk simulations in Chapter 5, I use the code PLUTO[2] (Mignone et al., 2007, 2012). PLUTO uses two forms for all physical variables, conservative and primitive variables. Conservative variables are those which are in the conservation equations and are updated at the final stages using the fluxes at the interfaces. However, primitive variables are much easier to use when calculating fluxes. Thus all quantities are mapped between the two depending on what stage of Figure B.4 the code is, using the following relation for conserved variables to primitive variables $U \rightarrow V$:

$$\begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho E \end{pmatrix} \rightarrow \begin{pmatrix} \rho \\ u \\ v \\ P \end{pmatrix} \tag{B.16}$$

where ρ is the mass density, u and v are the two-dimensional velocity components, total energy $E = P/(\gamma-1) + |\mathbf{m}|^2/2\rho$, momentum density $\mathbf{m} = \rho\mathbf{u}$ and pressure P. Incremental

[2]http://plutocode.ph.unito.it/

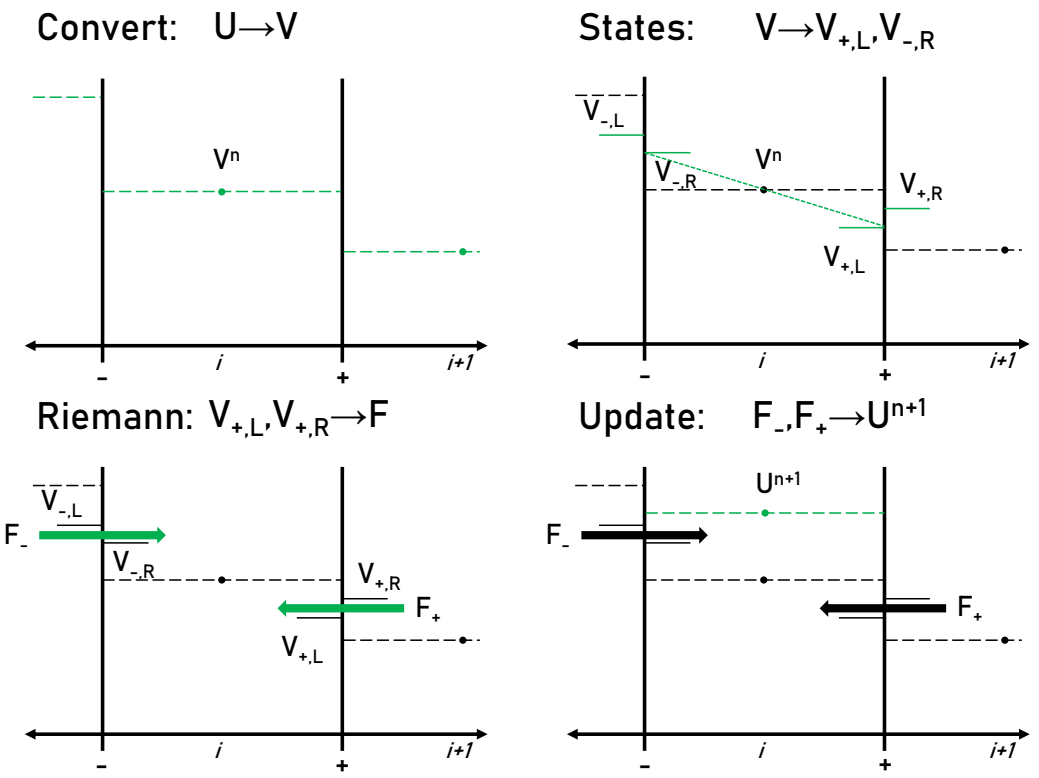


Figure B.4: An illustration of the reconstruct-solve-average strategy used by PLUTO to update a quantity within a cell. Variables are first converted to primitive analogs V^n at timestep n before the interface values $V_{+,L}, V_{i,R}$ are calculated by a interpolation/extrapolation scheme. The value on either side of a boundary are the conditions to a Riemann problem at that results in a flux F_-, F_+ across that boundary. Accounting for all the fluxes in all directions yields a net increase or decrease of the quantity U^{n+1} in that cell at the new timestep $n+1$.

values are thus calculated as

$$\begin{pmatrix} \partial\rho \\ \partial(\rho u) \\ \partial(\rho v) \\ \partial(\rho E) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ u & \rho & 0 & 0 \\ v & 0 & \rho & 0 \\ \frac{1}{2}q^2 & \rho u & \rho v & \frac{1}{\gamma-1} \end{pmatrix} \begin{pmatrix} \partial\rho \\ \partial u \\ \partial v \\ \partial P \end{pmatrix} = \begin{pmatrix} \partial\rho \\ u\partial\rho + \rho\partial u \\ v\partial\rho + \rho\partial v \\ \frac{1}{2}q^2\partial\rho + \rho u\partial u + \rho v\partial v + \frac{\partial P}{\gamma-1} \end{pmatrix} \tag{B.17}$$

To start, all quantities are converted to primitive value

Reconstruction Methods

While the the solution is discretized as a value at each grid cell of the domain, in order to determine the fluxes at the interfaces so the solution can be updated at the new timestep, the values need to be determined at the boundaries between cells even though they are defined at the cell centers. This is illustrated in the top right of Figure B.4 as 'States'. In the simplest approximation, and the one used in the simulation in this thesis, the value at the interface can be estimated through a linear interpolation between the central cell

and the nearest neighbors to either side. To avoid undesired oscillations in the solution, slope-limiters a.k.a. flux-limiters are used to keep the gradient in a cell from becoming too steep. I use the van Leer limiter (van Leer, 1974) with limit function $\phi_{vl}(r)$ as a function of the gradient on the mesh of quantity $r = (u_i - u_{i-1})/u_{i+1} - u_i$

$$\phi_{vl}(r) = \frac{r + |r|}{1 + |r|}; \qquad \lim_{r\to\infty} \phi_{vl}(r) = 2. \tag{B.18}$$

More sophisticated reconstructions of the solution between three points use quadratic fitting because it allows for steeper representation of discontinuities (Colella & Woodward, 1984). A weighted essentially non-oscillatory (WENO) method is one of the more sophisticated reconstruction methods, which uses several neighboring cell to construct a polynomial which can minimize the extrema introduced by shocks very well, even if it is more expensive (Jiang & Shu, 1996). Since this is done separately for each cell in each direction, there will be a discrepancy on the left and right of each interface between two cells. This sets up a Riemann problem at each cell, in each direction and for each substep of the Runge-Kutta integrator.

HLL Family of Riemann Solvers

PLUTO comes with a variety of approaches to the Riemann problem which needs to be solved at each cell interface at least once. In the case of the simulations used here, I used the Harten-Lax-van Leer-Contact (HLLC) solver, which is derived from the HLL solver (Harten et al., 1983). The Riemann problem is a discontinuity of state variables at a boundary for an equation of the form

$$\frac{\partial \boldsymbol{U}}{\partial t} + \frac{\partial \boldsymbol{F}(\boldsymbol{U})}{\partial x} = 0. \tag{B.19}$$

for some quantity $\boldsymbol{U}$ and flux of that quantity $\boldsymbol{F}(\boldsymbol{U})$. The initial conditions at position x

$$\boldsymbol{U}(x,0) = \begin{cases} \boldsymbol{U}_{\mathrm{L}}, & x < 0. \\ \boldsymbol{U}_{\mathrm{R}}, & x > 0. \end{cases} \tag{B.20}$$

A problem of this type is present at each interface between two cell for each variable, and multiple times per timestep for schemes that have multiple substeps (e.g. Runge-Kutta methods, Super-Timestepping methods).

General solutions to this problem consider that waves propagate away from this interface with eigenvelocities of $u + a$, $u - a$ and three degenerate waves of velocity of u. An HLL solver is an approximate solution $\boldsymbol{U}^{'}$ to the Riemann problem which has a single solution between two waves which propagate from the interface with characteristic wave speeds S_L and S_R (Toro et al., 1994).

$$\boldsymbol{U}^{'}(x,t) = \begin{cases} \boldsymbol{U}_{\mathrm{L}}, & x/t < S_L \\ \boldsymbol{U}_{\mathrm{HLL}}, & S_L < x/t < S_R \\ \boldsymbol{U}_{\mathrm{R}}, & x/t > S_R. \end{cases} \tag{B.21}$$

In an HLLC solver, where the 'C' stands for contact, the contact boundary between the two initial states is no longer stationary, but also moves with its own velocity S^*. This

introduces an additional solution to the HLL set

$$\boldsymbol{U}'(x,t) = \begin{cases} \boldsymbol{U}_{\mathrm{L}}, & x/t < S_L \\ \boldsymbol{U}^*_{\mathrm{L}}, & S_L < x/t < S^* \\ \boldsymbol{U}^*_{\mathrm{R}}, & S^* < x/t < S_R \\ \boldsymbol{U}_{\mathrm{R}}, & x/t > S_R. \end{cases} \tag{B.22}$$

where the solutions $\boldsymbol{U}^*_{\mathrm{L}}$ and $\boldsymbol{U}^*_{\mathrm{R}}$ are

$$\boldsymbol{U}^*_{\mathrm{L}} = \frac{1}{T(S^* - S_L)} \int_{TS_L}^{TS^*} \boldsymbol{U}(x,t)dx \tag{B.23}$$

and

$$\boldsymbol{U}^*_{\mathrm{R}} = \frac{1}{T(S_R - S^*)} \int_{TS^*}^{TS_R} \boldsymbol{U}(x,t)dx. \tag{B.24}$$

The fluxes in at the regions of $\boldsymbol{U}^*_{\mathrm{L}}$ and $\boldsymbol{U}^*_{\mathrm{R}}$ are then

$$\boldsymbol{F}^*_{\mathrm{L}} = \boldsymbol{F}_{\mathrm{L}} + S_L(\boldsymbol{U}^*_{\mathrm{L}} - \boldsymbol{U}_{\mathrm{L}}) \tag{B.25}$$

and

$$\boldsymbol{F}^*_{\mathrm{R}} = \boldsymbol{F}_{\mathrm{R}} + S_L(\boldsymbol{U}^*_{\mathrm{R}} - \boldsymbol{U}_{\mathrm{R}}). \tag{B.26}$$

PLUTO additionally makes use of the advection upstream splitting method (AUSM) which splits the flux at the cell boundary in two, a convective part and a pressure part (Liou & Steffen, 1993; Liou, 1996, 2006).

$$\boldsymbol{F} = \dot{m}\boldsymbol{\psi} + \boldsymbol{P} \tag{B.27}$$

This allows for the convective flux, where $\dot{m}$ is the mass flux and the vector quantity $\boldsymbol{\psi} = (1, u, H)^T$, to be calculated using an upwind stencil while the pressure term $\boldsymbol{P} = (0, p, 0)^T$, which does not necessarily travel in the same direction as the convective flux. This means that the convective flux can be updated with a specific flow direction (upwinded) using the convective velocity while the pressure term can be updated in a separate direction from the convective flux in a subsonic flow.

Runge-Kutte Integration

A Runge-Kutte integrator uses multiple stages (abbreviated RK2, RK3, RK4, etc. for methods using 2, 3, 4, etc. update stages, respectively) to update the contents of a cell is given by

$$U^{t_{n+1}} = U^{t_n} - \frac{\Delta t}{\Delta x}(F_+ - F_-) \tag{B.28}$$

where U is the quantity in the cell being updated over the time increment $\Delta t = t_{n+1} - t_n$, F_+ and F_- are the flux solutions to the Riemann problem and the cell interface and Δx is the distance between the cell center and the neighbor cell center.

These are explicit integrators, meaning they use only information from the previous step to calculate the value at the next step. Implicit integrators on the other hand, use information from both the previous step and the current step to determine the new value within the cell. This seems impossible on first glance, but it works if one calculates all the current values at the same time using a gigantic matrix which spans the entirely of the simulation domain. This is can be useful when there are significant hyperbolic (diffusive)

fluxes which are extremely time-consuming to calculate when doing each cell separately in an explicit scheme. In the simulations used here, an explicit RK2 method was used, which means the time step was divided into two parts, a predictor step U^* and a corrector step $U^{t_{n+1}}$.

$$U^* = U^{t_n} + (t_{n+1} - t_n)\mathcal{L}_n \tag{B.29}$$

$$U^{t_{n+1}} = \frac{1}{2}\left(U^{t_n} + U^* + (t_{n+1} - t_n)\mathcal{L}_n\right) \tag{B.30}$$

where $\mathcal{L}_n$ is the difference operator in direction d

$$\mathcal{L}(\mathbf{V}) = \frac{1}{\Delta\mathcal{V}^d}\left(A^d_+\mathbf{F}^d_+ - A^d_-\mathbf{F}^d_-\right) + S^d \tag{B.31}$$

for interface area A and cell volume V.

One way to avoid this problem with an explicit integrator is to use super-timestepping (STS). STS breaks down the Runge-Kutte timestep into s subdivisions, which do not need to individually meet the Courant-Friedrich-Levy (CFL) stability criterion, but instead only at the end of all substeps through Legendre or Chebyshev polynomials (Alexiades et al., 1996; O'Sullivan & Downes, 2007; Meyer et al., 2012). The advantage to this is that the hyperbolic fluxes are computed within the first steps before the CFL condition is applied. Future work with the PLUTO code, including the implementation of a flux-limited diffusion radiative transfer scheme using the elliptic solver implemented here, would benefit greatly from such a time marching scheme.

B.4 Adaptive Mesh Refinement Methods

Adaptive mesh techniques were developed to circumvent a traditional weakness of grid codes: the inability to dynamically adjust the performance of the code to changing conditions within a simulation. The overhead required to manage this continuous grid restructuring initially appears onerous and complex, but is often times far faster than running a simulation at the same effective resolution over the full domain.

When the problem domain contains physics on various scales which do not have the same grid resolution requirements, it is often useful to employ a adaptive mesh refinement (AMR) scheme. This involves creating a more finely resolved grid located at regions of the original grid where some criterion is met, usually the gradient of density, energy or velocity. The full details of the AMR method employed by PLUTO can be found in Mignone et al. (2012), and a few of the important points are reiterated here.

Refinement Strategies

AMR codes come in two varieties, those which refine with contiguous rectangular blocks within the domain and those which refine individual cells one at a time. The former, called block-structured AMR, breaks down each refinement level into a number of rectangular subdomains, i.e. blocks or patches, and further refinement is confined to within the blocks of the previous level. This makes parallelization more straightforward, since one need only divide the total number of blocks/cells as evenly as possible among the processors. The efficiency of the blocking process determines how many blocks are created and need to be distributed as well as how many cells are refined that did not need refinement. If one calls for a blocking efficiency of 99%, a large number of small patches will be created

In the PLUTO execution of AMR, the cells which will be refined are determined by

evaluating the norm of the first or second derivative of one of the conserved variables U (Löhner, 1987).

$$\chi(\boldsymbol{U}) = \sqrt{\frac{\sum_d |\Delta_{d,+1/2}\sigma - \Delta_{d,-1/2}\sigma|^2}{\sum_d (|\Delta_{d,+1/2}\sigma| - |\Delta_{d,-1/2}\sigma| + \epsilon(|\sigma_{i+1}| + 2|\sigma_i| + |\sigma_{i-1}|))^2}} \tag{B.32}$$

The forward and backwards differences of the conservative variable $\sigma \equiv \sigma(\boldsymbol{U})$ in direction d, $\Delta_{d,+1/2}\sigma$ and $\Delta_{d,-1/2}\sigma$, respectively, determine whether the norm is above the desired threshold $\chi(\boldsymbol{U}) > \chi_r$. AMR in PLUTO currently only refines based on the first and second derivatives of the conservative variable $\boldsymbol{U}$, but there are also good reasons tor refine based on the value of $\boldsymbol{U}$ so this will likely be implemented at some point.

In the case of the latter refinement strategy, named octree for the typical number of new cells made by refining a single cell and the cascade of eight branches that result, the only requirement for subdivision to a higher resolution is that the parent cell be refined to the penultimate level. This makes octree AMR schemes well-suited for simulations with irregular borders or embedded boundaries (Mignone et al., 2012).

Timestepping

The timestep in some AMR schemes is subcycled, which means that the finer grids are updated with a $\Delta t(r^{-l})$ that is a fraction of the timestep at the coarsest level Δt, where r is the refinement ratio from one level to the next (assumed the same for all levels here, although that need not be the case) and l is the level in question. In this way, only the grids at higher refinement are updated at the shorter timestep, rather than the entire domain. This is useful to speed up integration times with minimal additional overhead.

This undoubtedly will not be so simple when using introducing radiative transfer and parabolic fluxes which will require a diffusive timestep constraint dependent on $(\Delta x)^2$. This may then require the introduction of a super-timestepping method as mentioned above.

Coarse-Fine Interfaces

One important aspect of the regridding process is that the finer meshes need to be properly nested, meaning that there must always be one cell of the level l between the cells of levels $l-1$ and $l+1$. This simplifies the exchange of information between refinement levels, meaning there is only one refinement ratio used for refluxing operations and interpolation stencils.

For example, the multigrid solution ϕ is not necessarily smooth over the coarse-fine boundaries and the solution must be interpolated at these interfaces for a smooth solution. The normal procedure (see Figure 3 of Martin & Cartwright (1996)) is to interpolate the values corresponding to the fine cell centers using the coarse cells parallel to the boundary. These values within the coarse cells are then used to interpolate the value of a fine 'ghost' cell within the neighboring coarse cell and create a smooth solution.

This becomes complicated when the coarse-fine interface also abuts the physical boundary of the simulation, because the interpolation stencil lacks a cell to one side. The Dirichlet boundary condition is not necessarily appropriate for the solution within the grid domain and may interfere with a precise solution. In this case a little more effort may be required to obtain a smooth solution, which in CHOMBO means a special function converges the solution at boundary.

Appendix C

Lists

C.1 List of Figures

C.2 List of Tables

Appendix D

List of Author's Works

Those used which contribute to the thesis

Baehr H., Klahr H., Kratter K. M., 2017, The Fragmentation Criteria in Local Vertically Stratified Self-Gravitating Disks, The Astrophysical Journal, 484, 40

Baehr H., Klahr H., 2015, The Role of the Cooling Prescription in Disk Fragmentation: Numerical Convergence & the Critical Cooling Parameter in Self-Gravitating Disks, The Astrophysical Journal, 814 ,155

Baehr H., Klahr H., The Concentration and Growth of Solids in Fragmenting Circumstellar Disks, submitted to The Astrophysical Journal

Baehr H., Mignone A., Zanni C., Klahr H., A Multigrid Elliptic Solver for Adaptive Mesh Refinement in the PLUTO Code, in preparation

Other Works

Batzler D., **Baehr H.**, Klahr H., , in preparation

Appendix E

Bibliography

Adams, F. C., & Lin, D. N. C. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tucson: University of Arizona Press), 721

Adams, M., Colella, P., Graves, D. T., et al. 2015, Chombo Software Package for AMR Applications - Design Document, Tech. rep., Lawrence Berkeley National Laboratory

Ahmadi, A., Beuther, H., Mottram, J. C., et al. 2018, Astronomy & Astrophysics, 618, A46

Alexiades, V., Amiez, G., & Gremaud, P. A. 1996, Communications in Numerical Methods in Engineering, 12, 31

Allen, A., Li, Z.-Y., & Shu, F. H. 2003, The Astrophysical Journal, 599, 363

ALMA Partnership, T., Brogan, C. L., Pérez, L. M., et al. 2015, The Astrophysical Journal Letters, 808, L3

Andrews, S. M., & Birnstiel, T. 2018, in Handbook of Exoplanets, ed. H. J. Deeg & J. A. Belmonte (Springer International Publishing AG), 2205

Andrews, S. M., Wilner, D. J., Zhu, Z., et al. 2016, The Astrophysical Journal Letters, 820, L40

Ansdell, M., Williams, J. P., van der Marel, N., et al. 2016, The Astrophysical Journal, 828, 46

Armitage, P. J. 2011, Annual Review of Astronomy and Astrophysics, 49, 195

Backus, I., & Quinn, T. 2016, Monthly Notices of the Royal Astronomical Society, 463, 2480

Baehr, H. 2015, Master's thesis, University of Heidelberg

Baehr, H., & Klahr, H. 2015, The Astrophysical Journal, 814, 155

Baehr, H., Klahr, H., & Kratter, K. M. 2017, The Astrophysical Journal, 848, 40

Bai, X.-N., & Stone, J. M. 2010, The Astrophysical Journal Supplement Series, 190, 297

Balbus, S. A., & Hawley, J. F. 1991, The Astrophysical Journal, 376, 214

Balbus, S. A., & Papaloizou, J. C. B. 1999, The Astrophysical Journal, 20, 650

Barnes, J., & Hut, P. 1986, Nature, 324, 446

Baruteau, C., & Masset, F. S. 2008, The Astrophysical Journal, 678, 483

Baruteau, C., Meru, F., & Paardekooper, S.-J. 2011, Monthly Notices of the Royal Astronomical Society, 416, 1971

Behrendt, M., Burkert, A., & Schartmann, M. 2015, Monthly Notices of the Royal Astronomical Society, 448, 1007

Bell, K. R., & Lin, D. N. C. 1994, The Astrophysical Journal, 427, 987

Beltrán, M. T., Cesaroni, R., Neri, R., & Codella, C. 2011, Astronomy & Astrophysics, 525, A151

Berger, M. J., & Colella, P. 1989, Journal of Computational Physics, 82, 64

Berger, M. J., & Oliger, J. 1984, Journal of Computational Physics, 53, 484

Beuther, H., Mottram, J. C., Ahmadi, A., et al. 2018, Astronomy & Astrophysics, 617, A100

Bhandare, A., Kuiper, R., Henning, T., et al. 2018, Astronomy & Astrophysics, 618, A95

Birnstiel, T., Dullemond, C. P., & Pinilla, P. 2013, Astronomy & Astrophysics, 550, L8

Birnstiel, T., Ormel, C. W., & Dullemond, C. P. 2011, Astronomy & Astrophysics, 525, A11

Bitsch, B., Lambrechts, M., & Johansen, A. 2015, Astronomy & Astrophysics, 582, A112

Bodenheimer, P., & Lin, D. N. C. 2002, Annual Review of Earth and Planetary Science, 30, 113

Boffetta, G., & Ecke, R. E. 2012, Annual Review of Fluid Mechanics, 44, 427

Boley, A. C. 2009, The Astrophysical Journal, 695, L53

Boley, A. C., & Durisen, R. H. 2010, The Astrophysical Journal, 724, 618

Boley, A. C., Durisen, R. H., Nordlund, A., & Lord, J. 2007, The Astrophysical Journal, 665, 1254

Boley, A. C., Hayfield, T., Mayer, L., & Durisen, R. H. 2010, Icarus, 207, 509

Boley, A. C., Mejía, A. C., Durisen, R. H., et al. 2006, The Astrophysical Journal, 651, 517

Bolton, S. J., Adriani, A., Adumitroaie, V., et al. 2017, Science, 356, 821

Bonnefoy, M., Currie, T., Marleau, G.-D., et al. 2014, Astronomy & Astrophysics, 562, A111

Bonnefoy, M., Perraut, K., Lagrange, A.-M., et al. 2018, Astronomy & Astrophysics, 618, A63

Booth, R. A., & Clarke, C. J. 2016, Monthly Notices of the Royal Astronomical Society, 458, 2676

—. 2019, Monthly Notices of the Royal Astronomical Society, 483, 3718

Boss, A. P. 1997, Science, 276, 1836

—. 2001, The Astrophysical Journal, 563, 367

—. 2004, The Astrophysical Journal, 610, 456

—. 2015, The Astrophysical Journal, 807, 10

—. 2017, The Astrophysical Journal, 836, 53

Brandenburg, A. 2003, in Advances in Non-linear Dynamos, ed. A. Ferriz-Mas & M. Nunez (London: Taylor and Francis), 269

Brandt, A. 1973, in Proceedings of the Third International Conference on Numerical Methods in FLuid Dynamics, ed. H. Cabannes & R. Temam (Berlin, Heidelberg: Springer Berlin Heidelberg), 82

Butler, M. J., Tan, J. C., Teyssier, R., et al. 2017, The Astrophysical Journal, 841, 82

Cabral, B., & Leedom, L. C. 1993, Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH '93, 263

Cai, K., Durisen, R. H., & Michael, S. 2006, The Astrophysical Journal Letters, 636, L149

Casassus, S., Marino, S., Pérez, S., et al. 2015, The Astrophysical Journal, 811, 92

Cesaroni, R., Johnston, K. G., Maud, L. T., et al. 2017, Astronomy & Astrophysics, 602, A59

Chandrasekhar, S. 1960, Radiative Transfer, first edit edn. (New York: Dover Publications), 393

—. 1964, The Astrophysical Journal, 140, 599

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, The Astrophysical Journal, 568, 377

Cheetham, A. C., Bonnefoy, M., Desidera, S., et al. 2018, Astronomy & Astrophysics, 615, A160

Chiang, E., & Laughlin, G. 2013, Monthly Notices of the Royal Astronomical Society, 431, 3444

Chira, R.-A., Kainulainen, J., Ibáñez-Mejía, J. C., Henning, T., & Mac Low, M.-M. 2018, Astronomy & Astrophysics, 610, A62

Clarke, C. J. 2009, Monthly Notices of the Royal Astronomical Society, 396, 1066

Colella, P., & Woodward, P. R. 1984, Journal of Computational Physics, 54, 174

Cossins, P., Lodato, G., & Clarke, C. J. 2009, Monthly Notices of the Royal Astronomical Society, 393, 1157

D'Angelo, G., Durisen, R. H., & Lissauer, J. J. 2010, in Exoplanets, ed. S. Seager (Tucson: University of Arizona Press), 319

Darwin, G. H. 1906, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 206, 161

Deng, H., Mayer, L., & Meru, F. 2017, The Astrophysical Journal, 847, 43

Dubey, A., Almgren, A. S., Bell, J., et al. 2014, Journal of Parallel and Distributed Computing, 74, 3217

Dubrulle, B., Morfill, G., & Sterzik, M. 1995, Icarus, 114, 237

Dullemond, C. P., Birnstiel, T., Huang, J., et al. 2018, The Astrophysical Journal Letters, 869, L46

Durisen, R. H., Boss, A. P., Mayer, L., et al. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil No. 1 (Tucson: University of Arizona Press), 607

Espinoza, N., Fortney, J. J., Miguel, Y., Thorngren, D. P., & Murray-Clay, R. 2017, The Astrophysical Journal Letters, 838, L9

Fischer, D. A., & Valenti, J. A. 2005, The Astrophysical Journal, 622, 1102

Flaherty, K. M., Hughes, A. M., Rosenfeld, K. A., et al. 2015, The Astrophysical Journal, 813, 99

Flaherty, K. M., Hughes, A. M., Rose, S. C., et al. 2017, The Astrophysical Journal, 843, 150

Forgan, D. H., Price, D. J., & Bonnell, I. A. 2016, Monthly Notices of the Royal Astronomical Society, 466, 3406

Forgan, D. H., Ramón-Fox, F. G., & Bonnell, I. A. 2018, Monthly Notices of the Royal Astronomical Society, 476, 2384

Forgan, D. H., & Rice, K. 2011, Monthly Notices of the Royal Astronomical Society, 417, 1928

Fricke, K. 1968, Zeitschrift für Astrophysik, 68, 317

Fromang, S., & Papaloizou, J. C. B. 2007, Astronomy and Astrophysics, 468, 1

Gammie, C. F. 2001, The Astrophysical Journal, 553, 174

Gibbons, P. G., Mamatsashvili, G. R., & Rice, W. K. M. 2014, Monthly Notices of the Royal Astronomical Society, 442, 361

—. 2015, Monthly Notices of the Royal Astronomical Society, 453, 4232

Gibbons, P. G., Rice, W. K. M., & Mamatsashvili, G. R. 2012, Monthly Notices of the Royal Astronomical Society, 426, 1444

Goldreich, P., & Lynden-Bell, D. 1965, Monthly Notices of the Royal Astronomical Society, 130, 125

Goldreich, P., & Schubert, G. 1967, The Astrophysical Journal, 150, 571

Goodman, J. 1993, The Astrophysical Journal, 406, 596

Grady, C. A., Muto, T., Hashimoto, J., et al. 2013, The Astrophysical Journal, 762, 48

Guillet, T., & Teyssier, R. 2011, Journal of Computational Physics, 230, 4756

Hall, C., Rice, K., Dipierro, G., et al. 2018, Monthly Notices of the Royal Astronomical Society, 477, 1004

Harten, A., Lax, P. D., & van Leer, B. 1983, SIAM Review, 25, 35

Hartmann, L. W., & Kenyon, S. J. 1996, Annual Review of Astronomy and Astrophysics, 34, 207

Hawley, J. F., Gammie, C. F., & Balbus, S. A. 1995, The Astrophysical Journal, 440, 742

Helled, R., & Bodenheimer, P. 2010, Icarus, 207, 503

Helled, R., Podolak, M., & Kovetz, A. 2006, Icarus, 185, 64

Helled, R., & Schubert, G. 2008, Icarus, 198, 156

Hirose, S., & Shi, J.-M. 2017, Monthly Notices of the Royal Astronomical Society, 469, 561

Hockney, R. W., & Eastwood, J. W. 1981, Computer simulation using particles (Bristol, PA, USA: Taylor and Francis)

Hubeny, I. 1990, The Astrophysical Journal, 351, 632

Humphries, R. J., & Nayakshin, S. 2018, Monthly Notices of the Royal Astronomical Society, 477, 593

Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90

Huré, J.-M., & Pierens, A. 2009, Astronomy & Astrophysics, 507, 573

Ibáñez-Mejía, J. C., Mac Low, M.-M., Klessen, R. S., & Baczynski, C. 2017, The Astrophysical Journal, 850, 62

Ilee, J. D., Cyganowski, C. J., Brogan, C. L., et al. 2018, The Astrophysical Journal Letters, 869, L24

Ilee, J. D., Forgan, D. H., Evans, M. G., et al. 2017, Monthly Notices of the Royal Astronomical Society, 472, 189

Janson, M., Bonavita, M. A., Klahr, H., & Lafrenière, D. 2012, The Astrophysical Journal, 745, 4

Janson, M., Bonavita, M. A., Klahr, H., et al. 2011, The Astrophysical Journal, 736, 89

Jiang, G.-S., & Shu, C.-W. 1996, Journal of Computational Physics, 126, 202

Johansen, A., & Youdin, A. N. 2007, The Astrophysical Journal, 662, 627

Johnson, B. M., & Gammie, C. F. 2003, The Astrophysical Journal, 597, 131

Kalas, P. G., Graham, J. R., & Chiang, E. 2008, Science, 322, 1345

Kant, I. 1755, Allgemeine Naturgeschichte und Theorie des Himmels

Kennicutt, R. C., & Evans II, N. J. 2012, Annual Review of Astronomy and Astrophysics, 50, 531

Kim, J.-h., Agertz, O., Teyssier, R., et al. 2016, The Astrophysical Journal, 833, 202

Kim, W.-T., Ostriker, E. C., & Stone, J. M. 2002, The Astrophysical Journal, 581, 1080

Klahr, H. 2004, The Astrophysical Journal, 606, 1070

Klahr, H., & Bodenheimer, P. 2003, The Astrophysical Journal, 582, 869

Klee, J., Illenseer, T. F., Jung, M., & Duschl, W. J. 2017, arXiv preprint arXiv: 1704.02193v1, 8

Kley, W., Bitsch, B., & Klahr, H. 2009, Astronomy & Astrophysics, 506, 971

Kley, W., & Nelson, R. P. 2012, Annual Review of Astronomy and Astrophysics, 50, 211

Kratter, K. M., & Lodato, G. 2016, Annual Review of Astronomy and Astrophysics, 54, 271

Kratter, K. M., & Murray-Clay, R. A. 2011, The Astrophysical Journal, 740, 1

Kratter, K. M., Murray-Clay, R. A., & Youdin, A. N. 2010, The Astrophysical Journal, 710, 1375

Kreidberg, L., Bean, J. L., Désert, J. M., et al. 2014, The Astrophysical Journal Letters, 793, L27

Kuiper, G. P. 1951, Proceedings of the National Academy of Science, 37, 1

Kuiper, R., Klahr, H., Dullemond, C. P., Kley, W., & Henning, T. 2010, Astronomy & Astrophysics, 511, A81

Larson, R. B. 1969, Monthly Notices of the Royal Astronomical Society, 145, 271

Laughlin, G., Korchagin, V., & Adams, F. C. 1998, The Astrophysical Journal, 504, 945

Lavie, B., Mendonça, J. M., Mordasini, C., et al. 2017, The Astronomical Journal, 154, 91

Lee, J.-M., Heng, K., & Irwin, P. G. 2013, The Astrophysical Journal, 778, 97

Lenz, C. T., Klahr, H., & Birnstiel, T. 2019, The Astrophysical Journal, 874, 36

Lesur, G., Kunz, M. W., & Fromang, S. 2014, Astronomy & Astrophysics, 566, A56

LeVeque, R. J. 2007, Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems (Philadelphia: Society for Industrial and Applied Mathematics), 341

Levermore, C. D., & Pomraning, G. C. 1981, The Astrophysical Journal, 248, 321

Lichtenberg, T., & Schleicher, D. R. G. 2015, Astronomy & Astrophysics, 579, A32

Lin, C. C., & Shu, F. H. 1964, The Astrophysical Journal, 140, 646

Lin, M.-K., & Kratter, K. M. 2016, The Astrophysical Journal, 824, 91

Lin, M.-K., & Pierens, A. 2018, Monthly Notices of the Royal Astronomical Society, 478, 575

Liou, M.-S. 1996, Journal of Computational Physics, 129, 364

—. 2006, Journal of Computational Physics, 214, 137

Liou, M.-S., & Steffen, C. J. 1993, Journal of Computational Physics, 107, 23

Lissauer, J. J. 1993, Annual Review of Astronomy and Astrophysics, 31, 129

Lodato, G., & Clarke, C. J. 2011, Monthly Notices of the Royal Astronomical Society, 413, 2735

Lodato, G., & Rice, W. K. M. 2005, Monthly Notices of the Royal Astronomical Society, 358, 1489

Löhner, R. 1987, Computer Methods in Applied Mechanics and Engineering, 61, 323

Machida, M. N., Inutsuka, S.-i., & Matsumoto, T. 2011, The Astrophysical Journal, 729, 42

Maire, A.-L., Skemer, A. J., Hinz, P. M., et al. 2015, Astronomy & Astrophysics, 576, A133

Maire, A.-L., Rodet, L., Lazzoni, C., et al. 2018, Astronomy & Astrophysics, 615, A177

Malygin, M. G., Klahr, H., Semenov, D., Henning, T., & Dullemond, C. P. 2017, Astronomy & Astrophysics, 605, A30

Mamatsashvili, G. R., & Rice, W. K. M. 2010, Monthly Notices of the Royal Astronomical Society, 406, 2050

Marcy, G. W., & Butler, R. P. 1998, Annual Review of Astronomy and Astrophysics, 36, 57

Marois, C., Macintosh, B., Barman, T. S., & Zuckerman, B. 2008, Science, 322, 1348

Marois, C., Zuckerman, B., Konopacky, Q., Macintosh, B., & Barman, T. S. 2010, Nature, 468, 1080

Martin, D. F., & Cartwright, K. L. 1996, Solving Poisson's Equation using Adaptive Mesh Refinement, Tech. rep., U.C. Berkeley

Mathis, J. S., Rumpl, W., & Nordsieck, K. H. 1977, The Astrophysical Journal, 217, 425

Mayer, L., Lufkin, G., Quinn, T., & Wadsley, J. W. 2007, The Astrophysical Journal, 661, L77

McKee, C. F., & Ostriker, E. C. 2007, Annual Review of Astronomy and Astrophysics, 45, 565

Mejía, A. C., Durisen, R. H., & Pickett, M. K. 2005, The Astrophysical Journal, 619, 1098

Mercer, A., & Stamatellos, D. 2017, Monthly Notices of the Royal Astronomical Society, 465, 2

Meru, F., & Bate, M. R. 2011, Monthly Notices of the Royal Astronomical Society, 411, L1

Meru, F., Juhász, A., Ilee, J. D., et al. 2017, The Astrophysical Journal Letters, 839, L24

Meyer, C. D., Balsara, D. S., & Aslam, T. D. 2012, Monthly Notices of the Royal Astronomical Society, 422, 2102

Meyer, D. M.-A., Kuiper, R., Kley, W., Johnston, K. G., & Vorobyov, E. I. 2018, Monthly Notices of the Royal Astronomical Society, 473, 3615

Michael, S., Steiman-Cameron, T. Y., Durisen, R. H., & Boley, A. C. 2012, The Astrophysical Journal, 746, 98

Mignone, A., Bodo, G., Massaglia, S., et al. 2007, The Astrophysical Journal Supplement Series, 170, 228

Mignone, A., Zanni, C., Tzeferacos, P., et al. 2012, The Astrophysical Journal Supplement Series, 198, 7

Miotello, A., van Dishoeck, E. F., Williams, J. P., et al. 2017, Astronomy & Astrophysics, 599, A113

Mordasini, C., Alibert, Y., Benz, W., Klahr, H., & Henning, T. 2012, Astronomy and Astrophysics, 541, 97

Müller, T. W. A., Kley, W., & Meru, F. 2012, Astronomy & Astrophysics, 541, A123

Muto, T., Grady, C. A., Hashimoto, J., et al. 2012, The Astrophysical Journal, 748, L22

Nayakshin, S. 2010, Monthly Notices of the Royal Astronomical Society, 408, 2381

Nayakshin, S., Helled, R., & Boley, A. C. 2014, Monthly Notices of the Royal Astronomical Society, 440, 3797

Nelson, A. F. 2006, Monthly Notices of the Royal Astronomical Society, 373, 1039

Nelson, R. P., Gressel, O., & Umurhan, O. M. 2013, Monthly Notices of the Royal Astronomical Society, 435, 2610

Öberg, K. I., Murray-Clay, R. A., & Bergin, E. A. 2011, The Astrophysical Journal Letters, 743, L16

Okuzumi, S., Momose, M., Sirono, S.-i., Kobayashi, H., & Tanaka, H. 2016, The Astrophysical Journal, 821, 82

Ormel, C. W., & Klahr, H. 2010, Astronomy & Astrophysics, 520, A43

O'Sullivan, S., & Downes, T. P. 2007, Monthly Notices of the Royal Astronomical Society, 376, 1648

Paardekooper, S.-J. 2012, Monthly Notices of the Royal Astronomical Society, 421, 3286

Padoan, P., Haugbølle, T., Nordlund, Å., & Frimann, S. 2017, The Astrophysical Journal, 840, 48

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21

Pérez, L. M., Carpenter, J. M., Andrews, S. M., et al. 2016, Science, 353, 1519

Plummer, H. C. 1911, Monthly Notices of the Royal Astronomical Society, 71, 460

Pohl, A., Kataoka, A., Pinilla, P., et al. 2016, Astronomy & Astrophysics, 593, A12

Pohl, A., Benisty, M., Pinilla, P., et al. 2017, The Astrophysical Journal, 850, 52

Pollack, J. B. 1984, Annual Review of Astronomy and Astrophysics, 22, 389

Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, Icarus, 124, 62

Pringle, J. E. 1981, Annual Review of Astronomy and Astrophysics, 19, 137

Raettig, N., Klahr, H., & Lyra, W. 2015, The Astrophysical Journal, 804, 35

Rafikov, R. R. 2005, The Astrophysical Journal Letters, 621, L69

—. 2009, The Astrophysical Journal, 704, 281

Rameau, J., Chauvin, G., Boccaletti, A., et al. 2013, The Astrophysical Journal Letters, 772, L15

Ribas, Á., Bouy, H., & Merín, B. 2015, Astronomy and Astrophysics, 576, A52

Rice, K., Lopez, E. D., Forgan, D. H., & Biller, B. A. 2015, Monthly Notices of the Royal Astronomical Society, 454, 1940

Rice, W. K. M., Armitage, P. J., Mamatsashvili, G. R., Lodato, G., & Clarke, C. J. 2011, Monthly Notices of the Royal Astronomical Society, 418, 1356

Rice, W. K. M., Forgan, D. H., & Armitage, P. J. 2012, Monthly Notices of the Royal Astronomical Society, 420, 1640

Rice, W. K. M., Lodato, G., & Armitage, P. J. 2005, Monthly Notices of the Royal Astronomical Society, 364, L56

Rice, W. K. M., Paardekooper, S.-J., Forgan, D. H., & Armitage, P. J. 2014, Monthly Notices of the Royal Astronomical Society, 438, 1593

Riols, A., & Latter, H. N. 2016, Monthly Notices of the Royal Astronomical Society, 460, 2223

—. 2017, Monthly Notices of the Royal Astronomical Society, 474, 2212

Ruge, J. P., Flock, M., Wolf, S., et al. 2016, Astronomy & Astrophysics, 17, A17

Safronov, V. S. 1969, Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets (Jerusalem: Israel Program for Scientific Translations), 212

Schlaufman, K. C. 2018, The Astrophysical Journal, 853, 37

Schmitz, P. G., & Ying, L. 2014, Journal of Computational Physics, 258, 227

Schreiber, A., & Klahr, H. 2018, The Astrophysical Journal, 861, 47

Sengupta, D., Dodson-Robinson, S. E., Hasegawa, Y., & Turner, N. J. 2019, The Astrophysical Journal, 874, 26

Shakura, N. I., & Sunyaev, R. A. 1973, Astronomy & Astrophysics, 24, 337

Shi, J.-M., & Chiang, E. 2013, The Astrophysical Journal, 764, 20

—. 2014, The Astrophysical Journal, 789, 34

Shi, J.-M., Zhu, Z., Stone, J. M., & Chiang, E. 2016, Monthly Notices of the Royal Astronomical Society, 459, 982

Stamatellos, D., & Inutsuka, S.-i. 2018, Monthly Notices of the Royal Astronomical Society, 477, 3110

Stamatellos, D., & Whitworth, A. P. 2009, Monthly Notices of the Royal Astronomical Society, 392, 413

Stammler, S. M., Birnstiel, T., Panić, O., Dullemond, C. P., & Dominik, C. 2017, Astronomy & Astrophysics, 600, A140

Steinacker, J., Andersen, M., Thi, W.-F., et al. 2015, Astronomy & Astrophysics, 582, A70

Stephan, M., & Docter, J. 2015, Journal of large scale computing facilities, 1, A1

Stolker, T., Dominik, C., Avenhaus, H., et al. 2016, Astronomy & Astrophysics, 595, A113

Stone, J. M., & Norman, M. L. 1992, The Astrophysical Journal Supplement, 80, 753

Teague, R., Guilloteau, S., Semenov, D., et al. 2016, Astronomy & Astrophysics, 592, A49

Thorngren, D. P., Fortney, J. J., Murray-Clay, R. A., & Lopez, E. D. 2016, The Astrophysical Journal, 831, 64

Tobin, J. J., Chandler, C. J., Wilner, D. J., et al. 2013, The Astrophysical Journal, 779, 93

Tobin, J. J., Kratter, K. M., Persson, M. V., et al. 2016a, Nature, 538, 483

Tobin, J. J., Looney, L. W., Li, Z.-Y., et al. 2016b, The Astrophysical Journal, 818, 73

Tomida, K., Okuzumi, S., & Machida, M. N. 2015, The Astrophysical Journal, 801, 117

Tomley, L., Cassen, P. M., & Steiman-Cameron, T. Y. 1991, The Astrophysical Journal, 382, 530

Tomley, L., Steiman-Cameron, T. Y., & Cassen, P. M. 1994, The Astrophysical Journal, 422, 850

Toomre, A. 1964, The Astrophysical Journal, 139, 1217

Toro, E. F., Spruce, M., & Speares, W. 1994, Shock Waves, 4, 25

Truelove, J. K., Klein, R. I., McKee, C. F., et al. 1997, The Astrophysical Journal Letters, 489, L179

Turk, M. J., Oishi, J. S., Abel, T., & Bryan, G. L. 2012, The Astrophysical Journal, 745, 154

Turk, M. J., Smith, B. D., Oishi, J. S., et al. 2011, The Astrophysical Journal Supplement Series, 192, arXiv:1011.3514

van der Marel, N., van Dishoeck, E. F., Bruderer, S., et al. 2013, Science, 340, 1199

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22

van Leer, B. 1974, Journal of Computational Physics, 14, 361

Vandervoort, P. O. 1970, The Astrophysical Journal, 161, 87

Vaytet, N., Commerçon, B., Masson, J., González, M., & Chabrier, G. 2018, Astronomy & Astrophysics, 615, A5

Vigan, A., Bonavita, M. A., Biller, B. A., et al. 2017, Astronomy & Astrophysics, 603, A3

Vorobyov, E. I. 2013, Astronomy & Astrophysics, 552, A129

Vorobyov, E. I., & Basu, S. 2005, The Astrophysical Journal, 633, L137

—. 2006, The Astrophysical Journal, 650, 956

Vorobyov, E. I., Lin, D. N. C., & Guedel, M. 2015, Astronomy & Astrophysics, 573, A5

Wang, H. H., Klessen, R. S., Dullemond, C. P., van den Bosch, F. C., & Fuchs, B. 2010, Monthly Notices of the Royal Astronomical Society, 407, 705

Weidenschilling, S. J. 1977, Monthly Notices of the Royal Astronomical Society, 180, 57

Williams, J. P., & Cieza, L. A. 2011, Annual Review of Astronomy and Astrophysics, 49, 67

Wünsch, R., Walch, S., Dinnbier, F., & Whitworth, A. P. 2018, Monthly Notices of the Royal Astronomical Society, 475, 3393

Wurster, J., Bate, M. R., & Price, D. J. 2018, Monthly Notices of the Royal Astronomical Society, 480, 4434

Yang, C.-C., & Johansen, A. 2016, The Astrophysical Journal Supplement Series, 224, 39

Yang, C.-C., & Krumholz, M. R. 2012, The Astrophysical Journal, 758, 48

Youdin, A. N., & Johansen, A. 2007, The Astrophysical Journal, 662, 613

Young, M. D., & Clarke, C. J. 2015, Monthly Notices of the Royal Astronomical Society, 451, 3987

Zhu, Z., Hartmann, L. W., Nelson, R. P., & Gammie, C. F. 2012, The Astrophysical Journal, 746, 110

Zhu, Z., Stone, J. M., Rafikov, R. R., & Bai, X.-N. 2014, The Astrophysical Journal, 785, 122

"*Eppur si muove.*"

"And yet it moves."

Galileo Galilei (apocryphal)

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Declaration:

I hereby declare that this thesis is my own work and that I have used no other than the stated sources and aids.

Heidelberg, the 21st of May 2019 .....