From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

Doctoral Dissertations, 2011-2026 Astrophysical Sciences

2022

Formation and evolution of protostellar accretion disks: Towards a quantitative and predictive model

Xu, Wenrui

Princeton, NJ : Princeton University

http://arks.princeton.edu/ark:/88435/dsp018623j196s

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Formation and evolution of protostellar accretion disks: Towards a quantitative and predictive model

Wenrui Xu

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Astrophysical Sciences
Adviser: Matthew W. Kunz

September 2022

© Copyright by Wenrui Xu, 2022.

All Rights Reserved

Abstract

The formation and subsequent evolution of accretion disks around young (Class 0/I) protostars are very important events in the formation of stars and planets. Such disks constitute an evolutionary bridge between the earliest stages of star formation, during which a pre-stellar core fragments out of its natal molecular cloud and commences dynamical contraction, and the later (Class II) stage, when an optically visible pre-main-sequence star is surrounded by a protoplanetary disk, with planet formation likely already underway.

In this thesis, we attempt to develop a quantitative and predictive model of protostellar disks through a synergy of simulation and theory. We first develop and perform a 3D radiative non-ideal magnetohydrodynamic (MHD) simulation to investigate the formation and evolution of a young protostellar disk from a magnetized pre-stellar core. The simulation covers the first $\sim$10 kyr after protostar formation, and shows a hot, massive, and weakly magnetized disk that is gravitatinoally unstable. We then use our simulation results and a series of analytic arguments to construct a quantitative and predictive physical picture of Class 0/I protostellar disk evolution from several aspects, including (i) the angular-momentum redistribution in the disk, self-regulated by gravitational instability to make most of the disk marginally unstable; (ii) the thermal profile of the disk, well approximated by a balance between radiative cooling and accretion heating; and (iii) the magnetic-field strength and magnetic-braking rate inside the disk, regulated by non-ideal magnetic diffusion.

Using these physical insights, we build a simple 1D semi-analytic model of disk evolution. This 1D model, when coupled to a computationally inexpensive simulation for the evolution of the surrounding pseudodisk, can be used reliably to predict disk evolution in the Class 0/I phase. This model can also be used for inferring the properties of observed disks by fitting multi-wavelength dust-continuum observations. Using data from a recent observational survey of protostellar disks, we find that the majority of the sample agree well with our model, and that these disks could be significantly more massive than previous observational estimates (which could be biased by high optical depth), with typical disk-to-star mass ratio of order unity.

Acknowledgements

First and foremost I would like to express my gratitude to my advisor Matt Kunz, for patiently guiding me through the jungle of star formation, for teaching me to become a more careful and responsible researcher, for being extremely supportive of my career, and, most importantly, for always encouraging me to pursue whatever I am most passionate about. You set an excellent example of a researcher and of a mentor.

My gratitude extends to my committee members, Eve Ostriker, Jeremy Goodman, and Jim Stone, for your insightful comments and timely advice. Additionally, a lot of what I learned from semester projects with Jeremy and Jim have greatly benefited my later research projects, sometimes in unexpected ways.

I must also thank my undergraduate mentor, Dong Lai, who showed me the fun and beauty of theoretical astrophysics. I am constantly inspired by your ability of cutting through complex problems with simple physical intuitions.

I am also grateful to my collaborators and those who generously invested their time to answer and discuss my (sometimes naive) questions: Baobab Liu, Bruce Draine, Daniel Fabrycky, Geoffroy Lesur, Henrik Latter, John Tobin, Kengo Tomida, Lile Wang, Patrick Sheehan, Xuening Bai, Zhaohuan Zhu, Zhi-Yun Li, and many others. You grant me the ability and confidence to venture through unfamiliar territories.

Beyond that, I greatly appreciate the friendly and welcoming environment fostered by our department. Even though I only get to spend a little time at Princeton due to family reasons and the pandemic, it has been a great pleasure to share the life in Peyton Hall with my fellow graduate students, especially my class cohort, Erin F., Erin K-F., Goni, Lachlan, Scott, and Vassilis. I have also received much help from our departmental staff, especially Polly Strauss and Susan Duncan.

Special thanks to my cat, Eddy, for your innate cuteness, and your (sometimes involuntary) companionship.

Finally, my sincerest gratitude goes to my parents, and my wife Jing. It is my greatest fortune to have you as my family. Your constant love and support makes me feel grounded in this turbulent world, and inspires me to be a better version of myself.

For my family

Contents

Abstract **iii**

Acknowledgements **iv**

1 Introduction **1**

1.1 Background: from pre-stellar cores to protostars 1

1.2 Theory of protostellar disk formation and evolution 3

1.3 Observational constraints on protostellar disk properties 6

1.4 Summary and outline 8

2 How to build a realistic and reliable simulation of protostellar disk formation? **9**

2.1 Initial conditions of the pre-stellar core 10

2.1.1 Temperature profile 10

2.1.2 Density profile 10

2.1.3 Velocity profile 14

2.1.4 Magnetic field profile 15

2.2 Physical ingredients to be modeled 17

2.2.1 Chemistry and non-ideal MHD 18

2.2.2 The importance of chemistry and dust grains 21

2.2.3 Radiation and thermal evolution 24

2.3 Numerical considerations 26

2.3.1 Numerical convergence in existing studies 26

2.3.2 How to achieve numerical convergence in a long-term simulation? 28

2.4 Our simulation setup 29

2.4.1 Simulation overview 29

2.4.2 Initial conditions 30

2.4.3 Physical modules . . . 32
2.4.4 Numerical setup . . . 36

3 Simulation meets theory: assembling a quantitative physical picture 39

3.1 Definitions . . . 40
3.2 Overview of evolution . . . 41
3.2.1 Pre-stellar collapse and disk formation . . . 41
3.2.2 Disk evolution . . . 44
3.3 Angular-momentum budget and transport . . . 48
3.3.1 Mass and angular-momentum budget of the protostar-disk system . . . 48
3.3.2 Angular-momentum transport within the protostar-disk system: GI . . . 52
3.3.3 Size of the gravitationally unstable region . . . 54
3.4 Energy budget and disk temperature profile . . . 58
3.4.1 Rate of heating and cooling . . . 58
3.4.2 Relation between cooling rate and disk temperature . . . 58
3.4.3 Convection driven by spirals . . . 62
3.5 Magnetic field evolution . . . 64
3.5.1 Magnetic decoupling and field strength in the disk . . . 64
3.5.2 Magnetic braking in the disk . . . 69

4 A semi-analytic model of disk evolution 72

4.1 A semi-analytic 1D disk model . . . 73
4.1.1 Basic assumptions . . . 73
4.1.2 Local model for temperature and column density . . . 73
4.1.3 Estimating disk size and disk profile . . . 74
4.1.4 Summary and discussion . . . 75
4.2 1D+pseudodisk model: evolving the 1D model with a pseudodisk simulation . . . 76
4.2.1 Hierarchical evolution and simulating the pseudodisk . . . 76
4.2.2 Coupling the 1D model to pseudodisk simulation . . . 77
4.3 Comparison between the 1D+pseudodisk model and the 3D model . . . 78

5 Theory meets observation: testing our model on Orion protostellar disks 82

5.1 Model setup . . . 84
5.1.1 Physical assumptions . . . 84

5.1.2 Generating the disk profile . . . 85
5.1.3 Generating mock observation . . . 86
5.1.4 Dust opacity and disk truncation . . . 87
5.1.5 Choice of model parameters . . . 89
5.2 Fitting the model to observation . . . 91
5.2.1 Sample selection . . . 91
5.2.2 Fitting . . . 91
5.2.3 Quantifying systematic uncertainties . . . 92
5.3 Testing the model against observation . . . 92
5.3.1 Agreement between model and observation . . . 92
5.3.2 Is the good agreement coincidental? . . . 94
5.3.3 Does observation favor gravitationally self-regulated disks? . . . 97
5.4 Disk properties and physical implications . . . 101
5.4.1 Summary of disk properties and comparison with previous studies . . . 101
5.4.2 Implication on disk evolution . . . 102
5.4.3 Fragmentation . . . 104
5.5 Comparison with previous studies . . . 105
5.5.1 Summary of models . . . 106
5.5.2 Why can models get different results from the same observation? . . . 106
5.5.3 Which model should I trust? . . . 107
5.6 Discussion . . . 107
5.6.1 Visibility of GI-induced disk substructures . . . 107
5.6.2 Is the 9-mm flux contaminated by free-free emission? . . . 109
5.6.3 Uncertainties in the physical picture and other scenarios of disk formation . . . 111

6 Summary and outlook **112**

6.1 Summary . . . 112
6.1.1 Simulating disk formation . . . 112
6.1.2 The physical picture of disk formation and evolution . . . 112
6.1.3 Modeling protostellar disk with a semi-analytic model . . . 113
6.1.4 Observed properties of protostellar disks . . . 114
6.2 Open questions and possible future works . . . 115

6.2.1 Cloud/core scale ($\gtrsim$ 0.1 pc): whether cores are isolated, and how the main accretion phase ends . . . 116
6.2.2 Envelope/pseudodisk scale (100 au $\lesssim r \lesssim$ 0.1 pc): quantitative and predictive modeling of magnetic braking and the Hall effect . . . 117
6.2.3 Outer disk (few au $\lesssim r \lesssim$ 100 au): magnetic braking, outflow, and fragmentation . . . 117
6.2.4 Inner disk ($\lesssim$ few au): disk truncation, magnetic flux problem, and variability . . . 118
6.2.5 Application of semi-analytic disk model to planet formation: theoretical predictions and observational constraints on grain evolution . . . 120

A Chemical network and reaction rates **122**
A.1 Neutral collision timescales . . . 122
A.2 Species and reactions at low temperature . . . 122
A.3 Reactions at high temperature . . . 123
A.4 CR shielding and isotopic ionization . . . 125

B Angular momentum diffusion at the grid scale **127**

C Details of numerical setup **130**
C.1 A test of our self gravity solver . . . 130
C.2 Handling discontinuity in interpolation tables . . . 130
C.3 Optical depth estimate . . . 131
C.4 A test of our radiative cooling model . . . 133
C.5 Velocity and diffusivity caps . . . 135
C.6 Polar averaging . . . 136
C.6.1 Polar averaging algorithm in XK21a . . . 137
C.6.2 Polar averaging algorithm in XK21b and this thesis . . . 137
C.7 Polar correction for EMF . . . 138
C.7.1 Problem description and possible origin . . . 138
C.7.2 Polar EMF correction algorithm in XK21a . . . 138
C.7.3 Polar EMF correction algorithm for future studies . . . 140

D Numerical convergence **141**
D.1 Resolution . . . 141
D.2 The effect of having a 2π ϕ domain . . . 145
D.3 Inner boundary size and boundary condition . . . 145

E Detailed estimate of disc thermal budget **148**

E.1 Definitions 148

E.2 Basic equations 149

E.3 Perturbation amplitude estimates 151

F Testing the analytic estimate of energy transport by spiral convection **154**

G Are dust grains well-coupled to gas? **156**

H Systematic uncertainties in fitting observed protostellar disks **158**

I Scalings of radial disk profile and disk mass **160**

Bibliography **162**

Chapter 1

Introduction

The formation and subsequent evolution of protostellar disks – rotationally supported accretion disks around young stellar objects (YSOs) still embedded in their infalling natal envelopes – are important events in the formation of stars and planets. In the context of star formation, protostellar disks regulate accretion onto young stellar objects and ultimately affect stellar properties such as mass and initial magnetic flux; in the context of planet formation, they set the environment in which the earliest stages of planet formation (e.g., the coagulation of dust grains) take place, and establish the initial conditions for subsequent protoplanetary disk evolution. There has been a long history of theoretical studies of protostellar disk formation, and yet the many physical ingredients involved in this problem present myriad complications that make it difficult to predict quantitatively the outcome of disk formation. In this thesis, we revisit the problem of protostellar disk formation and attempt to develop a predictive and quantitative model of protostellar disks through a synergy of numerical simulation and analytic theory. We also test our model against recent observations of protostellar disks.

1.1 Background: from pre-stellar cores to protostars

The problem of star formation can be divided into two (interconnected) aspects. One is the "macrophysics" of star formation, which covers the formation and evolution of giant molecular clouds and the formation of overdensities that can collapse under its own gravity (pre-stellar cores) in a molecular cloud; the other is the "microphysics" of star formation, which covers the formation of a protostar (or multiple protostars) from a collapsing pre-stellar core. These two aspects are separated by a characteristic length scale of ~ 0.1 pc, which is the typical size of a pre-stellar core. In this thesis we restrict our attention to the microphysics of star formation, which is more directly relevant to the formation and evolution of protostellar disks. In this

section we give a brief overview of the (relatively well-established) physical picture of how protostatrs form in pre-stellar cores (also see reviews in Shu et al. 1987; Mouschovias & Ciolek 1999; McKee & Ostriker 2007).

Pre-stellar evolution: the formation of stars begins with the formation of dense pre-stellar cores in a turbulent, magnetized molecular cloud. A pre-stellar core is an overdensity that is massive enough so that thermal pressure and magnetic tension can no longer prevent its collapse under its own gravity. Pre-stellar cores have been observed extensively in molecular line emission (Myers & Benson, 1983; Jijina et al., 1999; Caselli et al., 2002), infrared absorption (Teixeira et al., 2005; Machaieie et al., 2017), and sub-millimeter dust emission (Ward-Thompson et al., 1994; Kirk et al., 2005; Könyves et al., 2015), which provide information on their mass, size, temperature, and rotation. The morphology and strength of the magnetic fields have also been inferred via observations of Zeeman splitting (e.g., Crutcher, 1999; Falgarone et al., 2008) and dust polarization (e.g., Ward-Thompson et al., 2000; Crutcher et al., 2004; Girart et al., 2006; Maury et al., 2018; Auddy et al., 2019). Generally, pre-stellar cores are found to be $\sim$0.1 pc in size, 10^4–10^5 cm^{-3} in density, $\sim$M$_\odot$ in mass (with an IMF similar to the stellar IMF but shifted to higher mass), and with slightly supercritical mass-to-flux ratio and a low level of rotation (with ratio between kinetic and thermal energy $\beta_{\rm rot} \sim 10^{-2}$); see a more detailed review in Section 2.1.

The collapse of the pre-stellar core generally produces a nearly self-similar profile, where the outer part of the core takes a r^{-1} column density profile (or r^{-2} density profile) which plateaus at small radii. As the collapse proceeds, the central density increases and this plateau shrinks. This profile has been predicted by theoretical models covering a wide range of collapse scenarios (Larson, 1969; Penston, 1969; Shu, 1977; Hunter, 1977; Basu, 1997), and is also consistent with observations (Alves et al., 2001; Teixeira et al., 2005). Meanwhile, the magnetic field threading the core is advected by the collapsing material to become hourglass shaped; when the magnetic field is dynamically important, the core also flattens along the fieldline to form a pseudodisk (Fiedler & Mouschovias, 1993; Galli & Shu, 1993; Girart et al., 2006; Tassis, 2007; Stephens et al., 2013). In Chapter 3 we provide a more quantitative discussion of the pre-stellar collapse.

Formation of the protostar: Eventually, the central part of the core becomes dense enough to be optically thick, and material there can no longer cool efficiently. This leads to the formation of a pressure-supported first hydrostatic core (FHSC), which marks the formation of the protostar. The FHSC typically forms at a number density of $n_{\rm n} \sim 10^{14}$ cm^{-3}, with an initial size of a few au (Larson, 1969; Kunz & Mouschovias, 2010). As the protostar continues to accrete mass, it shrinks in size and becomes effectively a point mass.

Evolution stages after protostar formation and observational classification: While the formation of a point-mass protostar happens quickly after FHSC formation, accretion onto the protostar persists for a much longer period of time. When the star first forms, most of the mass of the pre-stellar core is still

in the envelope (or pseudodisk) around the star, and the infall of this envelope drives protostar accretion. The stage when protostar accretion is mainly driven by this envelope infall is often referred to as the main accretion phase, as the star likely gains most of its mass during this phase. The main accretion phase ends when the envelope disperses (via accretion onto the central protostar, removal by outflow or turbulence, or competitive accretion from nearby protostars). Observationally, the main accretion phase lasts for $\sim 10^5$ yr (Andre et al., 2000); this implies a typical accretion rate of order 10^{-5} $M_\odot$/yr. After the dispersal of the envelope, the circumstellar disk can still persist for another 1 – 10 Myr (see a review in Williams & Cieza, 2011), and active accretion still occurs in the earlier part of this stage (T Tauri stage), but at a much lower rate ($\sim 10^{-8}$ $M_\odot$/yr, cf. Gullbring et al. 1998).

Observationally, YSOs are usually divided into Class 0, I, II, and III (Lada, 1987; Andre et al., 1993). Loosely speaking, Class 0 and I correspond to the main accretion phase (before envelope dispersal), Class II the T Tauri phase, and Class III the disk dispersal phase. Some studies also include a Flat Spectrum class between Class I and Class II (Greene et al., 1994). Traditionally, Class 0/I, Class II and Class III can be distinguished by the spectral index between 2 and 20 μm, with smaller spectral index towards later stages. Class 0 and Class I can be distinguished by the ratio between the bolometric luminosity and the luminosity above sub-mm wavelengths, which implies whether the luminosity is dominated by the contribution from the envelope (Class 0) or the central protostar (Class I). The exact boundaries between classes, however, could vary between different studies and are mostly empirical (see Evans et al., 2009, Section 5.1 for a review and Section 6 for further discussion). Statistics of observed YSOs suggest a lifetime of $1 - 3 \times 10^4$ yr for Class 0, $1 - 2 \times 10^5$ yr for Class I, and 1 – 10 Myr for Class II and III (Andre et al., 2000; Williams & Cieza, 2011).

In this thesis, our focus is on disk formation and evolution during the main accretion phase (Class 0/I), and we use the word "protostellar disks" to refer exclusively to disks in this phase. Meanwhile, the word "protoplanetary disks" will be used exclusively for Class II disks. This distinction is not a mere choice of terminology; the physical picture of disk evolution in Class 0/I is likely different from that in Class II (and III) because of the rapid accretion from the infalling envelope.

1.2 Theory of protostellar disk formation and evolution

Fundamentally, the formation of a rotationally supported circumstellar disk is a natural consequence of the pre-stellar core hosting a finite amount of angular momentum; this idea was already known to 18-century scholars like Kant and Laplace (see a review in Woolfson 1993). However, developing a more predictive and quantitative theory for prostellar disk formation and evolution remains challenging. Here we provide an overview of the different aspects of this problem in order to demonstrate its complexity; many of these

aspects have been explored extensively in the last couple of decades, mainly through numerical simulations (see a review in Zhao et al. 2020a Section 3.2). This overview also serves as a conceptual framework for how the formation and evolution of a protostellar disk will be analyzed in this thesis from the perspective of angular momentum budget and transport.

To begin with, there are a number of ways to provide the initial angular-momentum budget of the pre-stellar core. Angular momentum can be inherited from the rotation of the pre-stellar core, injected by turbulence in the molecular cloud (Seifried et al., 2012; Joos et al., 2013), and/or generated during the infall of a core with an asymmetric density profile (Verliat et al., 2020).

Next we need to relate this initial angular-momentum budget of the pre-stellar core to the angular-momentum budget of the protostellar disk. In hydrodynamic simulations, angular momentum is conserved, resulting in large, massive disks (e.g., Bate, 2018). In reality, however, the presence of a magnetic field leads to magnetic braking, which can remove most of the initial angular momentum of a pre-stellar core (Mouschovias & Paleologou, 1980; Basu & Mouschovias, 1994). The efficacy of magnetic braking depends on the strength and orientation of the initial magnetic flux (e.g., Basu & Mouschovias, 1995; Joos et al., 2012; Li et al., 2013; Masson et al., 2016; Tsukamoto et al., 2018; Hirano et al., 2020), as well as on the strength of the non-ideal magnetic diffusivity (e.g., Mouschovias & Paleologou 1986; Königl 1987; Li et al. 2011; Dapp et al. 2012), which decouples the magnetic field from the poorly ionized gas. Generally, magnetic braking is weaker when the initial magnetic field is energetically weak, the misalignment between the magnetic field and the rotation axis is significant, and/or the magnetic diffusivity is strong. These complex dependencies are compounded by the fact that, amongst the three types of non-ideal effects, Ohmic dissipation and ambipolar diffusion reduce the efficiency of the braking, whereas the Hall effect can either decrease or increase the angular momentum of the infalling gas depending on whether the magnetic field is aligned or anti-aligned with the rotation axis and which species is the dominant current carrier (Wardle & Ng, 1999; Wardle, 2004; Braiding & Wardle, 2012; Tsukamoto et al., 2015; Zhao et al., 2020b). In cores that are not initially rotating, the Hall effect can even induce rotation, although realistic values of Hall diffusion in dense cores may not be large enough to produce rotationally supported disks (Krasnopolsky et al., 2011). The exact strength of the magnetic diffusivity is also important, and depends on the level of ionization, which is related to the cosmic-ray (CR) ionization rate as well as the abundance and size distribution of dust grains (Zhao et al., 2018b). Low CR ionization, less dust mass, and the absence of very small dust grains generally increase magnetic diffusivity. The initial dust profile may be inferred from observations, but during the collapse one generally expects this distribution to change as grains coagulate and drift through the gas at size-dependent velocities (e.g., Guillet et al., 2020). However, there is only one protostellar disk formation simulation to date that includes (part of) this kind of active dust evolution (Lebreuilly et al., 2020), and the more common

practice is to use a (somewhat arbitrarily) modified dust size distribution with small grains removed (e.g., Zhao et al., 2018a).

Finally, in order to translate the angular-momentum budget to the disk properties (size, mass, radial profiles of surface density and temperature), we need to understand angular-momentum transport in the disk. There are several transport mechanisms that might be important. Magnetic braking (as discussed in the previous paragraph) and outflow (e.g., Machida & Hosokawa, 2013; Tomida et al., 2017) can remove angular momentum from the disk, while disk turbulence driven by instabilities such as the magnetorotational instability or gravitational instability (e.g., Vorobyov & Basu, 2007; Tomida et al., 2017; Kawasaki et al., 2021) can transport angular momentum radially outwards within the disk. The latter also allows the disk to grow beyond the centrifugal (or circularization) radius of infalling envelope material, which is where the azimuthal velocity reaches the local Keplerian velocity and is commonly assumed to be what defines the disk size (e.g., Ulrich, 1976). Although angular-momentum transport mechanisms have been studied extensively in the context of protoplanetary disks (see reviews in Armitage, 2011; Turner et al., 2014), they have received far less attention in the context of protostellar disks. This is probably related to a numerical challenge in simulating protostellar disk formation. The steep resolution requirements for accurately capturing the formation of the protostellar disk, which can initially be just a few au in size (comparable to the size of the first hydrostatic core), implies an extremely small timestep due to the short dynamical timescale and the rapid magnetic diffusion at these small radii. Meanwhile, understanding the evolution of the disk throughout the Class 0/I phase probably requires covering a nontrivial fraction of the main accretion phase, or at least until the disk settles into a quasi-steady state (if it does). Either way, this creates a large timescale separation, which limits the spatial resolution used in contemporary simulations; in the innermost few 10 au, an affordable cell size is often comparable to or even greater than the disk scale height. Under-resolving the disk could lead to incorrect transport rates; even worse, there could be unphysical angular-momentum diffusion and dissipation at the grid scale and/or at the inner boundary of the domain (or the central sink). As a result, many studies have demonstrated poor numerical convergence (i.e., the results are sensitive to resolution or numerical treatments of the inner boundary; Machida et al. 2014; Vorobyov et al. 2019; Hennebelle et al. 2020; Wurster & Bate 2019a), suggesting that existing simulations might be unfit for studying angular-momentum transport in the disk.

From this overview we can identify two main challenges in developing a predictive and quantitative theory for prostellar disk formation and evolution. In terms of numerical modeling, it is challenging to include all of the potentially important physical ingredients in a single simulation at a realistic level and to achieve good numerical convergence in long-term simulations. In terms of physical understanding, it is challenging to incorporate all the potentially relevant physics quantitatively into a single model; more precisely, we do

not fully understand what are the main mechanisms regulating the outcome of disk formation, nor are we able to model many of these mechanisms in a predictive and quantitative fashion. The first challenge, at least in part, contributes to the second by making it difficult to extract a complete and reliable physical picture from simulations.

1.3 Observational constraints on protostellar disk properties

Disks are observed nearly ubiquitously in relatively young (Class II dominated) stellar populations (Hernández et al., 2008, Fig. 11), suggesting that disk formation is a universal process in star formation and that disks might also be common among Class 0/I systems. However, observational knowledge of protostellar disks has long been limited by small sample size, as the lifetime of protostellar disks (especially Class 0) is much shorter when compared to other YSO evolutionary stages (e.g., Class II). In the past decade a growing number of Class 0 and Class I disks have been confirmed with a combination of resolved dust continuum emission and Keplerian rotation measurement (see a list of relevant observations in Section 1 of Tobin et al. 2020). While the sample of confirmed disks remains small and might be heavily biased, it reveals a trend that disk-like structures in resolved dust continuum emission around Class 0/I protostars usually, if not always, correspond to rotationally supported disks. This justifies surveys of protostellar disks in dust continuum, which in recent years have expanded greatly the observational sample of protostellar disks.

Surveys of protostellar disks in dust continuum have revealed a number of interesting trends across different wavelengths, evolutionary stages, and star-forming regions that have yet to be fully explained by theory. The VANDAM survey (Segura-Cox et al., 2018) identified 18 Class 0/I disk candidates in the Perseus molecular cloud at 8 mm with ∼12 au resolution. The detection rate of Class 0/I disks is quite low, which could be due to limited sensitivity. The detected disks are generally small and relatively massive, with disk sizes $\lesssim 30$ au and median masses of 0.07 $M_\odot$ in Class 0 and 0.03 $M_\odot$ in Class I (Tychoniec et al., 2018).[1] Andersen et al. (2019) also derived broadly similar disk masses for this sample using ∼mm observations. The CALYPSO survey (Maury et al., 2019) carried out multi-wavelength observation at 1 – 3 mm towards 16 Class 0 systems with 50 – 150 au resolution. They find that Class 0 disks might be common, with 11 of the 16 systems preferring models that include a disk. Meanwhile, large disks are rare, with only 4 systems hosting resolved disks with radii $\gtrsim 60$ au. The ODISEA program (Williams et al., 2019) surveyed 279 systems in the Ophiuchus region ranging from Class I to Class III at 1.33 mm, among which 28 are Class I and 50 Flat spectrum. They find very low protostellar disk masses, with a logarithmic mean Class I disk mass of 0.001 $M_\odot$. They also find a five-fold decrease in disk mass from Class I to Class II, consistent with the trend

[1] All disk mass estimates in this paragraph assume a dust-to-gas ratio of 0.01.

between Class 0 and Class I in the VANDAM survey. Finally, the VANDAM Orion survey (Tobin et al., 2020) observed protostellar systems in Orion with ~ 40 au resolution at 0.87 mm and 9 mm to assemble by far the largest sample of protostellar disks (379 detections with 87% detection rate). Based on the 0.87-mm data, the median disk sizes and masses are 45, 37, and 29 au (with $\lesssim$5 au statistical uncertainty) and 0.008, 0.004, and 0.003 $M_\odot$ for Class 0, I, and Flat Spectrum, respectively. Meanwhile, the 9-mm data suggest disk masses that are approximately five times higher (which may suggest that the 0.87-mm flux is optically thick) and disk sizes that are significantly smaller. This wavelength dependence is broadly consistent with the difference between the 9-mm VANDAM (Perseus) survey and other surveys at shorter wavelengths.

In summary, these studies show several trends:

- Disk mass peaks at Class 0 and decreases throughout later evolutionary stages (all studies);
- Disk size (or at least the apparent disk size) probably also peaks at Class 0 and decreases in Class I and Flat Spectrum (VANDAM Orion), but might start increasing again towards Class II given that surveys find protostellar disks to be generally smaller than young Class II disks (all studies);
- The apparent disk size likely decreases towards longer wavelengths, and the disks might be optically thick at $\sim$mm wavelengths (VANDAM Orion);
- There might be large systematic variation in disk size and mass across different star-forming regions, similar to what has been observed in Class II (cf. Eisner et al., 2018).

Before attempting to explain these trends, we also need to stress some important uncertainties in translating dust continuum observation to estimates of disk properties. The ratio between observed flux and disk mass depends on dust opacity, temperature profile (and the dominant dust heating mechanism), and whether the disk is optically thick at the observed wavelength; all of these remain highly uncertain. Meanwhile, observations provide only the apparent disk size (e.g., width of gaussian fits), and the agreement between apparent and actual disk size depends on the shape of the radial intensity profile, which could bias the comparison between different populations. Finally, we also need to consider whether a constant dust-to-gas ratio is appropriate; significant decoupling and radial drift is known to be common in Class II disks (e.g., Ansdell et al., 2018). We discuss these uncertainties in more detail in Chapter 5.

Overall, the current state of observation makes now a good time for theoretical studies of disk formation, as recent (and upcoming) observations offer an unprecedented amount of data that be used for testing theoretical predictions. Furthermore, a more quantitative and predictive physical picture of disk evolution might lead to new methods that more reliably translate observations to disk properties, and might offer explanations for observed trends in protostellar disk populations.

1.4 Summary and outline

Developing a quantitative and predictive theory of protostellar disk formation and evolution is crucial for studying the formation of stars and planets and for interpreting trends in recent observations of protostellar disk populations. However, it remains challenging to develop physically realistic and numerically reliable simulations of protostellar disk formation, and to identify and quantitatively model the main physical mechanisms that determine the outcome of disk formation.

This thesis attempts to address these challenges through a synergy of simulation and theory, and to develop a quantitative theory that allows the prediction of basic disk properties such as disk size, mass, and the radial profile of surface density and temperature. In Chapter 2 we discuss a series of physical ingredients and numerical issues that should be considered in order to develop a realistic and reliable simulation. We also introduce our simulation setup, including the implementation of various physical ingredients and numerical treatments. In Chapter 3 we present the results of our radiative non-ideal MHD simulation of protostellar disk formation, and compare the simulation results against a series of new and existing analytic arguments to identify and model quantitatively the various physical mechanisms that affect disk evolution. In Chapter 4 we take the physical insights from Chapter 3 to develop a predictive semi-analytic model of protostellar disks, and discuss how this model can be applied to predict long-term disk evolution and to model multi-wavelength disk observation. In Chapter 5 we test our model against multi-wavelength dust continuum observations from a large survey and discuss whether the predictions of our model agree with observed trends across different wavelengths and evolutionary stages. In Chapter 6 we conclude with a summary of main results and an outlook for future works on this subject.

Chapter 2

How to build a realistic and reliable simulation of protostellar disk formation?

Section 2.4 of this chapter is adapted from material published in:

Xu, W. and Kunz, M. W. (2021a). Formation and evolution of protostellar accretion discs – I. Angular-momentum budget, gravitational self-regulation, and numerical convergence. MNRAS, 502(4):4911–4929.

Xu, W. and Kunz, M. W. (2021b). Formation and evolution of protostellar accretion discs – II. From 3D simulation to a simple semi-analytic model of Class 0/I discs. MNRAS, 508(2):2142–2168.

2.1 Initial conditions of the pre-stellar core

The majority of protostellar disk formation simulations use an isolated pre-stellar core as the initial condition. In this section we discuss how such an initial condition should be constructed by reviewing the observational constraints on pre-stellar core properties and discussing the advantages and limitations of initial conditions adopted in existing simulations.

2.1.1 Temperature profile

Observationally, pre-stellar cores have temperatures ~ 10 K. The outer part of the core is generally hotter than the inner part of the core due to shielding of the interstellar radiation field that heats the core. The dust and gas temperatures can also be different because they are not thermally well coupled below a neutral number density of $\sim 10^5$ cm^{-3}. However, these differences are generally below order-unity and dynamically unimportant (di Francesco et al., 2007, Section 3.3). Therefore, simulations generally adopt pre-stellar cores with uniform temperature.

2.1.2 Density profile

Observational constraints

The azimuthally averaged density profiles of pre-stellar cores generally show a central plateau and a power-law fall-off with slope ~ -2 at larger radii (di Francesco et al., 2007, Section 3.2), but the details of the core density profile (central density and size of the plateau) vary significantly across different observational studies. For example, a collection of 10 different samples of pre-stellar cores reviewed in Ward-Thompson et al. (2007, Fig. 2) have mean densities ranging from $\sim 2 \times 10^3$ cm^{-3} to $\sim 4 \times 10^5$ cm^{-3}. Although systematic errors in how core properties are estimated could contribute to this difference, the diversity between different star-forming regions is probably the dominating factor. This is clearly demonstrated in the more recent Herschel Gould Belt Survey (André et al., 2010), which features a set of complete surveys of cores in many nearby star-forming regions using the same instrument and data pipeline. Derived core properties from part of this survey are shown in Fig. 2.1; we see that the typical central density is usually several 10^4 cm^{-3}, with a large spread similar to that found in the review by Ward-Thompson et al. (2007, Fig. 2).

Another important constraint on the properties of pre-stellar cores is that most bound cores appear to be in a near-critical balance between self-gravity and thermal pressure. For example, in Fig. 2.1 typical cores

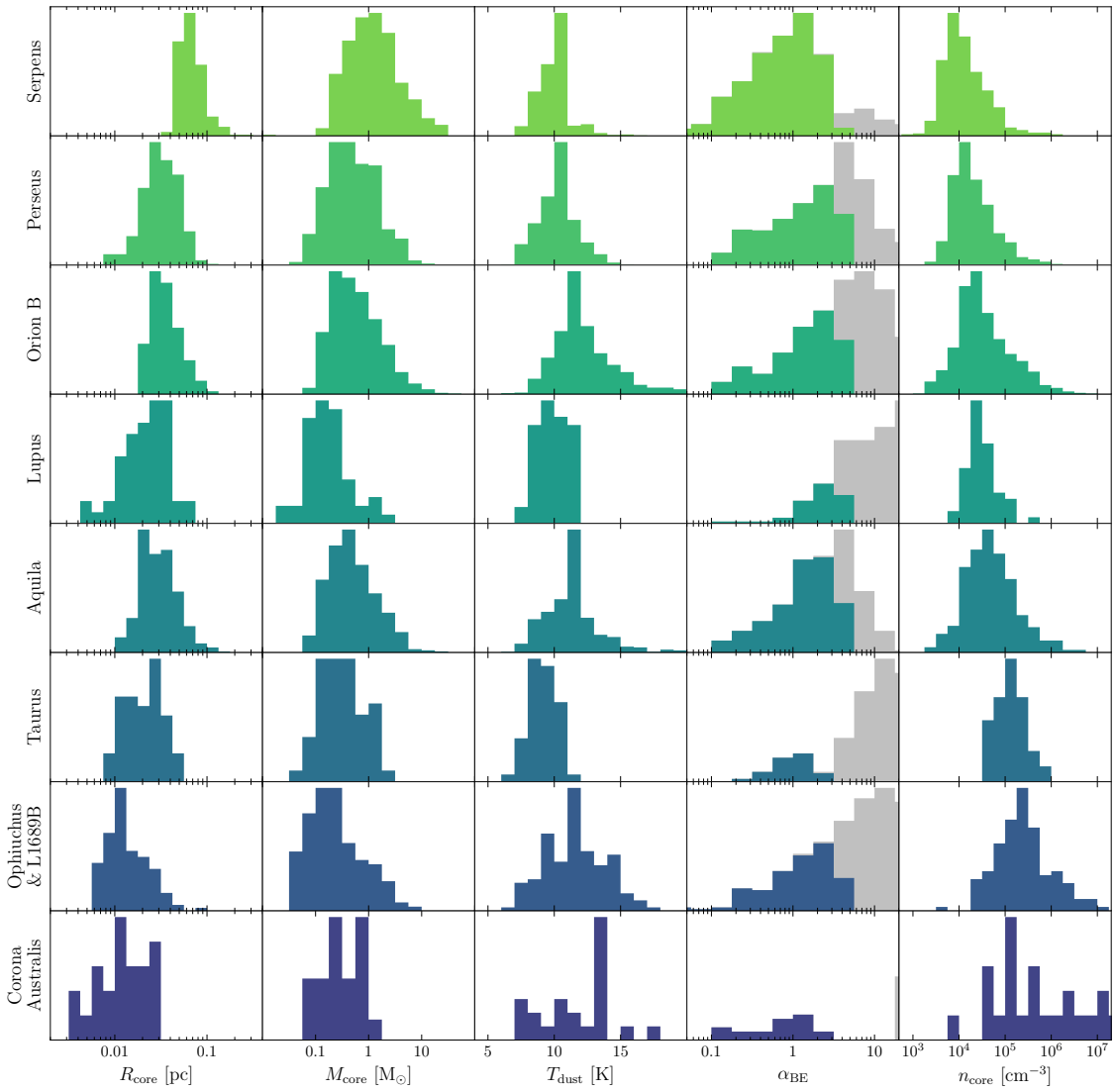

Figure 2.1: Pre-stellar core properties in different star-forming regions from the Herschel Gould Belt Survey (HGBS, André et al., 2010). Data are taken from the HGBS archive (http://www.herschel.fr/cea/gouldbelt/en/Phocea/Vie_des_labos/Ast/ast_visu.php?id_ast=66); see references therein. We only include datasets where a formatted table for derived core properties is available. Form left to right, columns show core size $R_{\rm core}$, mass $M_{\rm core}$, dust temperature $T_{\rm dust}$, ratio between Bonnor–Ebert mass (Eq. 2.1) and core mass $\alpha_{\rm BE}$, and inferred core mean density $n_{\rm core}$. Rows are sorted by ascending core density. In the $\alpha_{\rm BE}$ column, grey histograms show all starless cores, and colored histograms show bound (pre-stellar) cores only. All other columns show bound cores only. Typical cores are near critical in all regions. The core properties and the systematic differences between different populations are similar to those found in previous studies.

have $\alpha_{\rm BE} \sim 1$, where $\alpha_{\rm BE}$ is the ratio between the Bonnor–Ebert mass corresponding to the core size $R_{\rm core}$

$$M_{\rm BE} = \frac{2.4 R_{\rm core} k_{\rm B} T}{\mu m_{\rm p} {\rm G}}, \tag{2.1}$$

and the core mass. Here $M_{\rm BE}$ is the minimum mass required for collapse (in the absence of additional non-thermal support) and differs from the Jeans mass by a factor of order unity. (A more common choice is to define $M_{\rm BE}$ at a given ambient pressure; but here we follow the definition used in this survey.) Near-critical cores are also consistent with statistical estimates of core lifetimes, which show a core lifetime between 2 and 5 free-fall times Ward-Thompson et al. (2007, Section 3).

Finally, we stress that the observed pre-stellar core population includes pre-stellar cores that have undergone various degrees of collapse, yet for the initial condition of a simulation we prefer a core that is just beginning to collapse. Still, typical cores in the observed population should be a good approximation for cores that are just beginning to collapse. This is because during the collapse of a pre-stellar core, the timescale to increase the central density by a given factor decreases as the central density increases, and the core spends much more time in the early stages of the collapse (when the density profile is still similar to that at the beginning of the collapse) than the later stages (when the density profile has become significantly different). For an initially magnetically subcritical core, this contrast can be more significant (e.g., Basu & Mouschovias, 1994). Therefore one can use the typical properties of observed cores (perhaps with a slight reduction in density and an increase in size) as the initial condition of simulations.

Initial density profiles adopted in simulations

Two types of initial density profiles commonly adopted in recent simulations are an isolated sphere with uniform density (e.g., Li et al., 2011; Zhao et al., 2018a; Hennebelle et al., 2020) and an enhanced Bonnor–Ebert sphere in which the ratio between gravity and pressure gradient is a constant greater than 1 (e.g., Machida et al., 2018; Wurster & Bate, 2019b; Machida & Basu, 2019, 2020). In both cases the parameters of the profile are usually chosen such that the total mass in the density profile is $\sim {\rm M}_\odot$, in order to match the typical mass of observed cores (cf. Fig. 2.1). Both choices are simple and offer some flexibility on tuning the relative importance between pressure and gravity, which allows simulations to reproduce the observed near-critical cores. However, these assumed density profiles differ from typical observed cores in ways that might affect the evolution.

Observationally, the density profile around a pre-stellar core is flat at the center and falls off at approximately $n \propto r^{-2}$ at larger radii; and the core size is generally defined to be comparable to the size of the central plateau (e.g., using FWHM of flux or column density profile). In other words, we have a core at

approximately uniform density embedded in an $n \propto r^{-2}$ "envelope". For a uniform isolated sphere, this envelope is missing, and the sudden change in density and pressure leads to artifacts around the boundary of the sphere. The lack of ambient pressure also means that the core has to be more compact to collapse; for a $\sim M_\odot$ core, observations generally give a central density of $\sim 10^4$ cm^{-3} (cf. top two rows of Fig. 2.1), while simulations adopting a $\sim M_\odot$ uniform isolated sphere without ambient pressure support usually require densities $\gtrsim 10^5$ cm^{-3} to collapse. Meanwhile, the shape of an enhanced Bonnor–Ebert sphere would match well with the observed shape of the density profile, as it also has a central plateau surrounded by $n \propto r^{-2}$ falloff. However, simulations adopting enhanced Bonnor–Ebert spheres often try to match the observed core mass with the total mass in the Bonnor–Ebert profile, while a more reasonable choice would be to match the mass contained in the central plateau of the Bonnor–Ebert profile with the observed core mass. The two choices lead to significantly different densities: if we approximate the Bonner-Ebert profile with a broken power law with slope 0 at small radii and -2 at larger radii, the fraction of mass in the central plateau ("core") would be

$$\frac{M_{\rm core}}{M_{\rm total}} \approx \frac{1}{3\sqrt{n_0/n_\infty} - 2}. \tag{2.2}$$

Here n_0 is the central density and n_∞ is the ambient density which determines where this r^{-2} envelope ends. Simulations adopting enhanced Bonnor–Ebert spheres often assume $n_0/n_\infty \gtrsim 10$, corresponding to only $\lesssim 13\%$ of mass in the core. In other words, the "1M$_\odot$" Bonnor–Ebert profiles adopted in simulations are more similar to observed cores with a core mass $\sim 0.1 M_\odot$.

Motivated by the observed core profile (and observational definition of the core), a better option for the initial condition should contain a central plateau with mass $\sim M_\odot$, surrounded by a $n \propto r^{-2}$ envelope at larger radii. For example, to model a core with characteristic column density Σ_0 and size R_0, Dapp et al. (2012) in their 1D (flattened core) calculation adopted

$$\Sigma = \frac{\Sigma_0}{\sqrt{1 + (R/R_0)^2}}, \tag{2.3}$$

where R denotes the cylindrical radius. The drawback of this choice is also apparent. There would be a large amount of mass outside of the core ($\sim 30 M_\odot$ in Dapp et al. 2012), and if the simulation continues for sufficiently long time all this mass may eventually be accreted by the protostar. (In reality, collapse of the envelope around the core would probably be prevented by the magnetic field, turbulence, or competitive accretion; the latter two are generally not included in single-core simulations.) Still, this is less of a problem for simulating the formation and early evolution of protostellar disks because it only begins to affect the result at relatively late times. Material initially at $\gtrsim R_0$ reaches the protostar at a time $\gtrsim$Myr after point-mass

formation, if we estimate this time using the free-fall timescale for a central mass of 1 $M_\odot$ and $R_0 = 0.1$ pc. And once we simulate the system for such a long timescale, using an isolated core as initial condition becomes inappropriate anyway, unless the interaction with the surrounding environment can be modeled.

A note on asymmetry

So far our discussion only considers cores with a symmetric (axisymmetric or spherically symmetric) density distribution. In reality the core might be asymmetric (i.e., not having any rotational or reflectional symmetry), which allows angular momentum to be generated during its collapse (Verliat et al., 2020). However, observational characterizations of core geometry have mainly be focusing on modeling cores as oblate, prolate, or triaxial spheriods (e.g., Tassis, 2007), and asymmetries that may lead to the generation of angular momentum during collapse have not been characterized quantitatively.

2.1.3 Velocity profile

The velocity inside a pre-stellar core can be roughly divided into three components: infall, bulk rotation, and turbulent motion. As simulations generally use pre-stellar cores that just begin to collapse as their initial condition, the initial infall velocity is usually negligible. Meanwhile, the bulk rotation and turbulent motion can be chosen to match observed cores at an early stage of collapse.

Line-of-sight velocity measurements using molecular lines (Barranco & Goodman, 1998; Goodman et al., 1998; Caselli et al., 2002) show that the core hosts a low level of bulk rotation with $\Omega \sim 1$ km/s/pc and the ratio between kinetic and thermal energy being $\beta_{\rm rot} \lesssim$ few $\times 10^{-2}$. They also find a "transition to coherence", where the core shows a coherent length scale of ~ 0.01 pc (similar to the ambipolar cutoff scale of MHD turbulence; Mouschovias 1991; Caselli et al. 2002) and the level of turbulence in the core is generally subsonic with a size-linewidth relation that deviates from that at larger scales (Goodman et al., 1998). More recent studies (see a review in Ward-Thompson et al. 2007 Section 6 and Fig. 6) also find low level of turbulence with subsonic or transonic linewidths for single-star-forming cores (both isolated and within cluster-forming cores). However, the low level of turbulence does not imply that turbulence is negligible compared to bulk rotation. The ~ 0.01 pc coherence scale still allows the outer part of the core and the envelope around the core to be dominated by turbulent velocity. The linewidth measurements for single-star-forming cores (Ward-Thompson et al., 2007, Fig. 6) also show a large spread (both within and across samples), suggesting that cores with turbulent kinetic energy $\beta_{\rm turb} \lesssim \beta_{\rm rot}$ and those with $\beta_{\rm turb} \gtrsim \beta_{\rm rot}$ may both be common.

Despite these complexities, a common choice in simulations is to assign uniform rotation inside the core with $\beta_{\rm rot} \sim 10^{-2}$ to match observations. While this would yield a realistic level of rotation when the core

density profile is also realistic, we stress that $\beta_{\rm rot}$ may not be the best parameter to describe core rotation in the context of disk formation. Since disk formation is fundamentally a result of finite core angular momentum, one would expect the disk formation outcome to mainly depend on the angular momentum profile $L_z(M)$ (i.e., angular momentum within a given radius as a function of mass within that radius) as opposed to the total kinetic energy. For example, if the core has a flat density profile and uniform rotation at small radii, we expect $L_z(M) \propto M^{5/3}$ and the angular momentum budget can be characterized by the constant $L_z(M)/M^{5/3} \propto \Omega_0 n_0^{-2/3}$ where Ω_0 is core rotation and n_0 is core central density. This type of parameter would be helpful for comparing disk-formation outcomes across cores with different sizes and masses.

Some simulations have also attempted to include the effect of turbulence, but there are no clear conclusions on how (and at what range of $\beta_{\rm turb}$) turbulence affects the outcome of disk formation. It has been demonstrated that including transonic or supersonic turbulence helps disk formation (e.g., Seifried et al., 2012; Joos et al., 2013), yet the exact reason for this result is less well understood. It is conjectured that turbulence could reduce the effect of magnetic braking by perturbing the otherwise coherent fieldlines, or by causing misalignment between the angular momentum and the large-scale magnetic field. Another possibility is that turbulence simply provides more rotation. Turbulence adopted in these simulations often corresponds to turbulent kinetic energy with $\beta_{\rm turb} \gtrsim 0.1$ and is sometimes accompanied by bulk rotation with a similar value for $\beta_{\rm rot}$, which is much higher than typical $\beta_{\rm rot}$ adopted in simulations with uniform rotation only. It also remains unclear whether the spectrum of turbulence assumed in these studies (usually isotropic, Kolmogorovian, and without a cutoff scale) is reasonable, given the relatively large (~ 0.01 pc) observational coherence scales.

Due to the challenges in modeling turbulence and interpreting its effect, in this thesis we limit ourselves to the simple case of a non-turbulent core. Given the observational evidence that a nontrivial fraction of single-star-forming cores have highly subsonic and probably dynamically unimportant turbulence (Ward-Thompson et al., 2007, Fig. 6) while disk formation is ubiquitous, a non-turbulent setup is likely still sufficient for understanding the basics of disk formation.

2.1.4 Magnetic field profile

Field strength and geometry

Observations of pre-stellar cores usually show ordered, straight fields, whose mean direction is statistically consistent with being oriented parallel to the minor axis of the core in 3D. This is consistent with a collapsing core threaded by a dynamically important magnetic field. Some pre-stellar cores also show hourglass-shaped

fieldlines, consistent with an initially more uniform field that has been pinched radially by the core collapse (see a recent review of relevant observations in Section 5.1.1 of Pattle et al., 2022). The lack of turbulent magnetic field at (and below) core scales is consistent with the theoretical argument that ambipolar diffusion stops the propagation of Alfvén waves and straightens fieldines below a length scale of (Mouschovias, 1991)

$$\lambda_{\rm AD} \equiv \pi v_{\rm A} \tau_{\rm ni} \approx 0.03 \left(\frac{B}{30\ \mu{\rm G}}\right) \left(\frac{n_{\rm n}}{10^4\ {\rm cm}^{-3}}\right)^{-3/2} \left(\frac{\chi_{\rm i}}{10^{-7}}\right)^{-1}\ {\rm pc}. \tag{2.4}$$

Here $v_{\rm A}$ is the Alfvén speed, $\tau_{\rm ni}$ is the neutral-ion collisional timescale, B is the magnetic field strength, $n_{\rm n}$ is the neutral number density, and $\chi_{\rm i} = n_{\rm i}/n_{\rm n}$ is the ionization fraction.

The strength of the magnetic field threading the core is often characterized by the dimensionless mass-to-flux ratio μ_Φ, which is normalized by the critical mass-to-flux ratio for gravity to overcome magnetic tension in the central flux tubes of the core, $(3/2)(63G)^{-1/2}$ (Mouschovias & Spitzer, 1976). Observational estimates of core magnetic-field strengths range from ~ 10 μG to a few $\times 10^2$ μG, with the mass-to-flux ratio ranging from mildly supercritical $\mu_\Phi \sim 2$–3 to somewhat subcritical $\mu_\Phi \sim 0.1$–0.4 (Pattle et al., 2022, Section 5.1.2 and Fig. 2).

Overall, it would be reasonable to model the initial magnetic field in the core as a uniform field with near critical mass-to-flux ratio, give or take a factor of a few. Higher values of mass-to-flux ratio ($\gtrsim 5$) have also been used in some previous studies (often as part of a parameter space survey), although this regime is probably less relevant for typical low-mass stars.

Alignment with rotation

Given that pre-stellar cores form in a turbulent environment, and that sometimes the core itself may be turbulent, there should not be a strong preference for how magnetic field aligns with rotation. Indeed, misalignment between rotation and magnetic field is found in both observation (Hull et al., 2013, 2014) and simulations of core formation (Chen & Ostriker, 2018; Kuznetsova et al., 2020).

The effect of misalignment between magnetic field and rotation has been investigated in a number of recent simulations; see a detailed review of these simulations by Zhao et al. (2020a), Section 3.2.4. In summary, most simulations find that a misaligned magnetic field enhances disk formation to various degrees, with a few exceptions finding the opposite trend, and several studies find that the effect of misalignment is less pronounced once non-ideal MHD effects are included. Right now there is no consensus on how (and when) misalignment helps disk formation. However, a recent study by Tsukamoto et al. (2018) might offer some insight on this problem. They performed a small parameter survey comparing different misalignments, different levels of pressure support, and whether non-ideal MHD effects are included, and found that the

effect of misalignment could vary both quantitatively and qualitatively (i.e., vary between helping and suppressing disk formation) depending on the other variables. They also attempted to decompose the effect of misalignment into three different mechanisms: preferential accretion (i.e., material along the central flux tube is accreted first, and this flux tube has significantly higher angular momentum when field is misaligned with rotation), magnetic braking during the collapse of the core (which generally increases with misalignment, cf. Mouschovias & Paleologou 1979), and magnetic braking in the disk (or FHSC). While the strength of the last mechanism is difficult to estimate theoretically, using a competition between the first two one could already explain the complex dependence observed in their survey as well as trends observed in previous studies.

2.2 Physical ingredients to be modeled

The pre-stellar core and the protostar-disk system to be formed within it are a complex system that consists of many different species (atoms, molecules, dust grains in both neutral and charged forms) that interact with each other as well as the electro-magnetic and radiation fields. To simplify the problem, a common choice is to ignore the inertia of all charged species given that the system is predominately neutral, and to assume that all species are thermally coupled with approximately the same temperature. These assumptions are generally reasonable for the pre-stellar core and for the protostellar disk, with the caveat that they may become invalid in low-density regions (where thermal collisions are less frequent) such as the outflow cavity. Under these approximations, we can model the system as a single neutral fluid, which couples indirectly to the magnetic field via collisions with charged species (see, for example, the derivation in Kunz & Mouschovias 2009). The evolution of the system is then governed by the following set of equations:

$$\partial_t \rho + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}) = 0, \tag{2.5}$$

$$\partial_t (\rho \boldsymbol{v}) + \boldsymbol{\nabla} \cdot \left(\rho \boldsymbol{v}\boldsymbol{v} - \frac{1}{4\pi}\boldsymbol{B}\boldsymbol{B} + P^* \right) = -\rho \boldsymbol{\nabla} \Phi, \tag{2.6}$$

$$\partial_t E + \boldsymbol{\nabla} \cdot \left[(E + P^*)\boldsymbol{v} - \frac{1}{4\pi}\boldsymbol{B}(\boldsymbol{B}\cdot\boldsymbol{v}) + \frac{c}{4\pi}\boldsymbol{E}_{\rm nonideal} \times \boldsymbol{B} \right] = Q_{\rm rad}, \tag{2.7}$$

$$\partial_t \boldsymbol{B} - \boldsymbol{\nabla} \times (\boldsymbol{v} \times \boldsymbol{B} - c\boldsymbol{E}_{\rm nonideal}) = 0, \tag{2.8}$$

$$\nabla^2 \Phi = 4\pi \mathrm{G} \rho. \tag{2.9}$$

Here ρ, $\boldsymbol{v}$ and $\boldsymbol{B}$ are the mass density, fluid velocity, and magnetic field. $P^* = p + B^2/8\pi$ is the sum of thermal and magnetic pressures. Φ is the gravitational potential. $E = u + \frac{1}{2}v^2 + \frac{1}{8\pi}B^2$ is the total energy density, with $u = p/(\gamma - 1)$ being the internal energy density and γ being the ratio of specific heats. $Q_{\rm rad}$

is the heating (or cooling) rate due to radiation. Here we only consider the effect of radiation on thermal evolution, and ignore the effect of radiation pressure in Eq. (2.6) on the grounds that radiation pressure is negligible compared to thermal pressure for low-mass star formation. $\boldsymbol{E}_{\rm nonideal}$ is the electric field induced by non-ideal MHD effects, given by

$$\boldsymbol{E}_{\rm nonideal} = \frac{4\pi}{c^2}\left(\eta_{\rm O}\boldsymbol{j} + \eta_{\rm A}\boldsymbol{j}_\perp + \eta_{\rm H}\boldsymbol{j}\times\boldsymbol{b}\right). \tag{2.10}$$

Here $\boldsymbol{j} = \frac{c}{4\pi}\nabla\times\boldsymbol{B}$ is the current density, $\boldsymbol{j}_\perp$ is the component of $\boldsymbol{j}$ perpendicular to $\boldsymbol{B}$, $\boldsymbol{b}$ is the unit vector in the direction of $\boldsymbol{B}$, and $\eta_{\rm O}, \eta_{\rm A}$, and $\eta_{\rm H}$ are non-ideal magnetic diffusivities corresponding to Ohmic dissipation, ambipolar diffusion, and the Hall effect.[1]

There are two key physical ingredients that need to be modeled in order to close these equations. The first is the non-ideal MHD diffusivities $\eta_{\rm O}, \eta_{\rm A}$, and $\eta_{\rm H}$; determining these parameters realistically also requires modeling the chemical evolution of the system. The second is the heating and cooling by radiation, $Q_{\rm rad}$. These two physical ingredients deserve extra attention as they involve complex physics and are usually modeled with various degrees of approximation (and oversimplification). Below we discuss how these physical ingredients are modeled in existing studies, and whether various simplifying assumptions would significantly impact the evolution of the system. The goal of this discussion is to identify the maximum level of simplification one can adopt while still producing reasonably realistic simulation results.

2.2.1 Chemistry and non-ideal MHD

Non-ideal MHD effects result from the decoupling between the neutral and charged species in the mostly neutral core and disk. The contribution of non-ideal MHD effects can be divided into three effects: Ohmic dissipation, ambipolar diffusion, and Hall effect. These three effects differ in how the resulting electric field scales with the current; see Eq. (2.10).

Ohmic dissipation and ambipolar diffusion

In the context of star and disk formation, Ohmic dissipation and ambipolar diffusion both affect the evolution mainly by decoupling the magnetic field from the gas, which increases the mass-to-flux ratio in the central part of the system (the protostar and the disk) and suppresses magnetic braking. For typical conditions in the pre-stellar core and the protostellar disk, Ohmic dissipation is usually weaker than ambipolar diffusion ($\eta_{\rm O} \lesssim \eta_{\rm A}$; see discussion in Section 2.2.2). Therefore, we focus our discussion here on ambipolar diffusion.

[1] In the literature, the symbol η is commonly used for both resistivity with unit s and magnetic diffusivity with unit $\rm cm^2 s^{-1}$. The two differ by a factor of $c^2/4\pi$. In this thesis we always use η for magnetic diffusivity, and $\eta^{\rm R}$ for resistivity.

The effect of ambipolar diffusion is equivalent to a drift velocity between the field and the gas; i.e., the field evolution is identical to that in ideal MHD but with a fluid velocity $\boldsymbol{v} + \boldsymbol{v}_{\rm drift}$. This drift velocity is given by

$$\boldsymbol{v}_{\rm drift,A} = -\eta_{\rm A} \frac{\boldsymbol{B} \times (\nabla \times \boldsymbol{B})}{B^2}. \tag{2.11}$$

The right-hand side of Eq. (2.11) is proportional to the Lorentz force, and physically comes from a balance between the Lorentz force acting on the charged species and the collisonal drag between the neutrals and charged particles. In the context of a collapsing core, the collapse forms hourglass-shaped fieldlines with the Lorentz force and drift velocity pointing outwards and that allows the magnetic field to lag behind the gas during the collapse. This leads to an increased mass-to-flux ratio in the central region of the core, which could explain (partially) why stars contain many orders of magnitude lower magnetic flux than pre-stellar cores.

The decoupling of the magnetic field also directly affects the angular-momentum budget of the disk by reducing the efficacy of magnetic braking (Mouschovias & Paleologou, 1986; Königl, 1987; Li et al., 2011; Dapp et al., 2012). This behavior can be explained qualitatively using the simple model of magnetic braking presented by Basu & Mouschovias (1994). In steady state, the rate at which magnetic braking removes angular momentum per disk (or pseudodisk) area is given by

$$F_L(R) = \rho_\infty R_\infty^2 (\Omega_{\rm disk} - \Omega_\infty) v_{{\rm A},\infty} \frac{B_{z,{\rm disk}}}{B_\infty}. \tag{2.12}$$

Here the subscript ∞ denotes properties at $z \to \infty$ where the background medium is assumed uniform in density, rotation, and magnetic field. $R_\infty \sim R(B_{z,{\rm disk}}/B_\infty)^{1/2}$ is the radius at $z \to \infty$ of the fieldline through $(R, 0)$. Physically, this means that angular momentum is propagated along the flux tube at the Alfvén speed. The decoupling of the magnetic field reduces both the disk magnetic flux density $B_{z,{\rm disk}}$ and the "lever arm" R_∞ (for a given R) in Eq. (2.12). If diffusion is strong enough so that the field is also decoupled in the azimuthal direction, $\Omega_{\rm disk}$ will also be replaced by the rotation rate of the field in the disk which can be lower than the gas rotation. However, we stress that this discussion is only for conceptual understanding of how decoupling reduces magnetic braking; in our simulation the magnetic braking rate could be significantly different from Eq. (2.12), as the collapsing envelope and the disk are not in steady state all the way to $z \to \infty$ (or $\gtrsim 0.1$ pc, which is when the background may be considered as uniform); see discussion in Section 3.5.2.

The Hall effect

The Hall effect plays a more subtle role. As the electric field produced by the Hall effect is also perpendicular to the magnetic field, we can also associate the Hall effect with a drift velocity given by

$$\boldsymbol{v}_{\rm drift,H} = \eta_{\rm H}\boldsymbol{b} \times \frac{\boldsymbol{B} \times (\nabla \times \boldsymbol{B})}{B^2}. \tag{2.13}$$

We have written this drift velocity in a form very similar to Eq. (2.11) to highlight that it produces a drift velocity in the direction of $\boldsymbol{b} \times \boldsymbol{v}_{\rm drift,A}$. This allows the Hall effect to generate rotation in a initially non-rotating core as follows. The hourglass shaped fieldline generated during the collapse provides an azimuthal component of $\boldsymbol{v}_{\rm drift,H}$. Variation of azimuthal $\boldsymbol{v}_{\rm drift,H}$ (which is the azimuthal field velocity for a non-rotating core) along a fieldline shears the magnetic field to cause angular-momentum transport. This can be understood as a modified version of magnetic braking; magnetic braking tends to bring field rotation towards Ω_∞, and with Hall effect it now pushes gas rotation towards $\Omega_\infty - (v_{\rm drift,H})_\phi/R$, which can be nonzero even when the core is initially non-rotating as generating $(v_{\rm drift,H})_\phi$ only requires the radial collapse. We note that the Hall effect drives angular-momentum transport but cannot directly generate angular momentum, and the generation of rotation in the inner (or mideplane) region of the core is accompanied by the removal of angular momentum (or generation of counter-rotation) in the outer region of the core. Depending on the sign of $\boldsymbol{b}$ and $\eta_{\rm H}$ (which can be positive or negative, unlike $\eta_{\rm A}$ and $\eta_{\rm O}$ which are always positive), the Hall effect could either increase or decrease core rotation when the core is initially rotating (e.g., Wardle & Ng, 1999; Wardle, 2004; Braiding & Wardle, 2012; Tsukamoto et al., 2015; Zhao et al., 2020b). However, compared to Ohmic dissipation and ambipolar diffusion, the Hall effect is much less often included in numerical simulations, mainly because it is non-dissipative and more difficult to implement in a finite-difference or finite-volume code.

Computing the diffusivities

The strength of the non-ideal MHD effects directly depends on the chemical abundances of charged species. For charged species with density n_s, mass m_s, and charge q_s, we first compute the conductivity tensor $\boldsymbol{\sigma}$, which is defined through

$$\boldsymbol{j} = \boldsymbol{\sigma}\boldsymbol{E}_{\rm nonideal}. \tag{2.14}$$

With the magnetic field oriented in the z direction, the conductivity tensor takes the form

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_\perp & \sigma_{\rm H} & 0 \\ \sigma_{\rm H} & \sigma_\perp & 0 \\ 0 & 0 & \sigma_\parallel \end{pmatrix}. \tag{2.15}$$

The individual components are given by

$$\sigma_\parallel = \sum_s \sigma_s, \quad \sigma_\perp = \sum_s \frac{\sigma_s}{1+\omega_s^2\tau_{s\rm n}^2}, \quad \sigma_{\rm H} = -\sum_s \frac{\sigma_s\omega_s\tau_{s\rm n}}{1+\omega_s^2\tau_{s\rm n}^2}, \tag{2.16}$$

with

$$\sigma_s = \frac{n_s q_s^2 \tau_{s\rm n}}{m_s}, \quad \omega_s = \frac{q_s B}{m_s c}, \tag{2.17}$$

and $\tau_{s\rm n}$ is the mean (momentum exchange) collision times with neutrals, which depends on number density and temperature. Detailed expressions for $\tau_{s\rm n}$ of each species is given in Appendix A.1. Here we assume that $\tau_{s\rm n}$ is much shorter than the timescale of any inelastic collision; otherwise the effect of inelastic collisions also needs to be included in the conductivity calculation, as discussed in Kunz & Mouschovias (2009). We can then invert the conductivity tensor to get the resistivity tensor $\boldsymbol{\eta}^{\rm R}$ (here the subscript signifies that η represents resistivity instead of diffusivity)

$$\boldsymbol{\eta}^{\rm R} = \begin{pmatrix} \eta_\perp^{\rm R} & \eta_{\rm H}^{\rm R} & 0 \\ \eta_{\rm H}^{\rm R} & \eta_\perp^{\rm R} & 0 \\ 0 & 0 & \eta_\parallel^{\rm R} \end{pmatrix}, \tag{2.18}$$

with

$$\eta_\parallel^{\rm R} = \frac{1}{\sigma_\parallel}, \quad \eta_\perp^{\rm R} = \frac{\sigma_\perp}{\sigma_\perp^2+\sigma_{\rm H}^2}, \quad \eta_{\rm H}^{\rm R} = \frac{\sigma_{\rm H}}{\sigma_\perp^2+\sigma_{\rm H}^2}. \tag{2.19}$$

The resistivities are then related to the magnetic diffusivities by

$$\eta_{\rm O} = \frac{c^2}{4\pi}\eta_\parallel^{\rm R}, \quad \eta_{\rm A} = \frac{c^2}{4\pi}\left(\eta_\perp^{\rm R} - \eta_\parallel^{\rm R}\right), \quad \eta_{\rm H} = \frac{c^2}{4\pi}\eta_{\rm H}^{\rm R}. \tag{2.20}$$

2.2.2 The importance of chemistry and dust grains

The abundances of charged species can be computed with a chemical network. A minimum chemical network that includes the most relevant ingredients is provided in Kunz & Mouschovias (2009). This is an equilibrium chemical network where the active species are electrons, atomic and molecular ions (each modeled as only a

single species corresponding to Na^+ and HCO^+ respectively), and a family of neutral or singly (+/−) charged dust grains with different sizes. See details of the reactions in Kunz & Mouschovias (2009) or Appendix A.

The species abundances computed from this network [for a test $T(n)$ profile] and the resulting diffusivities [for a test $B(n)$ profile] is shown in Fig. 2.2. The figure shows results for several different dust size distributions, all with the same qualitative trends. In most cases the diffusivity is dominated by ambipolar diffusion. At low densities, the dominant charged species are electrons and atomic ions, and ionization and ambipolar diffusion generally decrease with density, with the balance between CR ionization ($\propto n_{\rm n}$) and recombination ($\propto n_{\rm i}^2$) giving an ionization fraction of $n_{\rm i}/n_{\rm n} \propto n_{\rm n}^{-1/2}$. At higher densities, dust grains begin to hold a larger fraction of charges and eventually become the dominant charged species, and ambipolar diffusion begins to increase with density. Meanwhile, the Ohmic diffusivity generally increases with density but remains less than the ambipolar diffusivity (for the test $B(n_{\rm n})$, $T(n_{\rm n})$, and $\zeta_{\rm CR}$ in Fig. 2.2). The evolution of the Hall effect is more complex, and $\eta_{\rm H}$ can change sign as the density increases. The relative importance of the Hall effect peaks at intermediate densities, where it may and may not be the dominate diffusivity depending on the dust properties.

Marchand et al. (2016) later modified the chemical network in Kunz & Mouschovias (2009) to include more atomic and molecular species; this does not significantly affect the resulting diffusivities. They also showed that the timescale of reaching chemical equilibrium is below free-fall time for $n \gtrsim 10^3$ cm^{-3}, thus chemical equilibrium should be a reasonable assumption in most of our calculation, with the possible exception being in the outflow region, which has low density and short dynamical timescale.

On the other hand, the properties of the dust grains, especially the grain size distribution, play an important role in determining the diffusivities; this is visible in Fig. 2.2, and has also been demonstrated in Zhao et al. (2018b). Grains in the general ISM follow a size distribution of $\mathrm{d}n/\mathrm{d}a \propto a^{-3.5}$ where a is the grain size with minimum and maximum grain sizes $a_{\rm min} \sim 0.005$–0.025 μm and $a_{\rm max} \sim 0.25$–1 μm, depending on the material. This is commonly referred to as the MRN distribution (Mathis et al., 1977). Star-formation calculations often use either the MRN distribution or a modified MRN distribution with the same power-law profile but different $a_{\rm min}$ and/or $a_{\rm max}$ to account for grain growth. An important property of this distribution is that both ends of the power-law grain size distribution matter. Most mass is concentrated in the largest grains, therefore $a_{\rm max}$ (at a given dust-to-gas ratio) sets the overall amplitude (intercept) of this power-law. Meanwhile, most of number density and grain surface area (which is proportional to the cross sections for reactions such as charge capture) are concentrated in the smallest grains, thus ionization and diffusivities are sensitive to $a_{\rm min}$. Therefore, both $a_{\rm min}$ and $a_{\rm max}$ need to be picked carefully and a single-sized dust distribution, which is sometimes adopted to reduce computational cost, may not be able to produce the correct evolution. A more self-consistent choice would be to directly evolve the dust size distribution,

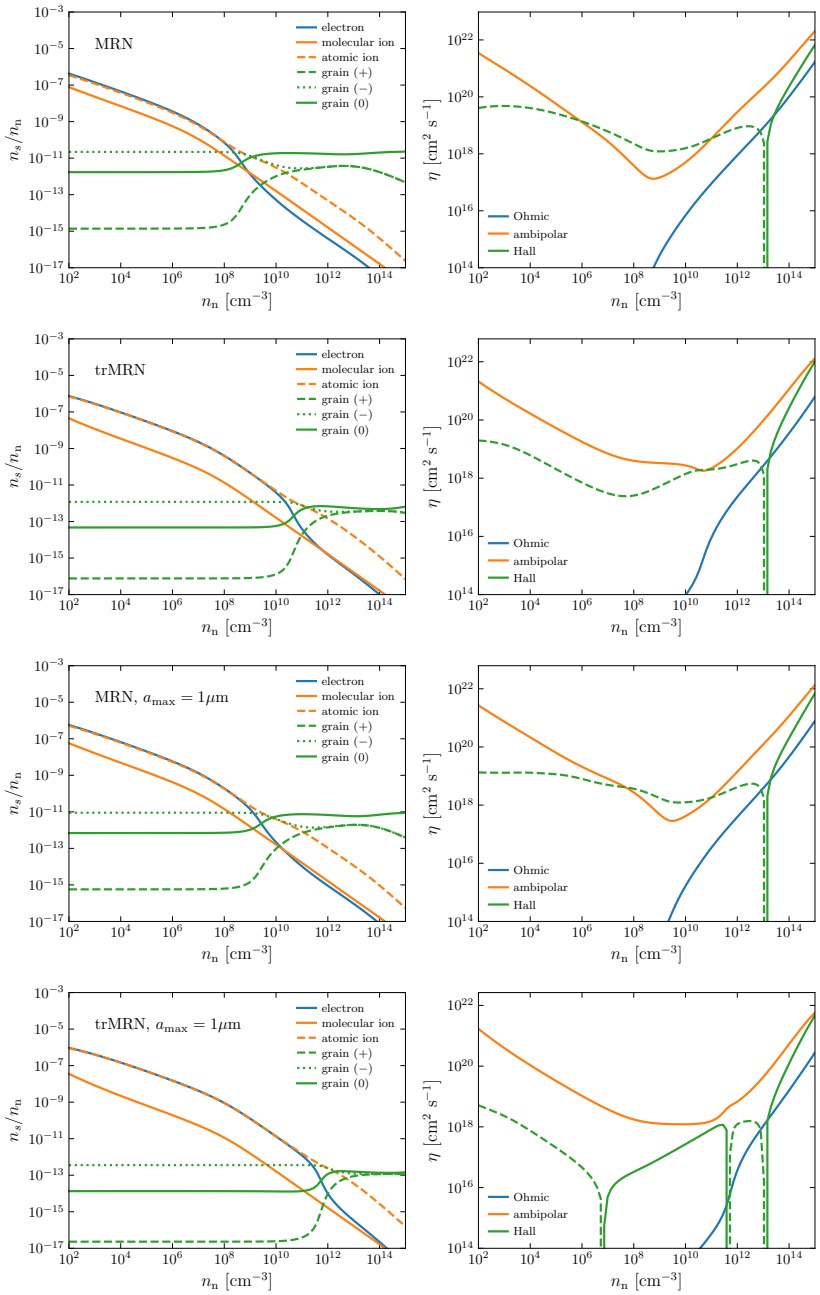

Figure 2.2: Species abundances and magnetic diffusivities computed using the chemical network in Kunz & Mouschovias (2009), using a few different grain size distributions. First row: MRN distribution ($\mathrm{d}n/\mathrm{d}a \propto a^{-3.5}$) with $a_{\rm min} = 0.025\mu$m and $a_{\rm max} = 0.25\mu$m. Second row: truncated MRN (trMRN) distribution with with $a_{\rm min} = 0.025\mu$m and $a_{\rm max} = 0.25\mu$m. Third row: $a_{\rm min} = 0.025\mu$m, $a_{\rm max} = 1\mu$m. Last row: $a_{\rm min} = 0.1\mu$m, $a_{\rm max} = 1\mu$m. For simplicity, all calculations assume $T = 10[1 + (n_{\rm n}/2 \times 10^{11}{\rm cm}^{-3})^{3/2}]$ K, $B = 0.1(n_{\rm n}/{\rm cm}^{-3})^{1/2}$ μG, and CR ionization rate $\zeta_{\rm CR} = 10^{-17}{\rm s}^{-1}$.

although existing attempts (e.g., Lebreuilly et al., 2020) have been unable to include a sufficiently realistic dust evolution model covering both coagulation and fragmentation (cf. Kawasaki et al., 2022).

Finally, we note that studies of the effect of non-ideal MHD diffusivities sometimes assume simple parametrized profiles of magnetic diffusivities as a function of density and temperature (e.g., Machida et al., 2007; Tomida et al., 2017), and this could be problematic. In some cases, the parametrized model simply does not have the right shape for the relation between diffusivity and density (which is non-monotonic, cf. Fig. 2.2). Additionally, ambipolar diffusion, which dominates diffusivities in most cases, is also sensitive to the magnetic-field strength, with $\eta_{\rm A} \propto B^2$ when all $\omega_s^2 \tau_{sn}^2 \gg 1$. These caveats need to be taken into consideration when interpreting the results of studies using a simple parametrization of η.

2.2.3 Radiation and thermal evolution

Realistic modeling of thermal evolution requires one to calculate $Q_{\rm rad}$ with radiative transfer calculations. Performing full radiative transfer [i.e. solve for intensity $I_\nu(\boldsymbol{x}, \boldsymbol{\mu})$ directly] is very expensive, in part because the frequency (ν) and direction ($\boldsymbol{\mu}$) dependence of intensity increases the dimensionality of the problem.

In the context of star formation, the effect of radiation is commonly modeled with grey flux-limited diffusion (FLD; Levermore & Pomraning 1981; for examples of star formation applications see Bodenheimer et al. 1990; Whitehouse & Bate 2006; Kunz & Mouschovias 2009, 2010; Tomida et al. 2015) Here grey means that the frequency dependence in the opacity is ignored (by replacing opacity κ_ν with, for example, the Planck mean opacity $\kappa_{\rm P}$), which allows I_ν to be replaced by a frequency-integrated intensity I. FLD assumes that the radiation flux $\boldsymbol{F}$ is proportional to the gradient of the isotropic intensity J (μ-averaged I), thereby removing the direction dependence. Radiative transfer then becomes a diffusion equation of J with diffusion coefficient $D = c\lambda/(\rho\kappa)$. Here λ is a dimensionless flux limiter that reduces D to preserve causality when the radiation flux would be too large without the limiter. FLD is essentially an interpolation between optically thin and optically thick limits; $Q_{\rm rad}$ computed with FLD is exact in these limits, and is only a rough estimate in the intermediate regime.

Grey FLD is already a big simplification compared to full radiative transfer. However, finding an efficient implementation of this method in numerical simulations remains challenging. The diffusion coefficient can be very large and that requires an implicit solver if one wants to keep the timestep manageable. Yet designing an implicit solver for parallel computation is challenging, especially since our problem requires a highly non-uniform (or refined) grid.

Given the difficulty of modeling radiative transfer, the majority of low-mass star formation simulations adopt a barotropic equation of state (EoS) to model the thermal evolution of the system. The simplest choice

is an isothermal EoS with temperature ~ 10 K. This is a reasonable approximation during early stages of the collapse, but this approximation eventually breaks down at high densities as the heat is unable to escape freely.

Consider the near self-similar spherical collapse of a core initially with size r_0 and uniform density n_0; when the central density reaches n_c the core will have approximately uniform density at $r < r_0(n_0/n_c)^{1/2}$ and a $\propto r^{-2}$ density profile further out. The optical depth at the central part of the core is then given by

$$\tau_c \sim \kappa\mu m_p n_c r_0 (n_0/n_c)^{1/2} \sim 1.2 \left(\frac{n_c}{10^{12}\ \mathrm{cm}^{-3}}\right)^{1/2} \left(\frac{n_0}{10^4\ \mathrm{cm}^{-3}}\right)^{1/2} \left(\frac{r_0}{0.1\ \mathrm{pc}}\right) \left(\frac{\kappa}{0.01\ \mathrm{cm}^2\mathrm{g}^{-1}}\right). \quad (2.21)$$

Here $\mu = 2.33$ is the mean molecular weight and κ is the opacity of the material. Meanwhile, the timescale for optically thin material to reach thermal equilibrium at temperature $T_{\rm eq}$ is given by

$$t_{\rm eq} \sim \frac{u}{4\rho\kappa\sigma T_{\rm eq}^4} \sim 7 \times 10^2 \left(\frac{T_{\rm eq}}{10\ \mathrm{K}}\right)^{-3} \mathrm{yr}. \quad (2.22)$$

Here $u = \rho(k_B T_{\rm eq}/\mu m_p)(\gamma - 1)^{-1}$ with $\gamma = 5/3$ being the internal energy density; σ is the Stefan–Boltzmann constant. Finally, the free-fall timescale at the center of the core is given by

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32 G\mu m_p n_c}} \sim 10^3 \left(\frac{n_c}{10^{12}\ \mathrm{cm}^{-3}}\right)^{-1/2} \mathrm{yr}. \quad (2.23)$$

Given the estimates above, we see that the pre-stellar core should remain optically thin and near thermal equilibrium with the background interstellar radiation field (at $\sim$ 10K) up to a density of $\sim 10^{12}$ cm^{-3}. Beyond this density, the central part of the core becomes optically thick and the temperature can increase; this leads to the formation of the FHSC. Inside the FHSC, the high optical depth makes cooling inefficient and that results in an approximately adiabatic equation of state (cf. Larson, 1969; Kunz & Mouschovias, 2010).

Many simulations have modeled this deviation from isothermal evolution using barotropic EoSs that switch from constant temperature to power laws with slopes chosen to match the $T(n)$ relation (at the center of the core) in star-formation calculations that resolve the protostar and include radiation. Such a barotropic EoS can be quite complicated and can capture features such as the change in the slope of $T(n)$ due to the freeing up of rotational and vibrational degrees of freedom and due to H_2 dissociation (e.g., Machida et al., 2006).

However, one should not expect any barotropic EoS to provide reasonable approximation to the disk temperature as the evolution of protostellar disk can be highly non-barotropic in nature. The disk tem-

perature is determined by a balance between heating and cooling that depends on many factors other than density, such as radial location and the dominant heating/cooling mechanism. There is no reason to expect the temperature in the disk to be similar to the temperature in the protostar (FHSC) when they have similar density, as the physical mechanisms determining the thermal budget are completely different. This has also been demonstrated in previous disk formation simulations that directly model radiation (e.g., Vaytet et al., 2018).

In summary, existing methods for modeling thermal evolution either are numerically expensive and difficult to implement (radiative transfer calculations, including FLD) or cannot produce realistic temperature estimates in the disk (barotropic EoS). This calls for new methods of estimating $Q_{\rm rad}$ with a better balance between realism and numerical cost.

2.3 Numerical considerations

2.3.1 Numerical convergence in existing studies

A major issue in existing simulations of disk formation is the lack of numerical convergence. Previous simulations covering a relatively long timescale ($\gtrsim$ 10 kyr after protostar formation), when a resolution study is available, usually (if not always) exhibit nontrivial differences in the outcome of disk formation across different resolutions. Here we provide a few examples of recent resolution studies.

In the radiative hydrodynamic simulation by Bate (2018) using smoothed-particle hydrodynamics (SPH), a resolution study suggests that the density profile at radii $\lesssim$ 100 au does not converge (i.e., remains resolution-dependent) even at the maximum tested resolution of 2×10^5 particles per solar mass, while their fiducial model has a resolution 3 times lower. The convergence for disk mass and size are slightly better at earlier times; their relative differences between resolutions of 2×10^4 and 2×10^5 particles per solar mass are $\lesssim 20\%$, but the differences increase at later times (Fig. B1, B2 in Bate 2018). A more recent simulation by the same group that further includes non-ideal MHD shows similar non-convergence, although this time increasing resolution leads to the formation of fewer stars and smaller, less massive disks (Wurster et al., 2019). Thermal properties of the disk also appear to be dependent upon resolution; in the radiative non-ideal MHD simulation by Vaytet et al. (2018) using a Cartesian grid with adaptive mesh refinement (AMR), reducing the resolution by a factor of 2 leads to the temperature being systematically higher also by a factor of ~ 2. Finally, Hennebelle et al. (2020) use a set of non-ideal MHD simulations on a Cartesian grid with AMR to survey different resolution (max refinement level), sink-particle size, and other sink-particle parameters (including the threshold density at which sink forms and the rate of removing mass in the sink

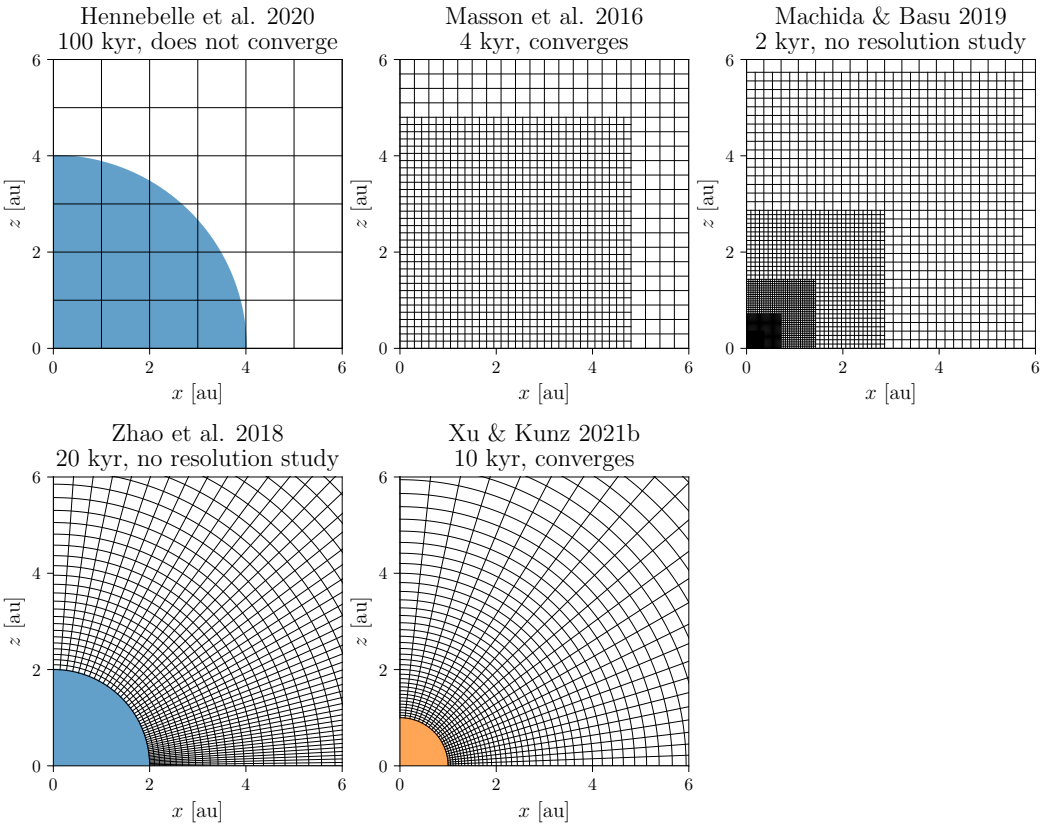

Figure 2.3: Grid and sink particle / inner boundary for representative recent grid-based simulations, including the one presented in this thesis (Xu & Kunz, 2021b). In the title of each panel we show how long the simulation is continued after protostar formation, and whether a resolution test (when available) demonstrates convergence. For studies that include multiple resolutions, the fiducial resolution is plotted. The last panel, which shows the setup of this thesis, will be discussed in Section 2.4.4.

region); both disk size and mass show significant variation when *any* of these parameters is varied. In the most extreme case, a disk simply disappears within ~ 10 kyr when restarting a fiducial resolution run while increasing the resolution at small radii by a factor of 4.

On the other hand, simulations with better resolution do demonstrate good numerical convergence, but they are often limited by small timesteps and generally run for at most a few kyr after protostar formation. For example, the non-ideal MHD runs in Masson et al. (2016) have a minimum cell size of ~ 0.1 au and show very good convergence, but only cover the first ~ 5 kyr after protostar formation. A more extreme example is the simulation by Machida & Basu (2019), which has excellent resolution with a minimum cell size $\sim \mathrm{R}_\odot$ but is only able to cover ~ 2 kyr after protostar formation using $\gtrsim 2$ yr of wall-clock time. In Fig. 2.3 we compare the numerical grids in a few recent grid-based simulations to demonstrate the trade-off between numerical convergence (resolution) and integration time.

2.3.2 How to achieve numerical convergence in a long-term simulation?

In order to achieve numerical convergence in a long-term simulation, we need to first understand why many of the existing simulations fail to achieve numerical convergence. We begin by discussing the contribution of resolution. There are several ways in which resolution can affect the outcome of disk formation. First, the resolution needs to be high enough to resolve at least the characteristic length of the dominant angular-momentum transport mechanism. For example, in the case of magnetic braking in the disk, one should at least resolve the disk scale height in order to capture the curvature of the fieldline correctly. The resolution constraint is more stringent if the main transport mechanism has a smaller scale; for example, the magnetorotational instability (MRI) generally requires at least 8–16 cells per scale height to be resolved (Simon et al., 2009).

Another important constraint on resolution comes from the numerical dissipation and diffusion of angular momentum at the grid scale (or particle scale, for SPH). Empirically, on a Cartesian grid the timescale at which grid-scale angular-momentum diffusion can significantly affect disk properties at a radius R is given by (Appendix B)

$$t_{\rm L,diff} \sim \Omega^{-1} \left(\frac{R}{{\rm d}x}\right)^3 \sim 2\times10^4 \left(\frac{M_\star}{\rm M_\odot}\right)^{-1/2} \left(\frac{R}{10\ {\rm au}}\right)^{3/2} \left(\frac{R/{\rm d}x}{16}\right)^3\ {\rm yr}. \tag{2.24}$$

Here dx is the cell size. If the timescale of disk evolution (which, roughly speaking, is comparable to the disk age multiplied by disk-to-star mass ratio) is below $t_{\rm L,diff}$, transport inside the disk will be dominated by non-physical angular momentum diffusion. Cartesian grid simulations generally have $R/{\rm d}x \sim$ 8–32 on large scales, and much lower values once the highest level of refinement is reached. Our estimates suggest that reasonable numerical convergence at $\sim$ 10 au is achievable for this kind of resolution but only when mesh refinement continues below this scale, consistent with the trend in previous studies (cf. Fig. 2.3). Meanwhile, numerical convergence at $\sim$ au scale would be very hard to achieve with a Cartesian grid. We note in passing that existing AMR disk formation simulations have not taken the above constraints into consideration; instead, they generally refine the grid according to the local Jeans length. While this criterion is reasonable for the near-radial collapse of pre-stellar core (and does not require very high resolution; cf. Commerçon et al., 2008), it does not capture the resolution requirement for correctly following disk evolution. This is because the latter requires the grid to conserve angular momentum over many orbits.

Given the limitations of a Cartesian grid, a better alternative might be to use a curvilinear grid with azimuthal symmetry; this provides a negligible level of angular momentum dissipation and diffusion at very modest resolution (Appendix B). Spherical coordinates have been used in a number of disk formation simulations (e.g., Zhao et al., 2018a); unfortunately these studies do not include resolution studies.

In addition to resolution, the size and numerical implementation of the inner boundary, or sink cell, are also important. For example, the resolution study by Machida et al. (2014) suggests that sink size $r_{\rm sink} \lesssim 1$ au and threshold density (the density above which the sink begins to accrete) $n_{\rm threshold} \gtrsim 10^{13}$ cm^{-3} might be necessary for producing the correct evolution. The reason for this is probably two-fold. First, if the sink or inner boundary size is too large, the FHSC may be unable to form in the computational domain and that could hinder disk formation because the FHSC, which is dense with high $\eta_{\rm AD}$ and largely decoupled from magnetic field, serves as a repository for angular momentum that can later be used to form the disk. Another factor to consider is that angular momentum can be lost through (or near) the inner boundary. Consider, for example, the inner boundary of a spherical domain. For a mass accretion rate $\dot{M}$, the angular-momentum flux through the inner boundary can be up to $\dot{M}\sqrt{GM_\star r_{\rm in}}$ where $r_{\rm in}$ is the inner boundary size. In reality, the star (and the disk at $r < r_{\rm in}$) can only hold a small amount of angular momentum, therefore most of the angular momentum advected through $r_{\rm in}$ needs to return to $r > r_{\rm in}$ via radial angular-momentum transport (e.g., via an effective viscosity driven by some instability). However, this "repay" of angular momentum cannot be captured by a simple outflow boundary condition, which is often used in simulations because of its simplicity. Some sink-particle implementations on a Cartesian grid attempts to avoid this by letting the sink accrete only mass but not momentum from active cells. This method could work when resolution is very high; but in reality the sink particle is often resolved with only a few cells, and angular momentum could be quickly diffused and dissipated near the sink.

In summary, there are at least three requirements for achieving numerical convergence. First, resolution needs to be good enough to resolve relevant transport mechanisms. Second, angular-momentum dissipation and diffusion at the grid scale needs to be sufficiently weak; this requires either Cartesian grids with resolutions that are higher than currently affordable for long-term integration, or the adoption of an azimuthally symmetric grid (e.g., spherical-polar). Finally, the inner boundary condition or sink particle has to be either sufficiently small or specially engineered to prevent excessive angular-momentum loss through the inner boundary or near the sink particle.

2.4 Our simulation setup

2.4.1 Simulation overview

We perform a radiative non-ideal MHD simulation that follows the evolution of a self-gravitating, magnetic, poorly ionized pre-stellar core until ∼10 kyr after protostar (i.e., point-mass) formation, covering the formation and early evolution of a massive, rotationally supported, protostellar accretion disc. The initial

conditions of the core will be described in Section 2.4.2.

The simulation is performed using the code `Athena++` (Stone et al., 2020) with a 3D spherical-polar (r, θ, ϕ) grid with an accreting point mass at the origin representing the protostar. `Athena++` is a general-purpose astrophysical fluid dynamics code, and its super-time-stepping feature (Meyer et al., 2014) allows significant acceleration for computing diffusive processes (e.g., Ohmic dissipation and ambipolar diffusion). In order to use `Athena++` for our simulation, we developed a few new physical modules and algorithmic adjustments; these modifications will be described in Section 2.4.3 and 2.4.4.

In light of previous discussions in this chapter, we focus on making our simulation capture the basic (and most necessary) physical ingredients of protostellar disk formation while maintaining good numerical convergence. On the other hand, physical ingredients that might affect disk formation but are likely not necessary for disk formation, such as core asymmetry, turbulence, misalignment between field and rotation, and the Hall effect, are left for future studies. Given our relatively idealized setup, we do not expect our simulation to reproduce accurately the evolution of a typical (or "median") protostellar system in nature. Instead, we aim to use this simulation to provide qualitative insights into what mechanisms regulate the evolution of protostellar disks, and to test quantitatively analytic theories that describe these mechanisms and predict disk evolution.

For this thesis project we have published two papers on simulating protostellar disk formation, Xu & Kunz (2021a, hereafter XK21a) and Xu & Kunz (2021b, hereafter XK21b). The main difference between the two is that XK21a adopts a barotropic equation of state while XK21b includes realistic radiative cooling. In this thesis we mainly discuss the simulation setup and results from XK21b; XK21a will only be used for discussing numerical convergence in Appendix D. Since XK21b, there have been some improvements and modifications to the code and simulation setup; these modifications will be used for future studies, and we document them at the end of each subsection.

2.4.2 Initial conditions

We start our simulation with an initially spherical pre-stellar core with uniform temperature $T_0 = 10$ K and radially (r) dependent number density of neutrals given by

$$n_{\rm n}(r) = \max\left\{\frac{n_0}{1+(r/r_0)^2}, n_\infty\right\}, \tag{2.25}$$

with initial central density $n_0 = 10^4$ cm^{-3} and characteristic scale $r_0 = 0.1$ pc. The background density is $n_\infty = 500$ cm^{-3}; it represents the ambient density in the parent molecular cloud and is excluded from the calculation of the self-gravitational potential. We also exclude n_∞ from all integrated diagnostics; for

example, a 'density-weighted' integration in the z direction is weighted only by $\max\{0, n_{\rm n} - n_\infty\}$. The mean mass per neutral particle $m_{\rm n} = 2.33 m_{\rm p}$ (accounting for molecular hydrogen with 20% He by number), so that the neutral mass density $\rho_{\rm n} = m_{\rm n} n_{\rm n}$ and the initial isothermal sound speed $c_{\rm s0} \equiv (k_{\rm B} T_0 / m_{\rm n})^{1/2} \simeq 0.188$ km s^{-1}. The total self-gravitating mass within r_0 is then 1.43 $M_\odot$. We set the core into uniform rotation with an initial angular frequency $\Omega_0 = 0.2 c_{\rm s0}/r_0 \simeq 1.22 \times 10^{-14}$ rad s^{-1} for $R < R_0$ (where $R = r \sin\theta$ is the cylindrical radius) and $\Omega = \Omega_0 (R/r_0)^{-2}$ for $R > r_0$. The core is threaded by a uniform magnetic field with strength $B_0 = 25$ μG that is aligned with the rotation axis, giving an initial mass-to-flux ratio in the central flux tubes of the core of $\simeq$1.5 times the critical value for collapse, $(3/2)(63{\rm G})^{-1/2}$ (Mouschovias & Spitzer, 1976). This initial condition is representative of early observations of NH_3 cores (Barranco & Goodman, 1998; Jijina et al., 1999; Crutcher, 1999). The choice of uniform rotation and aligned magnetic field is partly because the symmetry of the grid prefers an axisymmetric initial condition.

Planned modifications for future studies: For future studies, we plan to use an enhanced Bonnor–Ebert sphere as the initial density distribution. This is a density profile where the ratio between gravity and pressure gradient is a constant $f_{\rm BE}$ that is > 1. The profile can be specified by $f_{\rm BE}$ and central density n_0, and the mass of the core can be characterized by the density within the radius where the density profile transitions from approximately constant to $n \propto r^{-2}$ as we do in Eq. (2.25); for example, one could use the radius where density drops to $n_0/2$. (Meanwhile, most existing studies using a Bonnor–Ebert profile use the mass out to a much lower density as the core mass, which is less consistent with observational definitions as discussed in Section 2.1.) The main advantage of this profile (compared to our current choice) is that a smooth collapse is still possible when the core is very close to being thermally critical ($f_{\rm BE} \to 1$). The profile in Eq. 2.25 would produce artifacts near r_0 in that case. This features is especially useful for parameter surveys. We could also slightly increase the central density and decrease core mass and size to be more consistent with observations (e.g., Fig. 2.1). Finally, the adoption of an ambient density n_∞ might not be necessary, because the radius at which density drops to n_∞ is so far away that the properties there can barely affect evolution near the protostar. Meanwhile, dropping n_∞ has the advantage of reducing the number of free parameters in our setup.

2.4.3 Physical modules

Self-gravity

The gravitational potential in the domain can be separated into two components, one corresponding to the mass in the domain, and the other the point accretor (protostar) $M_{\rm acc}$:

$$\Phi = \Phi_{\rm s} + \Phi_{\rm acc}. \tag{2.26}$$

Here $\Phi_{\rm acc} = -GM_{\rm acc}/r$ and $\Phi_{\rm s}$ is obtained by solving the Poisson equation

$$\nabla^2 \Phi_{\rm s} = 4\pi G(\rho - \rho_\infty). \tag{2.27}$$

We adopt an open boundary condition for equation (2.27): $\Phi_{\rm s}$ is finite at $r \to 0$ and approaches zero at $r \to \infty$ if we were to continue solving $\nabla^2 \Phi = 0$ outside the domain.

On the right-hand side of equation (2.27), we have replaced the usual density term ρ with $(\rho - \rho_\infty)$, where ρ_∞ is the background mass density corresponding to the background number density n_∞ in equation (2.25). We also replace $\rho \boldsymbol{g}$ with $(\rho - \rho_\infty)\boldsymbol{g}$ in the momentum equation. This modification, together with an open boundary condition for $\Phi_{\rm s}$, corresponds physically to a situation where the system is embedded in an infinite uniform ambient density ρ_∞.

We solve equation (2.27) numerically in spherical-polar coordinates using spherical harmonic decomposition. The transformations to and from the spherical harmonic basis are performed using a fast Fourier transform (FFT) in the ϕ direction (necessary only for 3D simulations) and an inner product with the associated Legendre polynomials through direct summation. In this basis of spherical harmonics, the Poisson equation becomes a second-order ordinary differential equation for each Y^{lm} component. The (second-order accurate) finite-difference version of this differential equation corresponds to a tri-diagonal matrix equation, which we solve using LU factorization with pre-computed coefficients. This allows a solution cost of $\mathcal{O}(N_r)$ for each Y^{lm} component. When running the code in parallel, in each of these steps each thread needs to know data on columns of cells covering the whole domain of ϕ, θ, and r respectively; this is achieved by doing a different domain decomposition between each step. The infrastructure for such remaps have already been implemented in Steve Plimpton's parallel FFT package, which is currently used by `Athea++`'s FFT module. Overall, the computational cost for a single solve is $\mathcal{O}(N) + \mathcal{O}(N \log N_\phi) + \mathcal{O}(N N_\theta)$, where N_r, N_θ, N_ϕ are number of cells in each direction and $N \equiv N_r N_\theta N_\phi$ is the total number of cells. On paper, this scaling does not look very good. In practice, however, the simplicity of the algorithm lends it good performance, and the cost of solving $\Phi_{\rm s}$ becomes significant (compared to everything else) only when $N_\theta \gtrsim 100$. At our fiducial

resolution, the cost of the self-gravity solver constitutes only ∼15% of the total computational cost.[2]

A test of this self-gravity solver is presented in Appendix C.1.

Non-ideal MHD and chemistry

We include Ohmic dissipation and ambipolar diffusion in our simulations; the Hall effect is neglected for technical difficulties. The associated diffusivities are calculated from the abundances of charged species following Section 2.2.1.

The abundances of charged species are computed using an equilibrium chemical network that includes electrons, atomic and molecular ions, and a distribution of (neutral, singly negatively charged, and singly positively charged) dust grains divided into 5 bins of different sizes. The chemical model is identical to that in Kunz & Mouschovias (2009, §4; also documented in Appendix A.2), with the following modifications. First, we choose a smaller CR ionization rate of $\zeta_{\rm CR} = 10^{-17}$ s^{-1}, following other contemporary numerical models of pre-stellar core contraction (e.g., Li et al. 2011; Marchand et al. 2016; Tomida et al. 2015; Zhao et al. 2018a).[3] Observational estimates of the CR ionization rate are higher in diffuse clouds ($\sim 10^{-16}$ s^{-1}, cf. Indriolo et al. 2015), and decrease towards higher column densities (Neufeld & Wolfire, 2017). While $\zeta_{\rm CR} \sim 10^{-17}$ s^{-1} is a reasonable choice for column densities representative of pre-stellar cores, $\zeta_{\rm CR}$ might vary significantly between different cores and even within the same core (cf. Fig. 6 of Neufeld & Wolfire 2017). We also adopt a slightly different dust profile. Namely, we assume spherical dust grains distributed in radius a according to a truncated MRN (Mathis et al., 1977) size distribution, in which the number density of grains with radii between a and $a + {\rm d}a$ satisfies ${\rm d}n_{\rm g}/{\rm d}a \propto a^{-3/5}$ from a minimum grain size $a_{\rm min} = 0.1$ μm to a maximum grain size $a_{\rm max} = 0.25$ μm. The total dust mass is taken to be 1% of the gas mass, and we assume a grain density of 2.3 g cm^{-3}. The exclusion of dust grains smaller than 0.1 μm is motivated by studies showing that dust drift and coagulation can efficiently remove small dust grains during the collapse of the pre-stellar core (Rossi et al., 1991; Ossenkopf, 1993; Ormel et al., 2009; Hirashita, 2012; Guillet et al., 2020). The exclusion of small grains also increases ambipolar diffusion and promotes disk formation (Zhao et al., 2018a). The second row of Fig. 2.2 shows the abundance of different charged species and the resulting non-ideal diffusivities for our dust model (trMRN) for given $T(n_{\rm n}), B(n_{\rm n})$ profiles.

In order to save computational cost, we construct an interpolation table for equilibrium species abun-

[2] As a side note, algorithms with better scaling, such as multigrid [which is $\mathcal{O}(N)$], tend to require more communication between cells and become much less efficient when there are too few cells per thread (e.g., when each thread owns 8^3 or 16^3 cells); this problem is less severe for our spherical harmonic solver. This makes our solver suitable for reducing the real-world time of a simulation by using more cores.

[3] At high column densities ($\gtrsim 10^2$ g cm^{-2}, see Umebayashi & Nakano 1980), CRs will be shielded by optically thick gas, leading to lower $\zeta_{\rm CR}$ and higher magnetic diffusivities. We ignore this effect in our model and use a constant $\zeta_{\rm CR}$. This should not affect the evolution though, since in our simulation the magnetic field is already barely coupled to the gas (due to ambipolar diffusion) at such high densities.

dances on a 2D grid of density and temperature. Reading from the interpolation table is significantly faster than computing species abundances on-the-fly, which requires matrix inversion operations for every cell at every timestep. The interpolation is done for $\log n_s$, and therefore does not preserve charge neutrality. After interpolation for each n_s, we compute the total density of positive and negative charge (normalized by the elementary charge), $n_\pm$, and scale positively/negatively charged n_s by $(n_+ + n_-)/2n_{+/-}$. This correction is important for computing $\sigma_{\rm H}$ (which affects $\eta_{\rm AD}$ and $\eta_{\rm H}$) because when $\omega_s^2\tau_{sn}^2 \gg 1$ the terms in the summation for $\sigma_{\rm H}$ (Eq. 2.16) will be proportional to $n_s q_s$ and the summation goes to zero to lowest order due to charge neutrality.

Planned modifications for future studies: There are several changes planned for future studies. First, we plan to include the sublimation of dust grains in the chemical network. This can be achieved by using different grain density and dust-to-gas ratio for different temperature ranges, following our new dust model (see next subsection). In particular, at very high temperature (> 1200 K), all grains will be removed. Dust sublimation leads to sudden changes in abundances, and we discuss how this can be handled by our interpolation table in Appendix C.2. Second, we plan to include the effect of thermal ionization, which could be important for the innermost $\sim$au of the disk; the details of the reactions to be included are discussed in Appendix A.3. Finally, we plan to include the effect of CR shielding (Appendix A.4), as the algorithm for estimating optical depth in our radiation module is also capable of estimating the optical depth for CR shielding (cf. Appendix C.3). This also requires changing our abundance tables from 2D $(n_{\rm n}, T)$ to 3D $(n_{\rm n}, T, \zeta_{\rm CR})$. Initial testing shows that the impact on performance by this change is negligible.

Equation of state and radiative cooling

The equation of state in our simulation is an ideal equation of state $u = p/(\gamma - 1)$, where the ratio of specific heats γ is temperature-dependent to capture the temperature dependence of the vibrational and rotational degrees of freedom of H_2; we use the expreession of γ in Kunz & Mouschovias (2009, Eqs. 10 – 13 and Fig. 1).

We model the rate of radiative heating/cooling $Q_{\rm rad}$ as follows. In the optically thin limit (optical depth $\tau \to 0$), the gas should cool down towards the ambient temperature at $r \to \infty$, $T_0 = 10$ K. The corresponding radiative cooling rate is

$$Q_{\tau\to 0} = -4\rho\kappa_{\rm P}\sigma(T^4 - T_0^4), \tag{2.28}$$

where $\kappa_{\rm P}$ is the Planck mean opacity and σ is the Stefan–Boltzmann constant. In the opposite limit ($\tau \to \infty$), in which the gas is optically thick in all directions and the length scale of the temperature variation is $\gg (\rho\kappa_{\rm R})^{-1}$ (where $\kappa_{\rm R}$ is the Rosseland mean opacity), the radiative energy transport is diffusive and may

be modeled by

$$Q_{\tau\to\infty} = -\boldsymbol{\nabla}\cdot\boldsymbol{F}_{\tau\to\infty} \tag{2.29}$$

with the radiative energy flux

$$\boldsymbol{F}_{\tau\to\infty} = -\frac{4}{3}(\rho\kappa_{\rm R})^{-1}\sigma\boldsymbol{\nabla}T^4. \tag{2.30}$$

Because most of the system (per volume and per mass) is in one of these two limits, we may obtain a good estimate of $Q_{\rm rad}$ by interpolating between them using the optical depth:

$$Q_{\rm rad} \approx \mathrm{e}^{-\tau}Q_{\tau\to 0} - \boldsymbol{\nabla}\cdot(\mathrm{e}^{-1/\tau}\boldsymbol{F}_{\tau\to\infty}). \tag{2.31}$$

This provides a simple estimate of the rate of radiative cooling and diffusion without performing computationally expensive radiative transfer calculations. We expect this estimate to perform reasonably well, mainly because most of the system has either $\tau \ll 1$ (envelope and pseudodisk) or $\tau \gg 1$ (disk), and the region with $\tau \sim 1$ (only a thin layer on the disk surface) constitutes a very small volume (cf. Fig. C.2).

A key challenge in this method is to estimate optical depth (more precisely, $\mathrm{e}^{-\tau}$ and $\mathrm{e}^{-1/\tau}$) correctly. We develope a "pseudo-ray-tracing" method for optical depth estimation, where the optical depth of a cell is defined by the minimum optical depth along L-shaped rays that are aligned to the grid. These are rays that first go towards the pole along θ for a number of cells (can be 0), then go out along r to infinity (outer boundary). In Appendix C.3 we detail how this calculation can be efficiently parallelized and test our radiation model in Appendix C.4. In summary, we compare our estimate of τ in our simulation with the corresponding τ obtained via ray tracing, finding that the errors in our estimates of $\mathrm{e}^{-\tau}$ and $\mathrm{e}^{-1/\tau}$ are $\ll$1 except in a few cells. We also compare our model's prediction with full radiative transfer for a test problem, where we estimate the temperature profile in thermal equilibrium for a small patch of a geometrically thin disk (so the problem is effectively 1D in the vertical direction) subject to constant heating. We varied the heating rate and disk optical depth to survey a large parameter space, and found that our model accurately predicts the mean (density-weighted) temperature of the disk, with an error converging to zero when $\tau \ll 1$ or $\gg$1 and a maximum error $\sim$25% at intermediate optical depths. Additionally, where our simulation overlaps with previous simulations of protostellar core collapse using flux-limited diffusion to solve for the radiative transfer (e.g., Kunz & Mouschovias, 2010; Tomida et al., 2015), we find very similar temperatures.

Our radiation model does not include the effect of protostellar irradiation. This should not significantly affect the protostellar disk evolution for the evolutionary phases we simulate, because the innermost part of the disk generally has a larger aspect ratio H/R than the rest of the disk and of the pseudodisk farther out (hence shielding them from direct protostellar irradiation). That being said, it is possible for protostellar

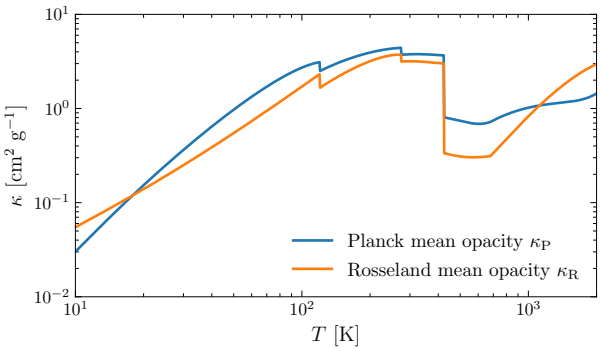

Figure 2.4: Planck and Rosseland mean opacities used in our simulation.

irradiation to affect the evolution of the outflow cone and the protostellar envelope high above the midplane (mainly through photoionization), and this may affect outflow propagation (but not launching, since disk surface is well shielded) and envelope dispersal (which should become important only towards the end of the Class I phase). It is also possible for the protostar irradiation to indirectly heat the disk by scattering off the envelope. Addressing these topics is beyond the scope of this thesis, but will be enabled by future improvements of and additions to our radiation and chemistry models.

For the range of temperature typical for the core and the disk, opacity is dominated by dust grains, and depends only on the temperature. The opacity table we use is similar to the iron-poor model of Semenov et al. (2003) except that species abundances are chosen to better reflect conditions in molecular clouds (D. Semenov, private communication). We show the opacity table we use in Fig. 2.4.

Planned modification for future studies: We plan to switch to the DSHARP dust model (Birnstiel et al., 2018) for computing grain opacities in future simulations. We have already used this model for fitting observation; see details in Chapter 5. It is more flexible than our current dust opacity tables as one can easily include the effect of grain sublimation and compute opacity tables for different grain size distributions. Moreover, the model is publicly available and that makes our simulations more reproducible.

2.4.4 Numerical setup

Computational domain, resolution, and numerical convergence

We use a spherical-polar grid (r, θ, ϕ) with an inner radial boundary at 1 au and an outer radial boundary at 10^5 au. Material flowing through the inner boundary is irreversibly accreted by a point-mass 'protostar'. The radial (r) grid is log-uniform with 240 cells, with $\mathrm{d}r/r \approx 0.05$. To reduce computational cost, we simulate only $\theta \in [0, \pi/2]$, with a reflecting boundary condition at $\theta = \pi/2$, and $\phi \in [0, \pi]$, with periodic boundaries at $\phi = 0$ and π. The polar (θ) grid is non-uniform with 24 cells, with the grid spacing decreasing towards

the midplane to be 1/3 of that at the pole. This gives a midplane angular resolution of $\simeq$0.036 rad. The azimuthal (ϕ) grid is uniform with 16 cells for $\phi \in [0, \pi]$.

In Appendix D we compare simulations covering a range of different $r_{\rm in}$ and resolutions, and demonstrate that our fiducial resolution is sufficient for achieving numerical convergence for basic disk properties.

Inner boundary condition

For the outer radial boundary, we use an open boundary condition (density, velocity, and magnetic field in the ghost cells are copied from the adjacent active cell). We note that no visible artifacts appear at that boundary, which is additionally placed far enough away ($\gg r_0$) so as not to influence the evolution of the contracting core during our simulation.

For the inner radial boundary, we also use an open boundary condition, but with the following modifications. To avoid unphysical mass flow into the active domain from the inner boundary, we cap the radial velocity in the boundary ("ghost") cells at zero. Still, this does not guarantee that the mass flux through the inner boundary is non-positive. Therefore, in addition to the velocity cap, we check the total mass flux through the inner boundary; if it is positive, we scale down the mass flux at all boundary cell interfaces with positive mass flux by the ratio between the total outward flux and the total inward flux so that the total mass flux becomes zero.[4] We also cap the energy flux through each boundary cell interface to zero, in order to avoid unphysical energy injection into the domain. Meanwhile, no special treatment for the magnetic field was necessary; a simple open boundary condition does not seem to cause any numerical artifacts.

One slightly unconventional feature of our inner boundary is that we set the flux of vertical angular momentum, $L_z = \int {\rm d}V\, \rho v_\phi r \sin\theta$, through all cell interfaces on the inner boundary to zero; this prevents any angular momentum from leaving the domain through the inner boundary. This condition is motivated by the following. Physically, most mass at $r < r_{\rm in} \sim 1$ au is eventually concentrated in a small protostar, which cannot hold much angular momentum. As a result, the net angular-momentum flux though $r = r_{\rm in}$ should be negligible compared to the angular-momentum flux that an open boundary condition would allow, which is on the order of the mass flux multiplied by the local Keplerian specific angular momentum. This particular angular-momentum-flux boundary condition is seldom, if ever, used in previous studies, but, as we show in Appendix D, it is crucial for achieving numerical convergence when $r_{\rm in}$ is not very small.

To account for the mass inside the inner radial boundary, we set a point mass at $r = 0$ and update its mass according to the mass flux through the inner boundary. The gravitational potential of this point mass accretor adds to that of the self-gravitating gas inside the domain.

[4]The code computes fluxes across interfaces by solving the Riemann problem and updates the density and momentum based on these fluxes. Our boundary conditions for fluxes are applied before updating the variables using the fluxes.

Planned modifications for future studies: Some minor changes to the inner boundary condition are planned to be applied to future simulations; although we do not expect them to significantly affect the simulation outcome, these changes might help to reduce artifacts near the boundary. First, instead of using a simple open boundary condition, we plan to extrapolate density and pressure into the ghost zone. This treatment is helpful for correctly simulating the formation of the FHSC (G. Lesur, private communication). Second, for v_ϕ in the ghost zones we plan to switch from constant v_ϕ across the inner boundary to constant $\Omega = v_\phi/R$ across the inner boundary. This could reduce unphysical shearing of magnetic field across the inner boundary. We might also need to reflect on the details of the zero-angular-momentum-flux boundary condition, such as whether the energy flux should be modified accordingly and whether it would be better to turn off this boundary condition when the disk is already much larger than the inner boundary.

In addition to these minor changes in the inner boundary condition, we have also implemented a moving inner boundary which allows the location of the inner boundary to move (discretely on the grid) according to pre-defined rules, such as in proportion to disk size or circularization radius. In other words, we change our domain by adding or removing cells at the inner boundary. This allows more flexibility when balancing between integration time and resolution.

Other numerical treatments

There are several additional numerical treatments for reducing numerical cost and avoiding artifacts. First, to alleviate the stringent timestep requirement imposed by diffusive processes and high Alfvén speed in the outflow, we implement velocity caps (in the form of a density floor) and diffusivity caps, whose values are chosen based on the local dynamical timescale so that applying these caps should not significantly affect evolution. Second, to avoid the timestep being limited by the small azimuthal cell size in the polar region, we implement a polar averaging algorithm that reduces the effective azimuthal resolution near the pole and thereby allows larger timesteps. Finally, we implement a geometric correction when calculating the electric field in order to avoid artifacts in the magnetic field due to geometric errors near the pole. These treatments are discussed in detail in Appendix C.5 – C.7, respectively.

Chapter 3

Simulation meets theory: assembling a quantitative physical picture

This chapter is adapted from material published in:

Xu, W. and Kunz, M. W. (2021b). Formation and evolution of protostellar accretion discs – II. From 3D simulation to a simple semi-analytic model of Class 0/I discs. MNRAS, 508(2):2142–2168.

In this chapter we discuss our fiducial simulation and compare it with a series of analytic arguments to assemble a quantitative physical picture of disk evolution. We begin by introducing some basic definitions in Section 3.1 and providing an overview of the simulation results in Section 3.2. We then construct a quantitative physical picture of disk evolution by focusing on several aspects, one at a time. In Section 3.3, we discuss the angular-momentum budget of the disk and how disk growth and protostar accretion are self-regulated by GI. In Section 3.4, we study the thermal evolution of the disk, and show how the disk temperature profile can be estimated using simple analytic arguments. In Section 3.5, we focus on the evolution of the magnetic field, which is modulated by non-ideal MHD effects, and provide estimates for the field strength and magnetic braking efficiency inside the disk. These physical insights will be used in the next chapter to construct a simple, semi-analytic model of disk evolution.

3.1 Definitions

Before we discuss the simulation, we go over some basic definitions that will be used in this discussion.

The protostar-disk system and disk radius: We use this term to define the region that is supported against gravity (either centrifugally or by pressure), plus the point mass enclosed within the inner radial boundary of the computational domain. When the protostar first forms, this region is the pressure-supported first hydrostatic core; later on, it consists of a protostar (inside the inner boundary) and a disk (which is mainly rotationally supported). In practice, we define the radial boundary of the protostar-disk system (or disk radius, $R_{\rm d}$) as the radius where the azimuthal kinetic energy begins to exceed half of the total kinetic energy (averaged over a spherical shell).

Protostar, accretor, and disk: Although the boundary of the protostar-disk system is easy to define, defining the boundary between the protostar and the inner disk can be tricky. For simplicity, we identify the protostar mass as the mass of the point-mass accretor within the inner boundary and count the remainder of the mass in the protostar-disk system as that of the disk. Note that this definition will count the part of the first hydrostatic core in the domain as the disk.

Pseudodisk: We use the term "pseudodisk" to refer to the pre-stellar material that is pressure-supported (and thus flattened) along magnetic-field lines but is not rotationally supported in the cylindrical-radial direction.

$\boldsymbol{\Omega_{\rm K}}$ and $\boldsymbol{Q_{\rm K}}$: At a given radius, we define the rotation rate required for full (Keplerian) rotational support as $\Omega_{\rm K}(R) \equiv \sqrt{-g_R R}$, where g_R is evaluated at the disk midplane. Inside of the disk, we generally find $\Omega \approx \Omega_{\rm K}$. Because $\Omega_{\rm K}$ exhibits significantly less fluctuations than does Ω, it is useful to replace Ω with $\Omega_{\rm K}$ when evaluating some variables. We use a subscript 'K' to denote variables evaluated using $\Omega_{\rm K}$. For

example, in this paper we often characterize the strength of GI with the Toomre Q parameter,

$$Q = \frac{\bar{c}_{\rm s}\kappa}{\pi {\rm G}\Sigma}, \tag{3.1}$$

where $\bar{c}_{\rm s}$ is the (density-weighted average) sound speed, κ is the epicyclic frequency, and Σ is the column density. Because κ is very sensitive to spatial fluctuations in Ω, we often approximate Q with $Q_{\rm K}$, in which case we replace Ω by $\Omega_{\rm K}$ when evaluating the epicyclic frequency.

3.2 Overview of evolution

In this section we provide an overview of the evolution of the protostellar core (and the disk formed within) using Figs 3.1–3.7, which show azimuthally averaged (Figs 3.1–3.4) and midplane (Fig. 3.5) profiles at different epochs, the evolution of midplane temperature with density (Fig. 3.6), and the time evolution of the protostellar disk mass and size (Fig. 3.7). We focus on describing the simulation results qualitatively; the detailed (quantitative) physical picture of disk formation and evolution is discussed in Sections 3.3–3.5.

3.2.1 Pre-stellar collapse and disk formation

During the first $\sim$1 Myr of evolution, the pre-stellar core flattens along magnetic-field lines into a (not rotationally-supported) pseudodisk and radially contracts dynamically (though slower than free fall). The magnetic field is dragged along with the collapsing core and becomes hourglass-shaped (first row of Fig. 3.1).

The pseudodisk acquires a flat column density Σ in its central region where thermal pressure smooths out any structure and a near-self-similar profile close to $\Sigma \propto R^{-1}$ outside this region. The magnetic field remains relatively well-coupled to the gas for most of the pre-stellar collapse, and the profile of the vertical magnetic field remains similar to that of the column density, with a uniform central region and a $B_z \propto R^{-1}$ profile further out. This indicates the maintenance of a near-constant mass-to-flux ratio during this phase of the collapse, consistent with a low level of ambipolar diffusion. The angular momentum of the infalling gas is also nearly conserved, because the magnetic-braking timescale is longer than the infall timescale (Section 3.3). The temperature remains close to 10 K during this phase, as the core remains optically thin.

At a central density of $n_{\rm n} \sim 10^{11}$ cm^{-3}, the central column density becomes sufficiently large and the gas in the innermost few au becomes optically thick and the temperature increases (see Figs 3.2 and 3.6 at $\leq$0.01 M$_\odot$). The additional pressure support leads to the formation of a first hydrostatic core with $n_{\rm n} \gtrsim 10^{12}$ cm^{-3}. At the same time, ambipolar diffusion "reawakens" to become increasingly important as the charged species are adsorbed onto the dust grains, thereby decoupling all species but the electrons from

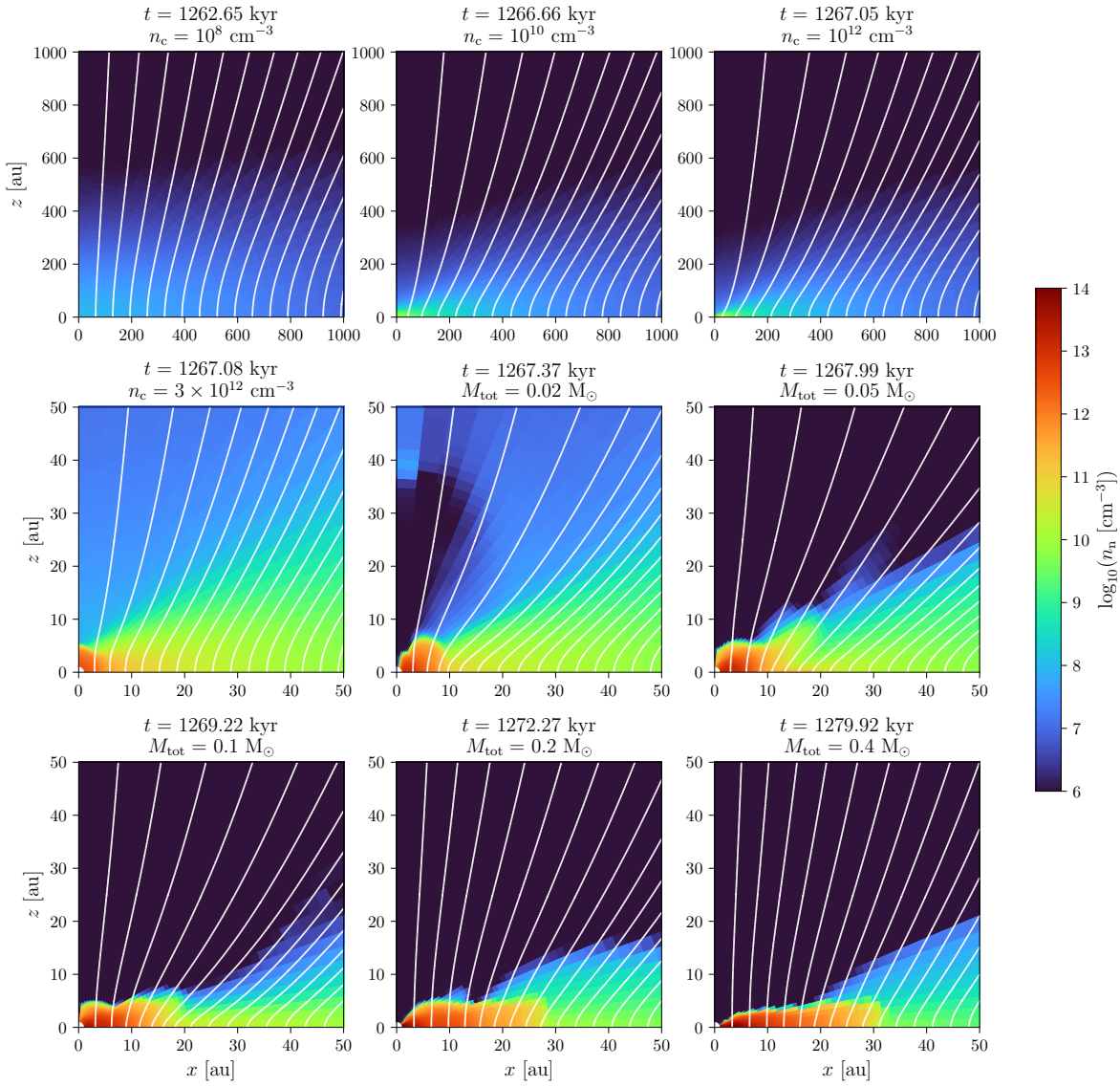

Figure 3.1: Zoom-in snapshots of the azimuthally averaged density profile (colour) and magnetic-field lines (white) at a few different epochs. For snapshots before protostar (first hydrostatic core) formation, we label the central density n_c; for snapshots after protostar formation, we label the total mass of the protostar and rotationally supported disk M_{tot}, which is frequently used as a proxy of time in this thesis. The three rows correspond to pre-stellar collapse, first hydrostatic core and disk formation, and disk evolution, respectively.

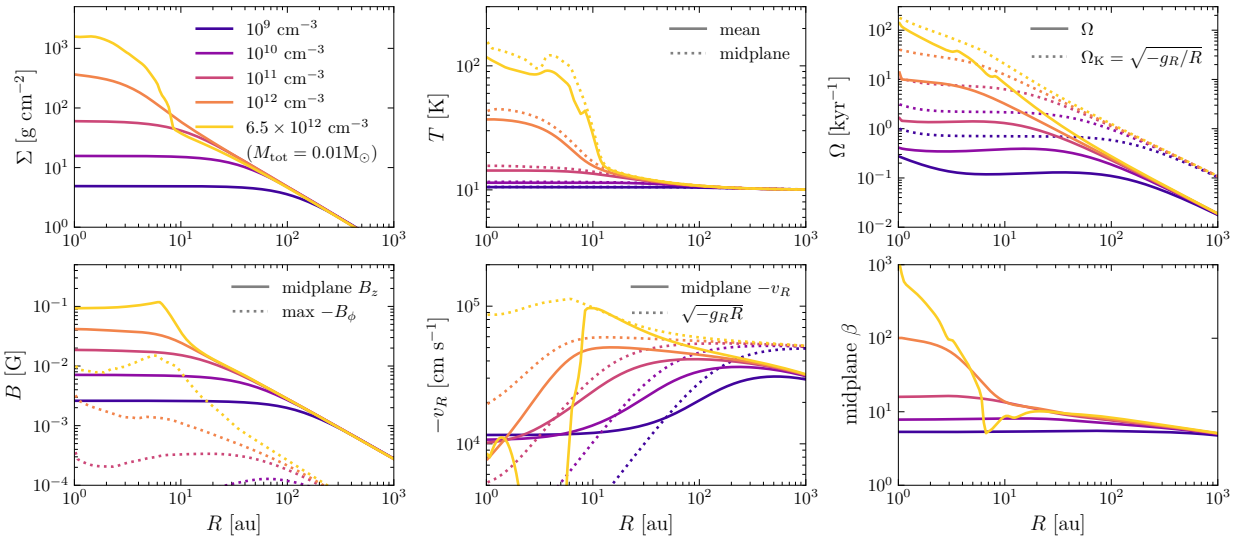

Figure 3.2: Azimuthally averaged column density (top left), mean (density-weighted) and midplane temperatures (top center), mean (density-weighted) rotation rate alongside the Keplerian rotation rate (top right), vertical and azimuthal components of the magnetic field (bottom left), midplane radial velocity (bottom center), and midplane plasma β (bottom right) in the innermost 10^3 au of the core at a few different epochs covering the self-similar collapse and the formation of the first hydrostatic core. The epochs are labeled by central density $n_{\rm c}$.

the magnetic field and causing the magnetic flux to pile up just outside of the hydrostatic core (see, e.g., Tassis & Mouschovias 2007 and Kunz & Mouschovias 2010).

The first hydrostatic core also marks the epoch of point-mass formation, which creates an expanding region (reaching $\gtrsim 10^3$ au at the end of our simulation) in which gravity is dominated by the point mass instead of the pseudodisk. Within this region, the pseudodisk profile flattens from the pre-stellar profile with $\partial \ln \Sigma / \partial \ln R \approx -1$ to one having $\partial \ln \Sigma / \partial \ln R \approx -1/2$ (Fig. 3.3; see also discussion in Dapp et al. 2012, section 7.1). The accelerated infall of the gas and the consequent pinching of the magnetic-field lines push the location of magnetic decoupling (and the resulting pile-up of magnetic flux) to larger radii (Section 3.5.1; see also Ciolek & Königl 1998 and Contopoulos et al. 1998).

The first hydrostatic core continues to accrete mass and angular momentum from the pseudodisk, and the accumulation of angular momentum soon shapes it into a rotationally supported torus (second row of Fig. 3.1). This torus quickly becomes gravitationally unstable, and the non-axisymmetric perturbations associated with GI (see Fig. 3.5) transport angular momentum outwards to allow the formation of a central protostar and a growing, rotationally supported disk with $\Omega \approx \Omega_{\rm K}$ (see Fig. 3.3).[1] The above evolution takes place within the first $\sim$kyr after first hydrostatic core formation.

[1]The gravitational instability of a self-gravitating torus has been discussed in Tohline & Hachisu (1990). It is different from the gravitational instability in a rotationally supported disk in that gravity is not dominated by a central point mass, but the mechanism of instability is largely similar.

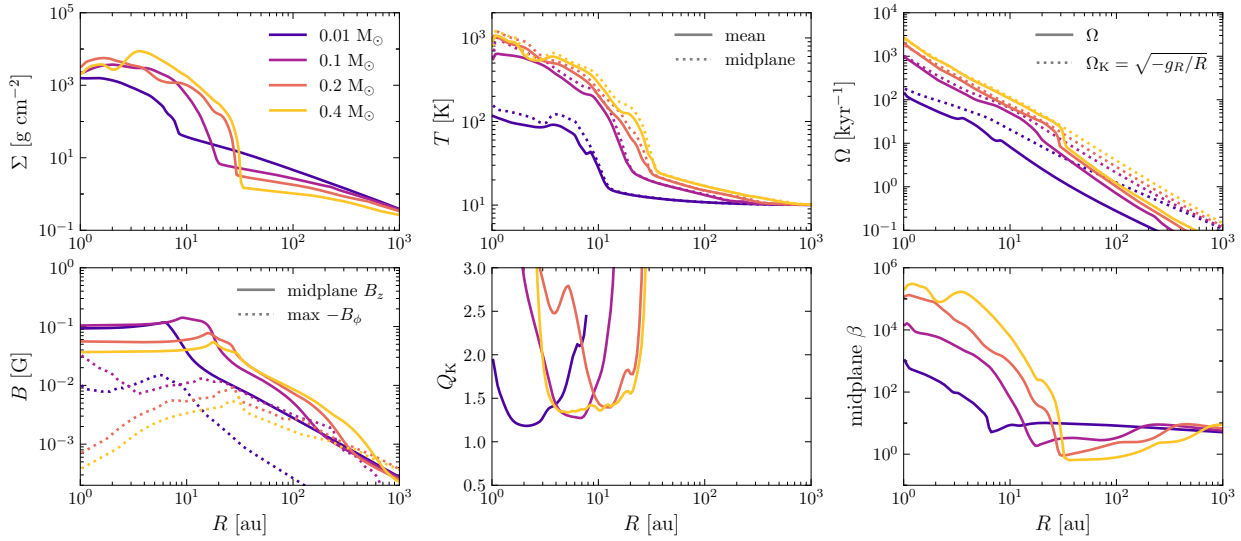

Figure 3.3: Similar to Fig. 3.2 but covers the formation and evolution of the protostellar disk after first hydrostatic core formation, with each epoch labeled by the total mass of the disk and the protostar, $M_{\rm tot}$; the first line shows the same epoch as the last line in Fig. 3.2. We also replaced the radial velocity panel with profiles of $Q_{\rm K}$ (a proxy of Toomre Q, bottom center). The disk that forms within $R \lesssim 30$ au is hot, near-Keplerian, and gravitationally unstable, with an approximately uniform, straight, and dynamically unimportant magnetic field.

3.2.2 Disk evolution

After the rotationally supported disk appears, the disk remains gravitationally unstable and the disk mass remains comparable to the protostar mass. Meanwhile, the disk size undergoes substantial growth and saturates at ∼30 au in radius (Figs 3.5 and 3.7; see also the last row of Fig. 3.1). The saturation of the disk size occurs mainly because increased magnetic braking in the pseudodisk at later times prevents the disk from accumulating enough angular momentum to grow further (see Section 3.3.1).

The disk size is significantly larger than the circularization radius of incoming material (Fig. 3.7; this also leads to the sudden change in Ω at the disk edge seen in Fig. 3.3). This is because the accreted material leaves most of its angular momentum in the disk before being accreted by the protostar, which can hold approximately zero angular momentum. This angular momentum is then transported outwards by GI to cause disk spreading (see Section 3.3.2). GI also gives rise to various substructures, such as prominent gravitationally excited spiral arms (see Fig. 3.5).

The temperature inside the disk is typically a few $\times 10^2$ K (Fig. 3.3). While the barotropic equation of state we employed in XK21a fits well the temperature evolution at early times (before disk formation), the relation between temperature and density is far from barotropic once the protostellar disk forms (Fig. 3.6), with the temperature being generally larger than predicted by the barotropic relation (which is based on an empirical central density-temperature relation motivated by radiative non-rotating core-collapse simulations).

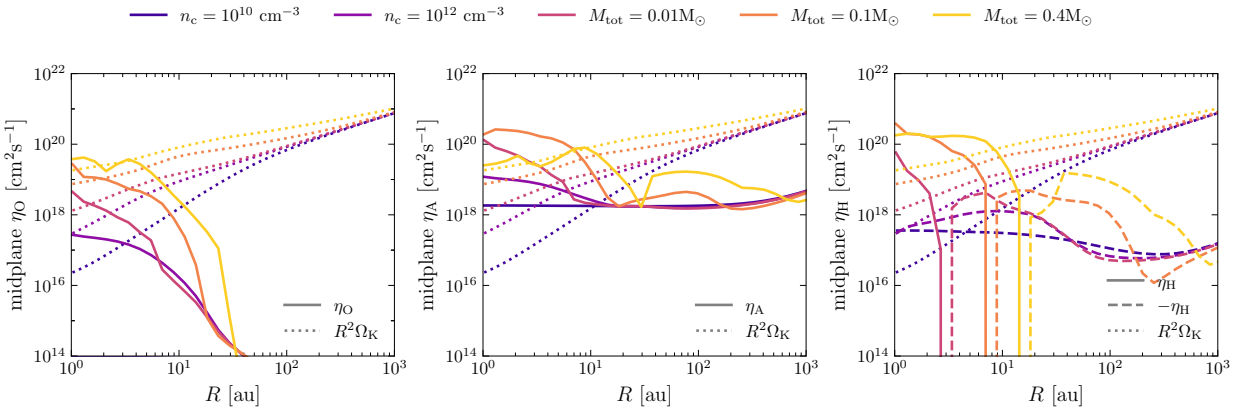

Figure 3.4: Midplane non-ideal MHD diffusivities at a few different epochs before and after disk formation. Although the Hall effect has not been included in the simulation, we still plot η_H in the right panel for reference.

This suggests that using a barotropic equation of state to model young protostellar disk evolution cannot produce the correct thermal evolution. We discuss what determines the disk temperature profile in Section 3.4.

The magnetic field inside the disk is approximately uniform, and the ratio between thermal and magnetic pressure $\beta \equiv 8\pi P/B^2 \gg 1$ (Fig. 3.3). This weak and nearly uniform field is mainly a result of non-ideal magnetic diffusion (Ohmic and ambipolar) decoupling the field from the gas in the radial direction, which we discuss in Section 3.5. The field is also approximately straight inside the disk, with a very small azimuthal component (Fig. 3.3); we discuss the generation of an azimuthal field and the strength of magnetic braking inside the disk in Section 3.5.2.

The magnetic field threading the disk also launches a relatively weak outflow, which removes little mass and angular momentum from the disk and barely affects disk evolution (Section 3.3.1). Qualitatively speaking, the weak outflow is because the generation of azimuthal magnetic field is inefficient in our disk given the strong ambipolar diffusion (Section 3.5.2). We refrain from a more detailed discussion of outflow properties (especially those far above the disk, which are more relevant to existing observations) in this thesis, as the propagation of the outflow far above the disk might be significantly impacted by numerical factors such as the low resolution in the polar region, the use of a density floor (Appendix C.5), and the omission of protostellar irradiation (Section 2.4.3). We also do not resolve possible outflow launching at small radii (< 1 au) where the disk could be thermally ionized and better coupled to the field. Such outflow is likely unable to significantly impact the angular-momentum budget of the disk given the small launching radii, but may impact the mass budget of the system.

A note on the significance of outflow: The relative unimportance of outflow in our simulation may

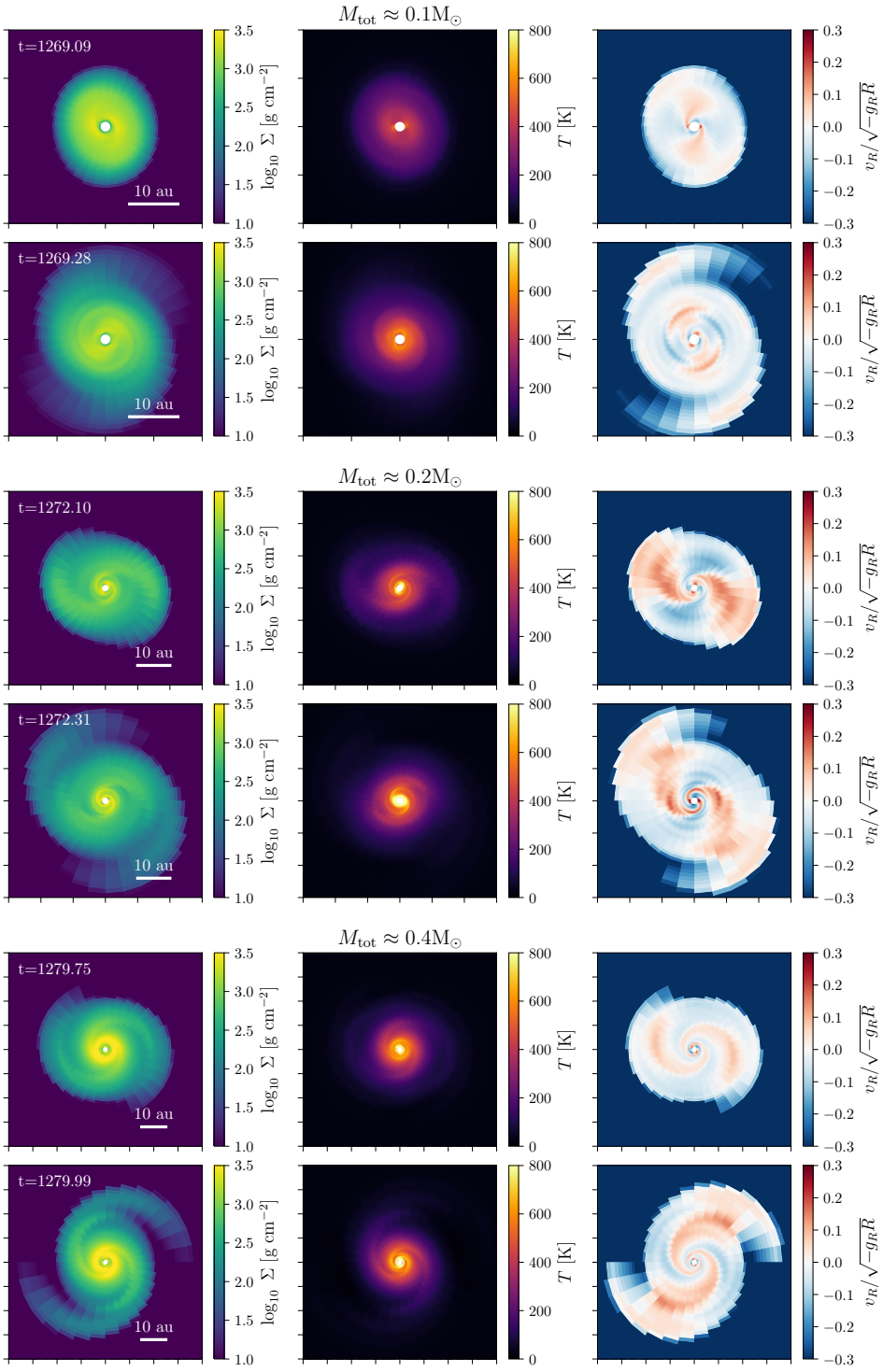

Figure 3.5: Disk column density (left), mean temperature (centre), and radial velocity (right) as viewed from above. Each row corresponds to a different epoch; to provide an unbiased selection, for each $M_{\rm tot}$ we plot the snapshots that exhibit the minimum and maximum disk size measured around that $M_{\rm tot}$. The disk size undergoes significant oscillations as spirals emerge and disperse. A rich display of substructures, mostly spirals, is visible.

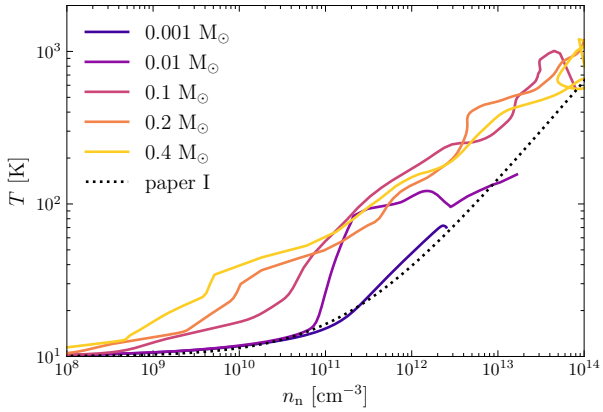

Figure 3.6: Relation between azimuthally averaged midplane temperature and density at a few different epochs. The barotropic equation of state used in XK21a (paper I) is shown for reference, and fits well the temperature evolution at early times. With our treatment of radiative cooling, the temperature in the disk is far from barotropic.

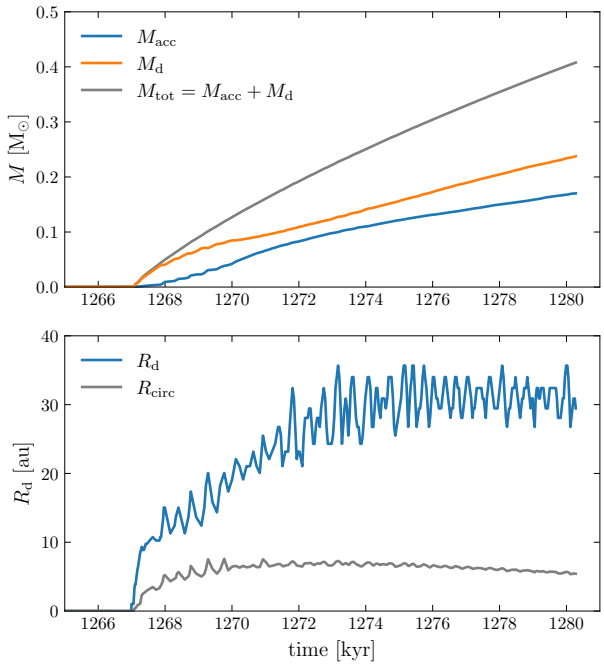

Figure 3.7: Time evolution of disk and accretor (protostar) mass (top panel) and disk radius (bottom panel). The circularization radius of accreted material, $R_{\rm circ} \equiv l^2/(\mathrm{G}M_{\rm tot})$, where l is the specific angular momentum of material accreted onto the disk, is shown for reference. The disk remains slightly more massive than the protostar. The disk radius is significantly larger than the circularization radius of incoming material, as a result of angular-momentum transport (redistribution) within the disk.

appear somewhat surprising, given that several previous simulations (e.g., Machida & Hosokawa, 2013; Tomida et al., 2017) found that outflow can remove a significant amount of mass and angular momentum. One important difference between our model and these simulations is that these studies adopt simple parametrizations of non-ideal diffusivity on density and temperature (cf. Machida et al. 2007 Eqs. 9–10 and Tomida et al. 2017 Eqs. 2–3) while we adopt a more realistic profile of magnetic diffusivity based on an equilibrium chemical network. These parametrizations generally underestimates the level of magnetic diffusion inside the disk, which could lead to a stronger torial field and more outflow launching. Our result is also consistent with recent observations by Sadavoy et al. (2019), which find no strong toroidal fields (at 35-au resolution) in a protostellar disk survey. There are also some differences in the initial conditions, including our adoption of a stronger, near-critical initial magnetic field, which could lead to a smaller disk and substantially weaker outflow as suggested by results from Machida & Hosokawa (2013). Overall, it is important to verify in future studies whether our result is applicable over a broader range of initial parameters. Finally, we note that, although observations of relatively high specific angular momentum in outflows (e.g., Bjerkeli et al., 2016) are sometimes used as evidence that outflows can remove a significant amount of angular momentum, this argument is not grounded firmly unless the mass-loss rate by outflow and the accretion rate of the disk are also known.

3.3 Angular-momentum budget and transport

In this section we study the angular-momentum budget of the disk. Considering the protostar and the disk as a whole, we separate the problem into a study of how much mass and angular momentum enters and leaves the protostar-disk system (Section 3.3.1), and a study of how angular momentum is redistributed within the protostar-disk system to allow accretion and disk growth (Section 3.3.2). We also include a discussion of how the redistribution of angular momentum by GI leads to formation of radial substructures and the spread of the gravitationally unstable region (Section 3.3.3).

3.3.1 Mass and angular-momentum budget of the protostar-disk system

Mass and angular momentum enters and leaves the protostar-disk system through several channels. Fig. 3.8 plots the mass and angular-momentum flux through a sphere of radius slightly larger than $R_{\rm d}$ from different channels; these are the mass flux from inflow/accretion[2] (region with $v_r < 0$) and outflow (region with

[2]The accretion rate here refers to accretion from the pseudodisk onto the protostar-disk system, rather than from the disk onto the protostar; the rate corresponding to the latter is similar to the former but exhibits much higher variability (cf. Machida & Basu, 2019). In this paper we forego any discussion of the variability of the accretion rate onto the central protostar, as it can be sensitive to the treatment of the inner boundary.

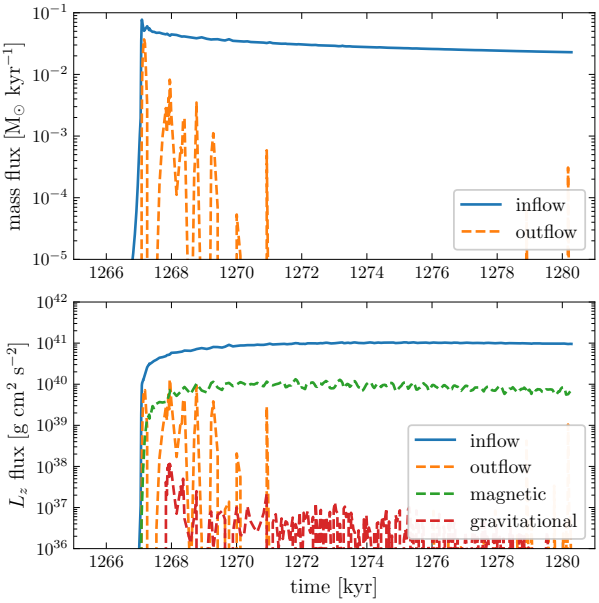

Figure 3.8: Mass (top panel) and angular-momentum (bottom panel) fluxes going into (solid) and out of (dashed) the protostar-disk system via different mechanisms. These fluxes are evaluated on a sphere with radius slightly larger than $R_{\rm d}$. Both the mass and angular-momentum fluxes are dominated by inflow (accretion).

$v_r > 0$), and the angular-momentum flux from radial advection ($R\rho v_r v_\phi$) by inflow and outflow and from transport by magnetic and gravitational stresses ($-RB_rB_\phi/4\pi$, $Rg_rg_\phi/4\pi$G). Both the mass and angular-momentum fluxes are dominated by accretion (inflow), which happens mostly through the pseudodisk instead of the envelope (because the latter has negligible density). The angular-momentum flux from the magnetic stress, which is approximately the rate of angular-momentum removal by disk magnetic braking, remains about one order of magnitude lower than the angular momentum injection by accretion. Other mechanisms of mass and angular-momentum removal are even weaker compared to accretion. Therefore, nearly all mass and angular momentum accreted by the protostar-disk system stays within the protostar-disk system. As a result, the mass and angular-momentum budget of the protostar-disk system can be understood mostly in terms of the accretion of mass and angular momentum from the pseudodisk. In the remainder of this subsection, we discuss how the infall of the pseudodisk determines the accretion rate of mass and angular momentum.

The mass accretion rate from the infalling pseudodisk peaks around the time of protostar formation, and decreases at later times, with a typical accretion rate of a few 10^{-5} M$_\odot$ yr^{-1} (Fig. 3.8). This accretion rate comes from the near self-similar radial collapse of the pseudodisk, which is driven by self gravity and is countered primarily by the pressure gradient and magnetic tension. Although the initial mass-to-flux ratio of the pre-stellar core is just slightly supercritical and B_z/Σ does not decrease significantly during

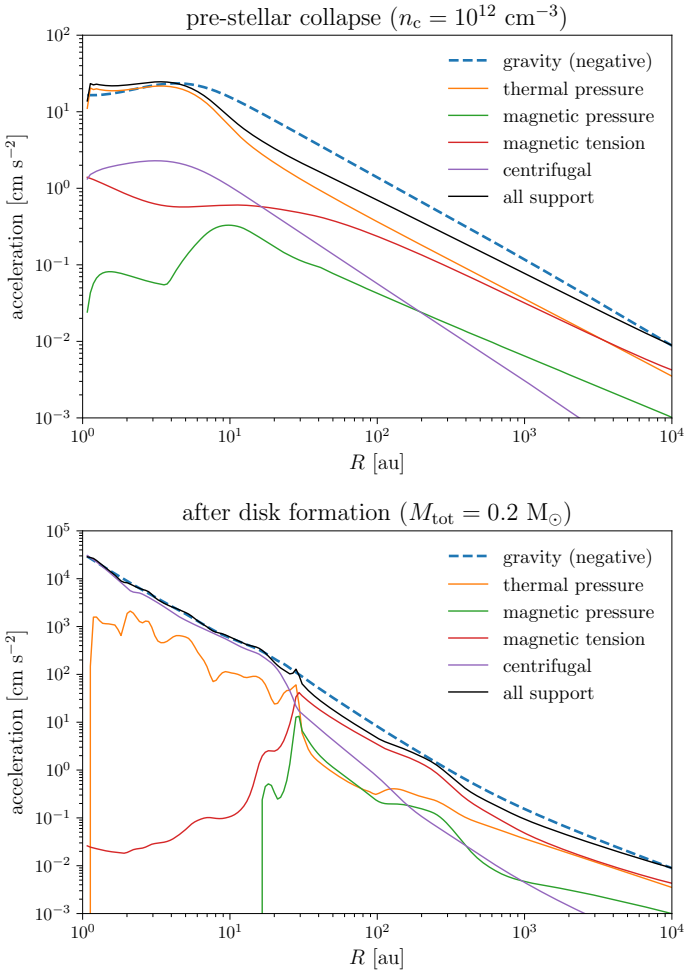

Figure 3.9: Radial acceleration at the midplane from different sources at representative epochs before (top panel) and after (bottom panel) disk formation. In the pseudodisk, although some supporting forces are of the same order of magnitude as gravity, the sum of them (black line) remains insufficient to balance gravity, and the pseudodisk is in near free-fall.

most of the infall, the pressure gradient and magnetic forces remain at least a factor of a few smaller than gravity, and the acceleration is $\gtrsim$50% of free fall in the bulk of pseudodisk (Fig. 3.9). Still, this is not to say that these retarding forces are unimportant. On the contrary, the formation of the self-similar column density profile requires a small but nontrivial ratio between these forces and self gravity (similar to the role of pressure in the spherical collapse model of Larson 1969), and this sets a characteristic accretion rate that depends only on the isothermal sound speed $c_{\rm s}$ and G. [More generally, the accretion rate may also depend on how thermally and magnetically supercritical the initial pre-stellar core is, and there exists a wide range of different self-similar solutions as discussed in Hunter (1977). Here we only consider cores that are initially near critical.] The mass accretion rate calculated from the self-similar model of Basu (1997) for the collapse of a slightly magnetically supercritical thin sheet is 13 $c_{\rm s}^3/G \sim 2 \times 10^{-5}$ M$_\odot$ yr^{-1}, similar to our result.

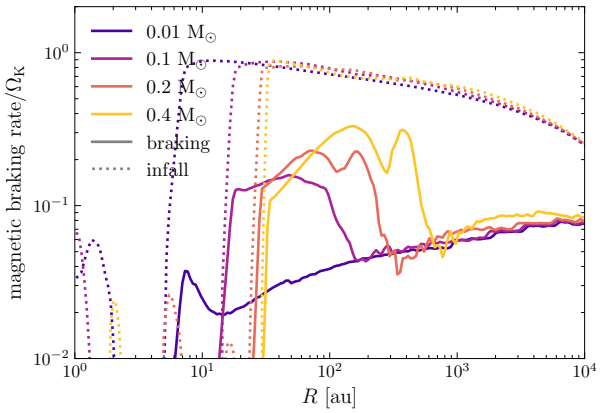

Figure 3.10: Comparison between pseudodisk magnetic-braking rate (magnetic torque divided by angular momentum) and infall rate ($-v_r/r$), both normalized by $\Omega_{\rm K}(R)$ (which is also approximately the local free-fall rate), at a few different epochs. (See Section 3.3.1 for how these quantities are obtained.) Magnetic braking is relatively weak in the outer pseudodisk. In the inner pseudodisk, however, the magnetic-braking rate can be comparable to the infall rate, and angular-momentum removal by magnetic braking becomes increasingly important at late times.

Next we consider the accretion of angular momentum. One informative diagnostic is the circularization radius of the material accreted from the pseudodisk, which is shown in the bottom panel of Fig. 3.7. In the absence of angular-momentum transport within the disk, this circularization radius would define the size of the disk. We see a roughly constant circularization radius, which implies that the specific angular momentum of material accreted at a given $M_{\rm tot}$ scales roughly as $l \propto M_{\rm tot}^{1.5}$. If angular momentum is conserved during the infall, this scaling should have a slope between 5/3 (when the magnetic field is relatively weak and infall is approximately in the spherical-radial direction) and 2 (when the magnetic field is very strong and all material first falls along magnetic-field lines to the midplane and then collapses radially). That the observed slope deviates from this range suggests the presence of magnetic braking in the pseudodisk.

In Fig. 3.10 we compare the rate of magnetic braking with the rate of infall. To compute the braking rate, we first label each (R, ϕ, z) location with the amount of magnetic flux threading a disk of radius R at height z, denoted $\Phi_B(R, z)$. [This procedure gives the same Φ_B for all ϕ at a given (R, z).] The braking rate at a given radius R_0 is then found by integrating the magnetic torque on an infinitesimal flux tube with $\Phi_B(R, z)$ between $\Phi_B(R_0, 0) \pm {\rm d}\Phi_B$; because the field is close to axisymmetry, this flux corresponds approximately to the region along the field lines starting from $R_0 \pm {\rm d}R$. However, we do not integrate through the entire flux tube – magnetic braking involves the transport of angular momentum along field lines from the collapsing pre-stellar core (including the disk that forms within the core) to the low density, magnetically dominated, non-collapsing background, and integrating through the entire flux tube would have the torques in these two regions cancel out. Instead, we separate the core from the background, and integrate only the

magnetic torque in the core. To do so, we define the core as the region where the magnetic energy is less than the sum of the kinetic and internal energies (i.e., not magnetically dominated). The braking rate at each cylindrical radius is then defined as the ratio of the magnetic torque and the angular momentum, both integrated in the "core" part of each corresponding flux tube. The main advantage of this definition is that it focuses on angular-momentum removal from the core and is unaffected by angular-momentum transport along magnetic-field lines within the core (which should barely affect the angular-momentum budget of the system). The infall rate is calculated as a density-weighted average of $-v_r/r$ in the same ("core" flux-tube) region.

The magnetic braking rate shown in Fig. 3.10 is consistent with the evolution of $R_{\rm circ}$ we observe. The braking rate generally increases in time, and at 0.4 $M_\odot$, braking is only a factor of $\approx$3 slower than infall in the inner part of the pseudodisk, which allows magnetic braking to affect the angular-momentum budget significantly. In Section 3.5.1 we show that the increased braking at late times can be explained by the radial decoupling of the magnetic field from the gas, which leads to a pile-up of magnetic flux in the inner part of the pseudodisk. We also show in Section 4.3 that we expect this trend of increased braking to continue, and that the pseudodisk magnetic braking should eventually become strong enough to affect significantly the angular-momentum budget and disk size evolution.

In summary, the mass and angular-momentum budgets of the disk are determined mainly by accretion from the pseudodisk. The angular-momentum budget of the disk is affected by magnetic braking in the pseudodisk, which is initially weak but becomes increasingly important at later times; the increased magnetic braking at later times is partly responsible for the saturated disk size observed in Figure 3.7.

3.3.2 Angular-momentum transport within the protostar-disk system: GI

Material in the disk has to lose angular momentum in order to be accreted by the protostar. Since mechanisms that remove angular momentum from the disk all appear to be inefficient, accretion requires some mechanism that transports angular momentum radially outwards (which also leads to disk spreading). In our simulation, such angular-momentum transport is facilitated mainly by GI.[3]

In the top panel of Fig. 3.11 we plot the angular-momentum flux through spherical (r) shells due to different mechanisms. Since studies in the literature often characterize angular momentum transport by an α viscosity (Shakura & Sunyaev, 1973), we also plot the effective α for these mechanisms, which are defined by the ratio between stress and pressure (both integrated on spherical shells). Angular-momentum transport (colored lines) inside the disk is dominated by turbulent advection (Reynolds stress, blue) and gravitational

[3]The magnetorotational instability, another popular mechanism for radial angular-momentum transport in differentially rotating disks (Balbus & Hawley, 1998), is fully suppressed by the strong magnetic diffusion ($\eta \gg \Omega H^2$) in our disk (cf. Kawasaki et al., 2021).

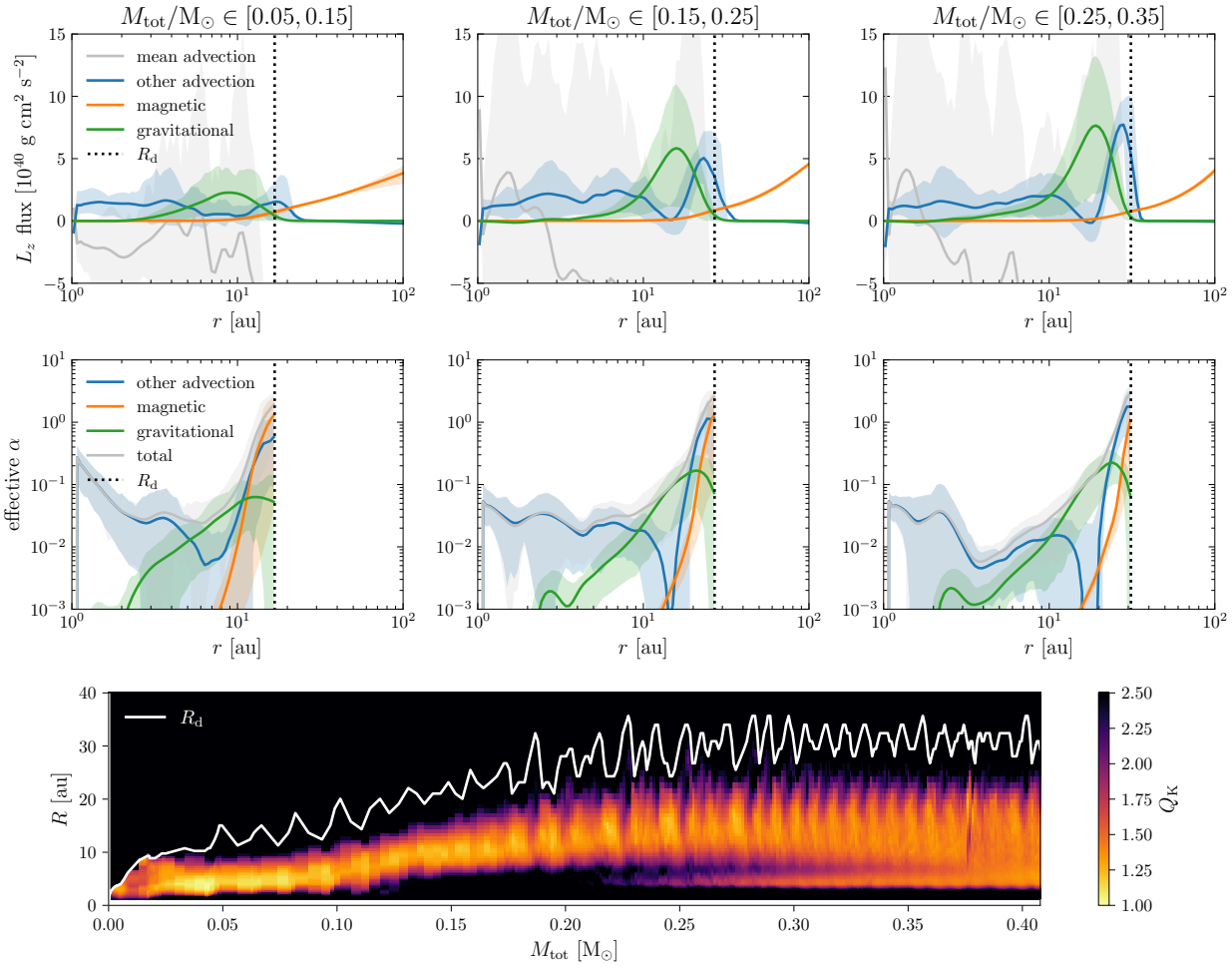

Figure 3.11: Top: Radial profiles of angular-momentum flux due to different mechanisms. Each line traces the mean value across all snapshots in the stated $M_{\rm tot}$ range; the shaded region indicates 1σ deviations about this mean. We decompose the advection contribution into that associated with the mean flow (viz., mass flux times density-weighted mean specific angular momentum; grey) and the residual "other advection" (blue), which is due mainly to transport by the spiral waves. Inside the disk, the angular-momentum transport not associated with mean advection is driven primarily by GI-induced gravitational stresses (green) and spiral waves (blue). Magnetic braking (orange) remains weak inside the disk. Middle: Effective α (the ratio between stress and pressure) in the disk corresponding to these angular momentum transport mechanisms. Bottom: Space-time evolution of $Q_{\rm K}$ (a proxy of Toomre Q) inside the disk; the disk edge $R_{\rm d}$ is plotted for reference. When the total protostar-disk mass $M_{\rm tot} \in [0.12, 0.24]$ M$_\odot$, the inner part of the disk is gravitationally stable and angular-momentum transport there is due to spiral waves excited radially farther out. Later on, a second gravitationally unstable region appears at smaller radii, and most of the disk eventually being marginally gravitationally unstable (see Section 3.3.3).

stress (green), which can both be interpreted as the outcome of spiral waves excited by GI. Note that the gravitational stress is confined mostly to regions that are locally gravitationally unstable (cf. bottom panel of Fig. 3.11), but the turbulent advection is not. The non-negligible amount of turbulent advection seen in the gravitationally stable regions arises from propagation of spiral waves into these regions (see Section 3.3.3).

The strength of GI, which determines the rate of angular-momentum transport, can be quantified using the Toomre Q parameter, with smaller Q giving stronger instability and faster angular-momentum transport, and hence leading to larger accretion rates. On the other hand, stronger instability and faster accretion tend to increase disk temperature (due to additional turbulent heating as well as additional release of gravitational energy via accretion) and reduce column density (by reducing disk mass and increasing disk size); this causes Q to increase. As a result, the disk should settle into a self-regulated state in which the disk is marginally unstable and GI produces just the right amount of angular-momentum transport and heating to maintain approximately constant Q (Vorobyov & Basu, 2007; Kratter & Lodato, 2016). Additionally, the strength of GI scales steeply with Q: at $Q \sim 2$, GI is turned fully off, but at $Q \sim 1$, GI becomes so strong that the resulting angular-momentum transport is generally much faster than needed for the self-regulated equilibrium. Therefore, without knowing any details of the evolution, the value of Q at this self-regulated state can be estimated to within a factor of 2.

An important implication of this idea of gravitational self-regulation is that one can easily estimate the disk profile (including disk size and mass) by assuming that the whole disk has some constant Q, if one knows the total mass and angular momentum of the disk (discussed in the previous subsection) and the thermal profile of the disk (Section 3.4). We apply this idea to build a simple 1D disk model in Chapter 4.

One caveat, however, is that the disk being gravitationally self-regulated does not directly imply that most of the disk needs to be marginally unstable. This is mainly because GI excites spiral waves that can propagate into gravitationally stable regions and transport angular momentum there. In the next subsection we discuss this process in more detail, and argue that most of the disk should eventually be marginally gravitationally unstable.

3.3.3 Size of the gravitationally unstable region

While we expect GI to account for most of the angular-momentum transport everywhere in the disk, this does not require the entire disk to be gravitationally unstable. Instead, as we show in Fig. 3.11 and discuss further below, it is possible for a wave excited in one unstable region to propagate to a stable region and deposit some of its angular momentum there. Furthermore, because the angular-momentum flux needed

to facilitate accretion generally decreases towards smaller radii, the spiral waves excited in the outer part of the disk in principle have sufficient angular-momentum flux to provide all angular-momentum transport needed at smaller radii. (Note that the angular-momentum flux of a spiral wave remains constant as the wave propagates if the wave is not damped through dissipative processes such as shocks.) Therefore, while the need for GI to transport angular momentum requires a gravitationally unstable outer disk, it is not certain whether the inner part of the disk needs to be unstable as well.

In our simulation, we do see the inner disk to be gravitationally stable for a period of time around $M_{\rm tot} \sim 0.2$ M$_\odot$. However, a second gravitationally unstable region appears later on at $\approx$4 au and eventually merges with the first, outer unstable region to make most of the disk unstable (Fig. 3.11). In the discussion below, we explain this behaviour and argue that the appearance of additional gravitationally unstable regions at yet smaller radii should be a generic behaviour and that, eventually, most of the disk should be unstable.

First, let us consider the evolution of column density at a locally stable patch of the disk in which angular-momentum transport is facilitated by spiral waves excited at some unstable region at larger radii. Using mass and angular-momentum conservation, we have

$$\partial_t(2\pi R\Sigma) = -\partial_R F_M, \quad \partial_t(2\pi R\Sigma l) = -\partial_R(F_L + F_M l). \tag{3.2}$$

Here F_L is the (positive) angular momentum flux by spirals though radius R, and $F_M = -\dot{M}(R)$ is the mass flux. Let us assume for simplicity that the specific-angular-momentum profile $l(R)$ stays constant in time, which is equivalent to assuming that the timescale of column-density evolution is $\gg\Omega^{-1}$ and $\ll M_{\rm tot}/\dot{M}_{\rm tot}$. This produces a simple relation between the accretion rate (mass flux) and the derivative of F_L; using equation (3.2),

$$-F_M = (\partial_R l)^{-1}\partial_R F_L. \tag{3.3}$$

In a gravitationally stable region, F_L stays constant when the amplitude of the spiral wave is small and damping (mainly through shock formation) is negligible, and $\partial_R F_L$ is mainly due to the damping of the wave. Therefore, in a gravitationally stable region, accretion (inward mass flux) through a radius is directly proportional to the amount of spiral wave damping at this radius.

Now we consider the radial variation of spiral wave damping and how that affects the disk mass distribution. A spiral wave excited in the gravitationally unstable region generally has a pattern speed $\Omega_{\rm p}$ comparable to Ω at the location of its excitation. Let us assume either that the unstable region is narrow or that the location where waves propagating outside the GI region are excited is close to the inner boundary of the GI region (which is a relatively good assumption when the GI region is extended, because spiral waves

cannot propagate far inside a GI region; cf. Béthune et al. 2021). In this case, the inner Lindblad resonance (ILR), which occurs at $m(\Omega - \Omega_{\rm p}) = \kappa$ (with $m = 2$ for the dominant mode), lies outside the GI region. The location of the ILR ($R_{\rm ILR}$) is important because we expect an increase in wave-disk interaction (which makes the wave more nonlinear and, for instance, more prone to shock formation) and wave damping around the resonance. Since mass flux is proportional to spiral-wave damping [equation (3.3)], the mass flux should also increase here, and so mass should tend to pile up at radii slightly smaller than $R_{\rm ILR}$.

Fig. 3.12 shows that the profile of the mass flux and the evolution of the mass distribution is consistent with this theory: the region around the ILR remains close to the local maximum of the mass flux, and there is a pile-up of mass at smaller radiii, which eventually becomes the second gravitationally unstable region seen in Fig. 3.11. Note that the ILR is not the only factor that affects the damping of spiral waves (and thus the mass flux). For example, we expect the waves to be more nonlinear and to undergo more damping (via steepening and shocking) when the column density is low (e.g., between 5 and 7 au in Fig. 3.12); this explains why the peak in the mass-flux profile is slightly to the left of the ILR marked in the bottom panel of Fig. 3.12.

Eventually, the mass pile-up at $R \lesssim R_{\rm ILR}$ causes that region to become gravitationally unstable as well. This changes how the column-density evolution is regulated, because now the excitation (or amplification) of waves by GI also affects $\partial_R F_L$. Additionally, waves are now excited at multiple locations, with different pattern speeds. We therefore expect the feature around $R_{\rm ILR}$ to be less significant, if not to disappear eventually, thus providing a reason for the eventual merging of the two GI regions we see near the end of the simulation.

In reality this process would likely repeat itself: after the two GI regions merge, a significant portion of the angular-momentum transport at yet smaller radii should come from spiral waves excited around the inner region of that extended GI region. Mass can then pile up beyond the ILR of this wave and create another unstable region, which again should merge with the outer GI region. Given enough time, we expect this process to repeat indefinitely and eventually the whole disk (except for the outer edge, where transport is dominated by the dissipation of outwardly propagating waves) should become unstable. It is difficult to verify this idea directly with our simulation, however, because the new location of the ILR is too close to the inner radial boundary of the computational domain.

The mass pile-up process we describe above happens on a characteristic timescale $\sim \pi R^2 \Sigma / \dot{M}(R)$, which can be comparable to the timescale of evolution of the entire system, $\sim M_{\rm tot}/\dot{M}_{\rm tot}$, when the disk-to-star mass ratio is order unity. If the evolution shown in our simulation is indeed generic, then it should be possible to find systems showing a gravitationally stable inner disk or exhibiting radial substructure in their (azimuthally averaged) column-density profile.

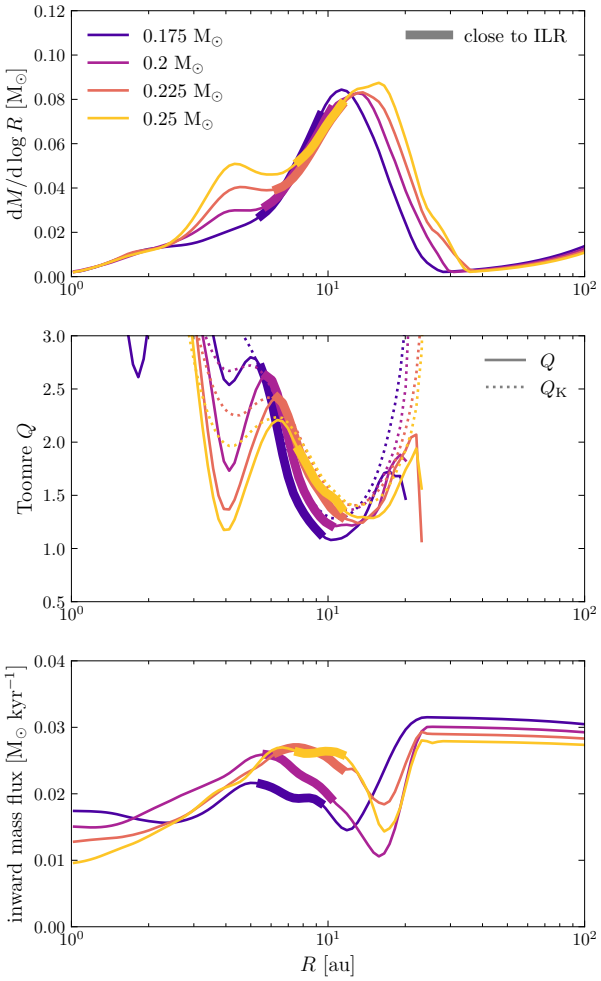

Figure 3.12: Mass distribution, Toomre Q profile (with $Q_{\rm K}$ plotted for reference), and inward mass flux during the development of the second gravitationally unstable region. Each curve is averaged using 10 snapshots to suppress random fluctuations. The pattern speed of the spirals, $\Omega_{\rm p}$, is measured directly from multiple snapshots. The region within 20% of the inner Lindblad resonance [ILR, where $m(\Omega - \Omega_{\rm p}) = \kappa$ with $m = 2$ for the dominant spirals] is marked with a thick line. (The region around the ILR is marked to account for fluctuations in κ.) Enhanced spiral-wave damping around the ILR increases the accretion mass flux, causing mass pileup at slightly smaller radii; this eventually leads to the formation of the second GI region.

3.4 Energy budget and disk temperature profile

The disk temperature profile is determined by a balance between heating and radiative cooling. We assess this balance by first estimating the heating rate (which is also approximately the rate of radiative cooling; Section 3.4.1) and then finding the relationship between the cooling rate and the disk temperature (Sections 3.4.2 and 3.4.3).

3.4.1 Rate of heating and cooling

We begin our estimate of the heating and cooling rates in the disk with a simple argument. As material in the disk moves radially inwards to be accreted, gravitational energy is released at a rate of $-g_R\dot{M}(R)$ per unit radius. A fraction of this energy (half of it, in the case of a Keplerian disk) is turned into increased rotational kinetic energy, and the rest should be mostly turned into heat. Therefore, the heating rate per unit radius is of order $-g_R\dot{M}(R) \approx -g_R\dot{M}_{\rm tot}$. The cooling rate should be similar, because the timescale of heating and cooling is generally short compared to the timescale of disk evolution and the disk should be in approximate thermal equilibrium.

In Appendix E, we perform a more rigorous analysis of the thermal budget of the disk. It turns out that the intuition above is only partially correct: the energy flux from turbulent perturbations (including spiral waves) can cause a significant amount of radial energy transport and affect the rate of heating (or, more accurately, energy injection). However, in Appendix E we also show that the energy injection due to such a turbulent flux is at most comparable to the energy injection by accretion, and $-g_R\dot{M}_{\rm tot}$ remains a good order-of-magnitude estimate of the disk heating and cooling rates.

We check the validity of this estimate by comparing it with the cooling rate measured from our simulation in Fig. 3.13. The actual cooling rate agrees relatively well with $-g_R\dot{M}_{\rm tot}$ inside the disk.

3.4.2 Relation between cooling rate and disk temperature

Now that we have an estimate of the disk cooling rate, the disk temperature profile can be estimated as well if the relation between cooling and temperature is known. Here we discuss this relation.

Again, we begin with a rough estimate, which ignores the effect of turbulent fluctuations altogether. In

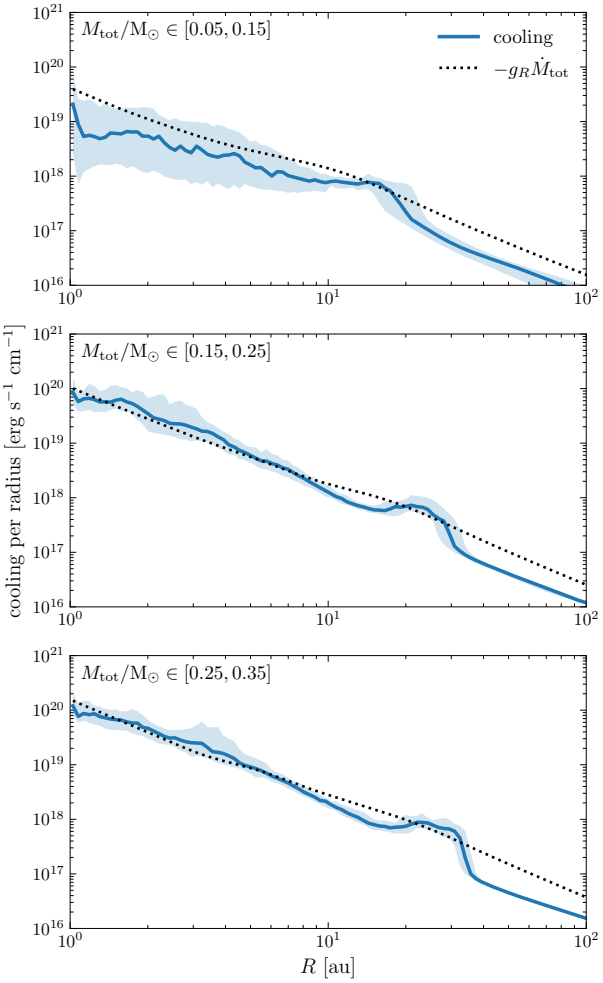

Figure 3.13: Cooling rate per radius at different epochs. Shaded area marks 1σ deviations. Inside the disk, the cooling rate required by thermal equilibrium is consistent with the (order-of-magnitude) analytic estimate $-g_R\dot{M}_{\rm tot}$.

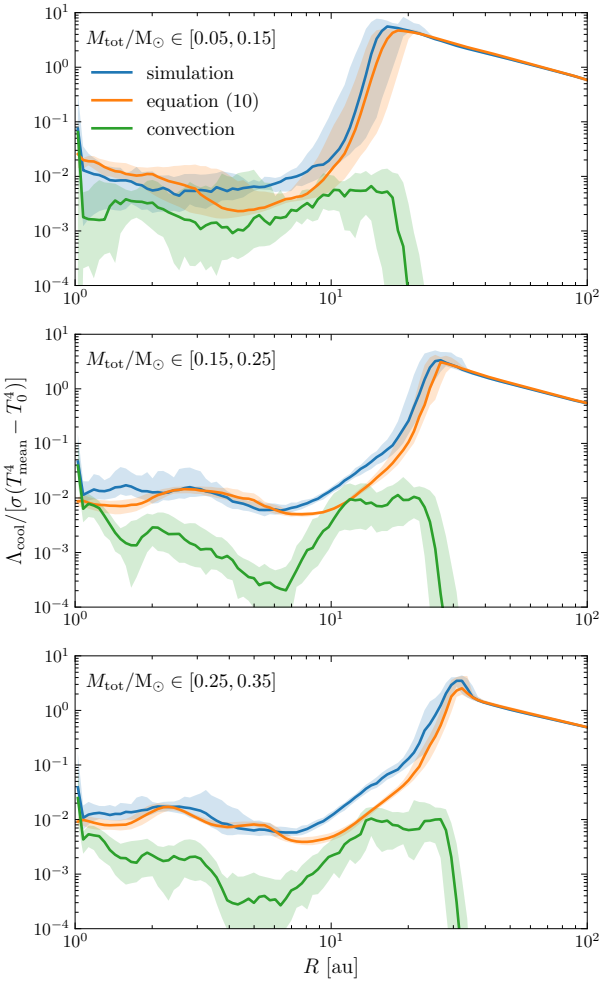

Figure 3.14: Normalized cooling rate per area at a few different epochs. The orange curve corresponds to the estimate in Eq. (3.4), which ignores the effect of turbulence (including spirals). This estimate fits reasonably well to simulation. The green curve measures vertical convection: We measure the vertical convective internal energy flux $\langle \rho u v_z' \rangle$ (where u is the specific internal energy and v_z' is the difference between v_z and its density-weighted azimuthal average) and take twice its maximum in z (at a given R; the factor of 2 accounts for both sides of the disk) as an estimate for the cooling due to vertical convection. Convection is always weak compared to the total cooling.

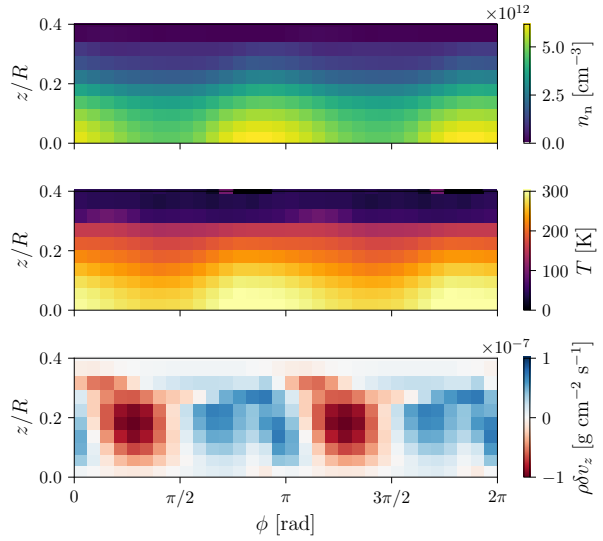

Figure 3.15: Density, temperature, and vertical velocity in the disk at 10 au, when $M_{\rm tot} = 0.3$ M$_\odot$. Both the temperature fluctuations and vertical motions correlates with the density fluctuations; this leads to convective heat transport that enhances cooling. However, as we show in Fig. 3.14 and Section 3.4.3, such an effect is never significant.

this case, the cooling rate (per area) of a disk in thermal equilibrium can be estimated as follows:[4]

$$\Lambda_{\text{no-conv}} \sim \frac{8\tau_{\rm P}}{1 + 0.875\tau_{\rm P}\tau_{\rm R}}\,\sigma(T_{\rm mean}^4 - T_0^4). \tag{3.4}$$

Here $\tau_{\rm P}$ and $\tau_{\rm R}$ are the Planck and Rosseland optical depths at the midplane. Equation (3.4) is exact when $\tau_{\rm P} \ll 1$ or $\tau_{\rm P}, \tau_{\rm R} \gg 1$ and provides a smooth transition between these two limits. We compare this estimate with the actual cooling rate in Fig. 3.14. While there are some small differences (at most by a factor of $\sim$2), this estimate stays very close to the actual cooling rate.

Next we consider how the inclusion of turbulent fluctuations – mainly the spirals – might affect the cooling rate. First, the temperature and optical depth are no longer uniform at a given radius. But we anticipate the effect of this variation to be relatively unimportant, since these fluctuations, and the variation in cooling rate they induce, should at most be of order unity. A second, and potentially more important, effect of the spirals is the vertical convection associated with them (e.g., Riols & Latter, 2018; Béthune et al., 2021).[5] An example of such convective motions is shown in Fig. 3.15. In our 3D simulation, we find that this kind of convection is generally weaker than the total radiative cooling by a factor of a few or more (Fig. 3.14, green

[4]In the literature, $(16/3\tau)\sigma T^4$ is frequently quoted as the cooling rate for an optically thick disk. However, the T in this expression should be treated as the mid-plane temperature; the mean (density-weighted average) temperature is smaller than the mid-plane temperature by a factor of 0.874 (in the limit $\tau \to \infty$), if we assume the disk has constant opacity (Hubeny, 1990, eq. 3.11). The $8/(16/3 \times 0.874^4) = 0.875$ factor in equation (3.4) takes this into account.

[5]This should not be confused with convection due to a vertical entropy gradient; our disk is generally stable against the latter, since for typical disk temperatures (a few 100 K) we have $\gamma \approx 1.4$ and $\kappa_{\rm P} \sim$ const (Rafikov, 2007, eqn. 4).

curve), and thus does not significantly affect the cooling rate. In the next subsection we argue that the same should be true for spiral convection in any Class 0/I disk. Therefore, even in the presence of turbulence and spirals, equation (3.4) should remain a reasonably good estimate for the relation between disk temperature and cooling rate.

3.4.3 Convection driven by spirals

In this subsection we derive an analytic estimate of the convection by spirals and use that to argue that such convection should not dominate cooling in a Class 0/I disk.

First, because there is a strong correlation between vertical velocity and temperature (as both correlate with the spirals), the amount of convective cooling should be proportional to the characteristic amplitude of the velocity perturbations times the amplitude of fluctuations in the internal energy density:

$$\Lambda_{\rm conv} \sim \frac{1}{2} U (\delta v_z / H)(\delta u / u). \tag{3.5}$$

Here U is the vertically-integrated internal energy, δv_z is the (density-weighted) rms velocity perturbation, H is the scale height (measured by dividing the density-weighted rms sound speed by Ω), and δu is the (density-weighted) rms specific internal energy perturbation. The factor of $1/2$ is chosen to fit the simulation result (see Appendix F).

Next we estimate δv_z and $\delta u/u$. The vertical motion is driven mainly by the pressure gradient associated with temperature perturbations (note that spiral waves in an isothermal disk do not generate vertical motion; see Riols & Latter 2018); increased temperature leads to vertical expansion of the disk by increasing the scale height needed for hydrostatic equilibrium $H_{\rm eq}$. The frequency at which the temperature is perturbed by the spirals is (in the frame corotating with the gas)

$$\Omega_{\rm s} \equiv |\Omega_{\rm p} - \Omega|. \tag{3.6}$$

Meanwhile, the characteristic rate at which the disk returns to vertical hydrostatic equilibrium can be estimated using the vertical sound crossing rate, $c_{\rm s}/H \sim \Omega$. Comparing Ω with the "forcing" frequency $\Omega_{\rm s}$ gives two different regimes. When $\Omega \gtrsim \Omega_{\rm s}$, the disk can adjust itself into vertical hydrostatic equilibrium quickly, and the amplitude of vertical oscillation is just $\delta H_{\rm eq} \sim H(\delta u/u)$. This gives

$$\delta v_z \sim \Omega_{\rm s} \delta H_{\rm eq} \sim c_s (\Omega_{\rm s}/\Omega)(\delta u/u) \quad (\text{for } \Omega \gtrsim \Omega_{\rm s}). \tag{3.7}$$

When $\Omega \lesssim \Omega_{\rm s}$, the disk cannot adjust to vertical hydrostatic equilibrium, and the amplitude of the vertical acceleration remains $a_z \sim g_z(\delta u/u) \sim \Omega^2 H(\delta u/u)$. The typical vertical velocity is then

$$\delta v_z \sim \Omega_{\rm s}^{-1} a_z \sim c_s(\Omega/\Omega_{\rm s})(\delta u/u) \quad (\text{for } \Omega \lesssim \Omega_{\rm s}). \tag{3.8}$$

Combining equations (3.7) and (3.8), we get

$$\delta v_z \sim c_{\rm s} \min(\Omega_{\rm s}/\Omega, \Omega/\Omega_{\rm s})(\delta u/u). \tag{3.9}$$

Finally, we estimate $\delta u/u$ by assuming that all of the heating happens near the spiral shock, while cooling is more uniformly distributed in ϕ. Gas in the disk encounters spiral shocks at a frequency of $\Omega_{\rm s}/\pi$ (for $m = 2$ spirals), so the energy injection (per disk area) at each encounter is $\Lambda_{\rm heat}(\pi/\Omega_{\rm s})$, and the jump in internal energy across the shock is $\delta u \sim \Lambda_{\rm heat}(\pi/\Omega_{\rm s})\Sigma^{-1}$. We therefore expect

$$\delta u/u \sim \min\left(\frac{1}{2}, \frac{\Lambda_{\rm heat}(\pi/\Omega_{\rm s})}{U}\right). \tag{3.10}$$

Here we cap $\delta u/u$ to 1/2, corresponding to the case where under-dense material contains nearly zero specific internal energy.

Altogether then, we have

$$\Lambda_{\rm conv} \sim \frac{1}{2} U \Omega_{\rm s} \min(1, \Omega^2/\Omega_{\rm s}^2) \min\left(\frac{1}{2}, \frac{\pi \Lambda_{\rm heat}}{U \Omega_{\rm s}}\right)^2. \tag{3.11}$$

This gives $\Lambda_{\rm conv} \lesssim (\pi/4)\Lambda_{\rm heat}$, with the equality being achieved only when $\Omega_{\rm s} \sim \Omega$ (i.e., at radii $\lesssim$ the radius of spiral excitation) and $U\Omega \sim \Lambda_{\rm heat}$ (i.e., the heating/cooling timescale is comparable to orbital timescale). In Appendix F we compare the above analytic estimates with the convective energy transport measured from the simulation, and show that these estimates model the simulation measurement accurately. The relation between vertical velocity and heating rate proposed above is also consistent with the result observed in the higher resolution simulation (of an initially gravitationally unstable disk) from Béthune et al. (2021), which shows $\delta v_z \sim 0.1 c_{\rm s}$ for $\Lambda_{\rm heat} \sim 0.1U$ and $\Omega \sim \Omega_{\rm s}$.

In summary, in Class 0/I disks, spiral convection should not produce enough vertical-energy transport to account for most of the cooling.

3.5 Magnetic field evolution

3.5.1 Magnetic decoupling and field strength in the disk

In the ideal-MHD limit, the magnetic field is frozen into and advected with the gas, and the field strength is simply proportional to the column density. In reality, however, the presence of non-ideal magnetic diffusion (Ohmic dissipation and ambipolar diffusion for our simulation) allows the field to become decoupled from the gas and left behind as the bulk-neutral fluid flows inwards. Neglecting the Hall effect, this decoupling is described by the non-ideal induction equation

$$\partial_t \boldsymbol{B} = \boldsymbol{\nabla} \times \left(\boldsymbol{v} \times \boldsymbol{B} - \tfrac{4\pi}{c} \eta_{\rm O} \boldsymbol{j} - \tfrac{4\pi}{c} \eta_{\rm A} \boldsymbol{j}_\perp \right), \tag{3.12}$$

where $\boldsymbol{j} = (c/4\pi)\boldsymbol{\nabla} \times \boldsymbol{B}$ is the current density and $\eta_{\rm O}$ and $\eta_{\rm A}$ are the Ohmic and ambipolar diffusivities. For simplicity, we ignore the contribution from the field-parallel current $\boldsymbol{j}_\parallel$ to the Ohmic term, which is a good approximation because $\boldsymbol{j}$ is nearly perpendicular ($\perp$) to $\boldsymbol{B}$ everywhere and $\eta_{\rm O} \ll \eta_{\rm A}$ in most of the protostellar core. Under this approximation, equation (3.12) may be rewritten as

$$\partial_t \boldsymbol{B} = \boldsymbol{\nabla} \times [(\boldsymbol{v} + \boldsymbol{v}_{\rm d}) \times \boldsymbol{B}], \tag{3.13}$$

where the non-ideal drift velocity $\boldsymbol{v}_{\rm d}$ is defined by

$$\boldsymbol{v}_{\rm d} \equiv -(\eta_{\rm O} + \eta_{\rm A}) \frac{\boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{B})}{B^2}. \tag{3.14}$$

The combination $\boldsymbol{v} + \boldsymbol{v}_{\rm d}$ may be interpreted as the velocity of the field lines. Comparing the drift velocity $\boldsymbol{v}_{\rm d}$ with the gas velocity $\boldsymbol{v}$ therefore allows us to quantify the decoupling between the field and the gas. Note that the second part of equation (3.14) describes how non-uniform the field is; having the same form as the Lorentz force, it goes to zero when the field is straight and uniform, and increases as the field strength becomes more non-uniform or the field lines become more pinched.

Different regimes of field decoupling

In Fig. 3.16 we show the radial drift velocity at the midplane and compare it with the radial velocity of the (predominantly neutral) gas. We divide the domain into three different regimes:

Outer pseudodisk: well-coupled field. At large radii, $|v_{{\rm d},R}| \ll |v_R|$ and the field is well-coupled to the gas. The flux evolution is determined by (near) flux freezing, with a residual amount of ambipolar diffusion.

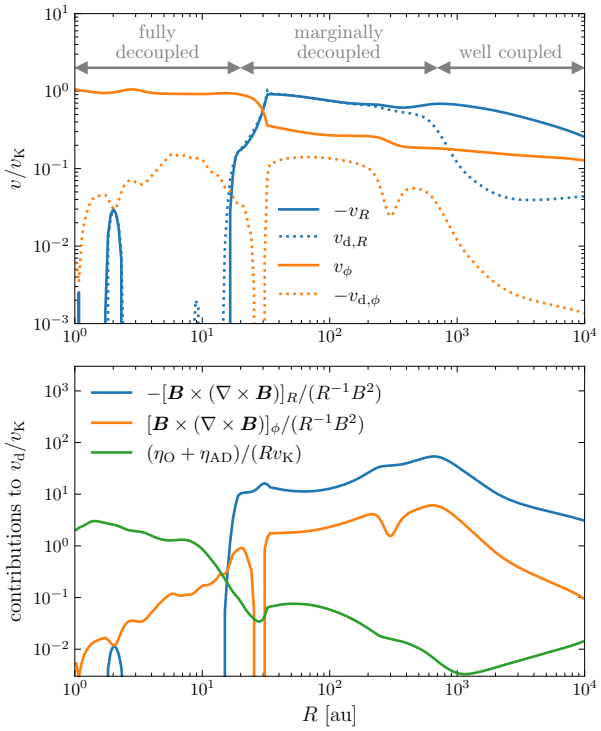

Figure 3.16: Top panel: Comparison between gas velocity v (solid) and non-ideal drift velocity $v_{\rm d}$ (dotted) measured in the midplane in the radial (blue) and azimuthal (orange) directions, all azimuthally averaged (weighted by midplane B_z) and normalized by $v_{\rm K} \equiv \sqrt{-g_R R}$ (top panel). In the radial direction, the magnetic field is well coupled in the outer pseudodisk, marginally decoupled in the inner pseudodisk, and fully decoupled in the disk (as marked in the figure); in the azimuthal direction, the magnetic field remains relatively well coupled (i.e., azimuthal field velocity is similar to azimuthal gas velocity) in the whole domain, in part because the field curvature in the azimuthal direction cannot produce a sufficiently large azimuthal drift velocity (cf. Section 3.5.2). Bottom panel: Contributions to the midplane $v_{\rm d}/v_{\rm K}$ from magnetic-field non-uniformity (blue, orange) and from non-ideal magnetic diffusivity (green), all azimuthally averaged. Both panels are evaluated when $M_{\rm tot} \approx 0.4$ M$_\odot$, with the behaviour being qualitatively similar at other epochs.

Inner pseudodisk: marginally decoupled field. In the inner part of the pseudodisk, as the field becomes more pinched and the ambipolar diffusivity increases, $v_{\mathrm{d},R}$ becomes comparable to $-v_R$ and the field radially decouples from the gas. The decoupling of the magnetic field tends to straighten out the field lines (reducing field non-uniformity), which in turn reduces $v_{\mathrm{d},R}$ and makes the field better coupled to the gas. Therefore, the system tends towards a self-regulated, marginally decoupled state in which $v_{\mathrm{d},R}$ is kept comparable to $-v_R$ by regulating field non-uniformity. This idea is consistent with the profiles of drift velocity and field non-uniformity shown in Fig. 3.16.

The decoupling also slows the inward advection of the magnetic field, leading to a local enhancement in magnetic-field strength, which is visible in the B_z panel of Fig. 3.3. Since the magnetic field remains relatively well coupled in the azimuthal direction, this increased field strength can help explain the increased braking in the inner pseudodisk seen in Fig. 3.10 (also see the discussion on magnetic braking in Section 3.5.2).

We note in passing that this local increase in magnetic-field strength due to decoupling is physically similar to the "magnetic wall" suggested in Li & McKee (1996), although in our case the wall is caused by ambipolar diffusion rather than Ohmic dissipation (as in Contopoulos et al. 1998, Ciolek & Königl 1998, and Tassis & Mouschovias 2005) and there is no clear evidence of magnetic interchange instability occurring behind the wall. In our simulation, the maximum growth rate of the interchange (Lubow & Spruit, 1995), which is proportional to the gradient of the local mass-to-flux ratio, is always smaller than the local infall rate (which is close to free-fall). Meanwhile, significant asymmetries due to interchange have been observed in some simulations of protostar formation (e.g., Krasnopolsky et al., 2012; Zhao et al., 2018a; Machida & Basu, 2019), and so its occurrence (i.e., whether it can occur on a timescale shorter than both infall and magnetic diffusion) likely depends upon the details of the simulation setup and various physical conditions.

Protostellar disk: fully decoupled field. In the protostellar disk, the even higher density further increases the non-ideal diffusion; more importantly, the typical radial velocity is significantly lower compared to the pseudodisk due to increased column density. As a result, the field is fully decoupled radially and becomes nearly uniform and straight.

We comment that the marginally decoupled regime and the fully decoupled regime are essentially controlled by the same physics, in which the magnetic field self-regulates its non-uniformity (for a given magnetic diffusivity) to give a drift velocity that approximately cancels the gas velocity. The main difference between these two regimes is just whether the field non-uniformity required in this self-regulated state is negligibly small.

Where does decoupling begin?

One interesting trend we observe is that the location where magnetic decoupling begins (i.e., the boundary between the well-coupled regime and the marginally decoupled regime) shifts outwards in time. This trend causes the expansion of the region in the inner pseudodisk with enhanced B_z (Fig. 3.3) and increased magnetic braking (Fig. 3.10), thereby contributing to the increased importance of pseudodisk magnetic braking at later times. Here we discuss the origin of this trend.

We can obtain a rough estimate for the location of decoupling by considering the curvature of the pseudodisk magnetic field in the ideal-MHD limit. Decoupling happens approximately when the drift velocity corresponding to this ideal-MHD field non-uniformity becomes comparable to the infall velocity. From the perspective of the pseudodisk, the central protostar-disk system is effectively a point mass. We therefore divide the pseudodisk into an outer, "pre-stellar" region, where gravity is dominated by the self-gravity of the pseudodisk [$M(<R) \gg M_{\rm tot}$], and an inner, "post-stellar" region, where gravity is dominated by the point mass [$M(<R) \approx M_{\rm tot}$].[6] The pre-stellar region acquires a near-self-similar profile during the pre-stellar collapse, one which holds as well during the Class 0/I phase, with (approximately) $\Sigma \propto R^{-1}$ and $M(<R) \propto R$. The post-stellar region also converges to a near-self-similar profile, but with $\Sigma \propto R^{-1/2}$ (e.g., Contopoulos et al., 1998; Dapp et al., 2012). The boundary between these two regions, $R_{\rm post}$, occurs roughly at $M(<R) \sim 2M_{\rm tot}$ and expands outwards as $M_{\rm tot}$ increases.

We use these profiles to obtain the scaling of the field non-uniformity $-[\boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{B})]_R/B^2$ in the pseudodisk midplane (assuming ideal-MHD). Because the pseudodisk is geometrically thin, this non-uniformity is dominated by the field-line curvature, and so

$$-\frac{[\boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{B})]_R}{B^2} \approx \frac{1}{B_z}\frac{\partial B_R}{\partial z} \sim \frac{B_R^{\rm d}}{B_z H}. \tag{3.15}$$

Here $B_R^{\rm d}$ is the strength of the radial field at the pseudodisk surface; the pseudodisk scale height H may be estimated as

$$H \sim \frac{c_{\rm s}^2}{g_z}. \tag{3.16}$$

The sound speed $c_{\rm s} \approx c_{\rm s0}$, because disk is approximately isothermal in the regions of interest here. The vertical component of the gravitational acceleration g_z is $\sim 2\pi {\rm G}\Sigma$ in the pre-stellar region, and $\sim({\rm G}M_{\rm tot}/R^2)(H/R)$ in the post-stellar region. Meanwhile, previous semi-analytic calculations (e.g., Contopoulos et al., 1998;

[6] The post-stellar region is often called the "expansion wave" region in the literature, following Shu (1977). Here we avoid this terminology because in our case the outer (pre-stellar) envelope is not a near-hydrostatic singular isothermal sphere (as in Shu, 1977), but rather a dynamically contracting pseudodisk. Moreover, the boundary between the pre-stellar region and the post-stellar region ("expansion wave") is not a physical, outward propagating wave/shock but a transition between two self-similar profiles.

Galli et al., 2006) predict, in the flux-freezing limit, $B_z, B_R^{\rm d} \propto R^{-1}$ in the pre-stellar region and $B_z \propto R^{-1/2}$, $B_R^{\rm d} \propto R^{-2}$ in the post-stellar region, the latter corresponding to a split monopole. Therefore, at a given time, the field curvature should scale as R^{-1} and $R^{-5/2}$ for the pre-stellar and post-stellar regions, respectively. Since the field curvature must be continuous around $R_{\rm post}$ and the pre-stellar profile remains approximately constant during the Class 0/I stage (because the local dynamical timescale is $\gtrsim$ the timescale after point mass formation), we find

$$\frac{B_R^{\rm d}}{B_z H}\ (\text{ideal-MHD}) \propto \begin{cases} R^{-1} & (R \gtrsim R_{\rm post}) \\ R^{-5/2} R_{\rm post}^{3/2} & (R \lesssim R_{\rm post}) \end{cases} . \tag{3.17}$$

This profile suggests that, at a given location R, the field curvature remains constant until $R_{\rm post}$ reaches this radius, and quickly increases ($\propto R_{\rm post}^{3/2}$) afterwards (until decoupling sets in). The infall velocity also increases after $R_{\rm post}$ exceeds R, but it increases more slowly, with a scaling $\propto R_{\rm post}^{1/2}$ at a given radius. Therefore, the growth of $R_{\rm post}$ (driven by accretion onto the point mass $M_{\rm tot}$) promotes decoupling by increasing field curvature and thereby lowering the magnetic diffusivity required for decoupling. This argument is similar to that made in appendix C of Ciolek & Königl (1998) (see also Contopoulos et al. 1998 and Krasnopolsky & Königl 2002). While the increase in field curvature is the main cause of decoupling in our pseudodisk, we note that in general magnetic decoupling may also be driven mainly by increased diffusivity (e.g., Desch & Mouschovias, 2001) and/or decreased radial velocity (e.g., our protostellar disk).

An example of decoupling driven by field curvature increase in the post-stellar region is shown in the bottom panel of Fig. 3.16, with the transition between the pre- and post-stellar regions occurring at $R_{\rm post} \sim 2000$ au and decoupling starting around ∼800 au, after field curvature has increased by a factor of a few in the post-stellar region. A similar correlation between the location of decoupling and $R_{\rm post}$ is observed throughout our simulation, suggesting that the expansion of the marginally decoupled region (and the increased importance of braking at later times – see Section 3.5.2) is largely due to the expansion of the post-stellar region.

What determines the field strength in the protostellar disk?

Now we discuss how to obtain a rough estimate of the protostellar disk magnetic-field strength using the field strength in the pseudodisk, which will be used for our semi-analytic disk model in Chapter 4.

One complicating factor is that the boundary between the marginally decoupled regime and the fully decoupled regime does not lie exactly at the disk boundary (defined as the location where the kinetic energy becomes dominated by its azimuthal component), but is rather slightly inside the disk. For example, in

Fig. 3.16 the disk boundary is at $R_{\rm d} \sim 30$ au while the boundary of the fully decoupled regime is at $R_{\rm dec} \sim 20$ au. This is because the increase in magnetic diffusivity and column density (which produces the decrease in typical radial velocity) happens gradually across the transition region between the dense, gravitationally self-regulated region and the pseudodisk.

For $R < R_{\rm dec}$, the magnetic-field strength should be nearly constant. Between $R_{\rm dec}$ and $R_{\rm d}$, the magnetic field is non-uniform and slightly pinched to provide the finite drift velocity necessary for canceling the gas velocity. This leads to the bump in field strength at the disk edge shown in Fig. 3.3. The transition region between $R_{\rm dec}$ and $R_{\rm d}$ is generally narrow, with $R_{\rm dec}/R_{\rm d} \sim 1$; the difference in magnetic-field strength between these two locations should be of order unity. Therefore, we argue that the field strength in the pseudodisk around $R_{\rm d}$ can serve as a reasonable estimate of the field strength inside the disk, with an order-unity uncertainty due to the marginally decoupled region that occurs between $R_{\rm dec}$ and $R_{\rm d}$.

3.5.2 Magnetic braking in the disk

Another important aspect of magnetic-field evolution is the generation of an azimuthal field component and the resulting magnetic braking. We showed earlier that angular-momentum removal from the protostar-disk system by magnetic braking is weak compared to angular-momentum injection by accretion. In this subsection we aim to understand what determines the strength of magnetic braking in the disk. This could help us determine in the future whether disk magnetic braking should be similarly unimportant in other systems.

Inside the disk, while the magnetic field is decoupled from the gas in the radial direction, it is relatively well-coupled to the gas in the azimuthal direction, with a non-ideal drift velocity significantly smaller than the azimuthal velocity (Fig. 3.16). Consider a field line threading the disk; the rotation rate of the field line near the disk should then be much greater than its rotation rate at $z \rightarrow \infty$. Traditionally, the winding of field lines (and the associated angular momentum transport) due to this differential rotation between the disk and infinity is often considered as the main source of magnetic braking. In this case, the strength of the braking may be estimated using the steady-state model presented in Basu & Mouschovias (1994), in which one effectively assumes that the region of a flux tube casually connected to the disk by torsional Alfvén waves is brought into corotation with the disk. This kind of model may not apply to our disk, however, given that the gas dynamics occurring along the field lines – a combination of outflow launching near the disk and pre-stellar infall further out – is quite different than is assumed in this model, and the magnetic diffusivity in the outflow cavity and envelope could also have a nontrivial impact on how the field lines wind up. In fact, the method from Basu & Mouschovias (1994) provides an estimate for the magnetic-braking

rate that is 1–2 orders of magnitude greater than what we measure in the simulation.

We propose a different model to explain the strength of magnetic braking we observe, which assumes that the main origin of disk magnetic braking is not the differential rotation between the disk and infinity, but rather the (vertical) differential rotation *within* the disk. Under this assumption, the disk magnetic braking rate can be estimated as follows.

We observe the magnetic field to be dynamically unimportant inside the disk, as magnetic energy remains well below the kinetic and thermal energy. Therefore, the rotation of the gas is approximately the same as that in a hydrodynamic disk, and so along a field line at R (which is approximately vertical inside the disk) the amount of differential rotation inside the disk (due to the vertical variation of radial gravity and pressure gradient) is approximately

$$\Delta\Omega \sim (z_{\rm d}/R)^2\Omega. \tag{3.18}$$

Here Ω is the rotation rate of disk at R, and $z_{\rm d}$ is the half-thickness of the disk.[7]

In steady state, the field must rotate uniformly along each field line, and so the vertical differential rotation of the gas in the disk needs to be canceled by the vertical variation of the non-ideal drift velocity. The amplitude of the drift velocity then has to be at least

$$v_{{\rm d},\phi} \sim R\Delta\Omega. \tag{3.19}$$

Meanwhile, using equation (3.14), the azimuthal drift velocity is approximately

$$v_{{\rm d},\phi} \sim -\frac{B_\phi^{\rm d}}{B_{\rm z} z_{\rm d}}\eta, \tag{3.20}$$

where $B_\phi^{\rm d}$ is B_ϕ evaluated at disk surface and η is the typical $\eta_{\rm O}+\eta_{\rm AD}$ inside the disk, which can be estimated using $\eta_{\rm O}+\eta_{\rm AD}$ at the disk midplane. Equations (3.19) and (3.20) provide an estimate of $B_\phi^{\rm d}$. The resulting braking rate per unit area is

$$F_{L,{\rm brake}} = -2RB_\phi^{\rm d}B_z \sim 2B_z^2\eta^{-1}\Omega z_{\rm d}^3. \tag{3.21}$$

We verify this physical picture in Fig. 3.17, which shows the rotation rate of the field along a field line (green), and its contribution from the gas velocity (blue) and the drift velocity (orange). To compute the rotation rate of the field, we first find the horizontal component of the field velocity, $\boldsymbol{v}_{\rm f}^{\rm h}$, which is defined such that $v_{{\rm f},z}^{\rm h} = 0$ and $\boldsymbol{v}_{\rm f} - \boldsymbol{v}_{\rm f}^{\rm h}$ (where $\boldsymbol{v}_{\rm f} \equiv \boldsymbol{v} + \boldsymbol{v}_{\rm d}$ is the field velocity) is parallel to the magnetic field

[7] $z_{\rm d}$ is defined as the height where magnetic energy exceeds the sum of kinetic and thermal energy, and should be comparable to the scale height $H = c_{\rm s}/\Omega$. However, since the braking rate estimate has a steep dependence on $z_{\rm d}$, we do not use H to approximate $z_{\rm d}$ here. For our simulation, $z_{\rm d}$ is typically $\sim 1.5H$.

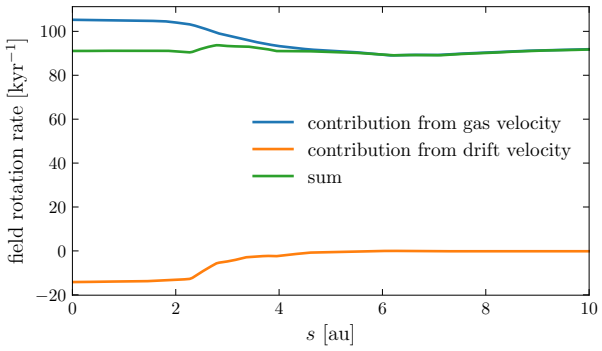

Figure 3.17: Azimuthally averaged field-rotation rate (green; see definition in Section 3.5.2) and its contributions from gas velocity (blue) and drift velocity (orange), evaluated when $M_{\rm tot} = 0.4\ {\rm M}_\odot$ along a field line going through the midplane at 10 au. The abscissa is distance along the field line ($s = 0$ at midplane). The field is in nearly uniform rotation, and the vertical differential rotation of the gas within the disk ($s \lesssim 3$ au) is countered by the non-ideal drift velocity.

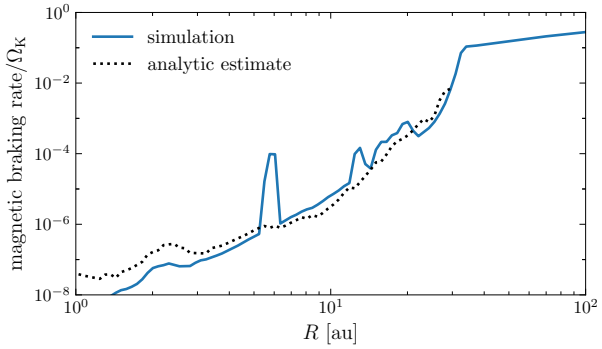

Figure 3.18: Comparison between the magnetic-braking rate when $M_{\rm tot} = 0.4\ {\rm M}_\odot$ (same definition as in Fig. 3.10) and the analytic estimate (3.21). The two show good agreement inside the disk. The occasional peaks in the measured braking rate (such as the one at $\sim$6 au) are random fluctuations likely associated with turbulent motions and spiral waves in the disk.

(so $\boldsymbol{v}_{\rm f} \times \boldsymbol{B} = \boldsymbol{v}_{\rm f}^{\rm h} \times \boldsymbol{B}$). The field rotation rate can then be defined as $\Omega_{\rm f} \equiv v_{{\rm f},\phi}^{\rm h}/R$. The contribution to $\Omega_{\rm f}$ from the gas velocity and the drift velocity can be computed similarly by replacing $\boldsymbol{v}_{\rm f}$ with $\boldsymbol{v}$ or $\boldsymbol{v}_{\rm d}$. We comment that inside the disk, both $\boldsymbol{v}$ and $\boldsymbol{v}_{\rm d}$ are nearly horizontal, therefore we do not have to distinguish the velocities from their horizontal components in the earlier discussion. Fig. 3.17 shows that the field is indeed in nearly uniform rotation, and the differential gas rotation inside the disk is offset by non-ideal drift velocity.

We also compare the estimated braking rate using equation (3.21) with the actual braking rate inside the disk in Fig. 3.18. Our model predicts the braking rate relatively well, suggesting that it is likely the main mechanism determining the strength of disk magnetic braking in our simulation.

Chapter 4

A semi-analytic model of disk evolution

This chapter is adapted from material published in:

Xu, W. and Kunz, M. W. (2021b). Formation and evolution of protostellar accretion discs – II. From 3D simulation to a simple semi-analytic model of Class 0/I discs. MNRAS, 508(2):2142–2168.

In this chapter we formulate a 1D semi-analytic model of a gravitationally self-regulated protostellar disk with the physical insights acquired from the previous chapter (Section 4.1). We then discuss how this model can be used together with simulations that do not directly resolve the disk to predict the long-term evolution (throughout the main accretion phase) of the disk (Section 4.2), and test this method against our 3D simulation (Section 4.3). We note that this is not the only application of such a model; in Chapter 5, we discuss how this model can be used to interpret dust continuum observation after some minor modifications.

4.1 A semi-analytic 1D disk model

4.1.1 Basic assumptions

Consider a gravitationally self-regulated disk similar to the one found in our simulation, with most of the disk marginally unstable. Let $R_{\rm GI}$ be the outer boundary of the gravitationally unstable region. While we expect the disk size $R_{\rm d}$ to be $\sim R_{\rm GI}$ from gravitational self-regulation (Sections 3.3.2 and 3.3.3), there is still some order-unity difference between $R_{\rm GI}$ and $R_{\rm d}$, as the region between $R_{\rm GI}$ and $R_{\rm d}$ serves as a transition region for spiral waves excited within $R_{\rm GI}$ to disperse (Fig. 3.11). For simplicity, we assume

$$R_{\rm d} = f_{\rm d} R_{\rm GI} \tag{4.1}$$

with $f_{\rm d}$ being a constant of order unity. To match the typical value of $R_{\rm d}/R_{\rm GI}$ realized in our 3D simulation, we set $f_{\rm d} = 1.5$. We also assume that most of disk mass and angular momentum are contained in $R \leq R_{\rm GI}$ (because the column density falls off rapidly outside $R_{\rm GI}$) and that the Toomre Q parameter is approximately constant for $R \leq R_{\rm GI}$, i.e.,

$$Q(R \leq R_{\rm GI}) \approx Q_0. \tag{4.2}$$

We set $Q_0 = 1.5$ by default. $f_{\rm d}$ and Q_0 will be the only free parameters of our model.

4.1.2 Local model for temperature and column density

Next we model the mean temperature $T_{\rm mean}$ and column density Σ at a given location inside the gravitationally self-regulated region ($R \leq R_{\rm GI}$). We assume the disk to be near Keplerian, so that

$$Q \approx Q_{\rm K} = Q_0. \tag{4.3}$$

(Recall that $Q_{\rm K}$ is Toomre's Q evaluated with $\Omega = \Omega_{\rm K} \equiv \sqrt{-g_R/R}$.) We estimate g_R as $g_R \approx -{\rm G}M(<R)/R^2$ where $M(<R)$ is the total mass enclosed within R, including the protostar. Meanwhile, we showed earlier that thermal equilibrium in the disk requires a cooling rate of order $-g_R\dot{M}_{\rm tot}$ with $\dot{M}_{\rm tot}$ being the accretion rate of the protostar-disk system (Section 3.4.1)[1], and that the relation between the cooling rate and mean temperature is approximated well by equation (3.4) (Section 3.4.2). Therefore, we can estimate $T_{\rm mean}$ by requiring

$$2\pi R \frac{8\tau_{\rm P}}{1+0.875\tau_{\rm P}\tau_{\rm R}} \sigma(T_{\rm mean}^4 - T_0^4) = -g_R\dot{M}_{\rm tot}. \tag{4.4}$$

Here, the midplane optical depths $\tau_{\rm P,R}$ can be estimated using $\tau_{\rm P,R} \approx \Sigma\,\kappa_{\rm P,R}(T_{\rm mean})/2$, where $\kappa_{\rm P,R}(T)$ is the Planck/Rosseland mean opacity evaluated at temperature T and $T_0 = 10$ K is the ambient temperature. Note that the steep dependence of the cooling rate on $T_{\rm mean}$ ensures that we get a reasonable temperature estimate even though $-g_R\dot{M}_{\rm tot}$ is only an order-of-magnitude estimate of the cooling rate (cf. Fig. 3.14). Using the two constraints above, we can then solve for Σ and $T_{\rm mean}$ if $M(<R)$ and $\dot{M}_{\rm tot}$ are known.

4.1.3 Estimating disk size and disk profile

The local model above cannot be directly used to predict the disk profile because $M(<R)$ depends on $\Sigma(R)$ when the disk-to-star mass ratio is finite. Here we describe how $\Sigma(R)$, $M(<R)$ and $R_{\rm GI}$ can be calculated using the total mass ($M_{\rm tot}$), angular momentum ($L_{\rm tot}$), and accretion rate ($\dot{M}_{\rm tot}$) of the protostar-disk system.

For a given $\Sigma(R)$ and $M(<R)$, we can solve for $R_{\rm GI}$ by requiring that

$$\int_0^{R_{\rm GI}} 2\pi R{\rm d}R\ \Sigma(R)R^2\Omega_{\rm K} = L_{\rm tot}. \tag{4.5}$$

Meanwhile, for a given $\Sigma(R)$ and $R_{\rm GI}$, we can also obtain $M(<R)$ by solving

$$\frac{\rm d}{{\rm d}R}M(<R) = 2\pi R\Sigma(R) \tag{4.6}$$

with the boundary condition $M(<R_{\rm GI}) = M_{\rm tot}$.

To solve for $\Sigma(R)$, $M(<R)$ and $R_{\rm GI}$, we perform the following iteration:

1. Initialize $M(<R)$ to $M_{\rm tot}$;
2. Calculate $\Sigma(R)$ (and also $T_{\rm mean}$) using $M(<R)$ and equations (4.3) and (4.4), and solve $R_{\rm GI}$ from equation (4.5);

[1]The same result (up to an order-unity coefficient) can be derived using the relation between cooling rate and effective α presented in Gammie (2001).

3. Update $M(<R)$ using equation (4.6);

4. Repeat (2) and (3) until the disk profile converges.

The disk size is simply $R_{\rm d} = f_{\rm d} R_{\rm GI}$.

Finally, the field in most part of the disk should be approximately uniform due to strong diffusion. The field strength can be estimated using the field strength at the inner edge of the pseudodisk $B_z^{\rm pd}$ (Section 3.5.1), giving

$$B_z \sim B_z^{\rm pd}(R_{\rm d}). \tag{4.7}$$

4.1.4 Summary and discussion

In summary, we can estimate the temperature $T_{\rm mean}(R)$, column density $\Sigma(R)$, and disk size $R_{\rm d}$ if the total mass and angular momentum of the protostar-disk system, $M_{\rm tot}$ and $L_{\rm tot}$, and the accretion rate, $\dot{M}_{\rm tot}$, are known. We can further estimate the (approximately constant) field strength B_z inside the disk if the pseudodisk magnetic-field profile $B_z^{\rm pd}(R)$ is also known. This model contains only two free parameters, Q_0 and $f_{\rm d}$, both of which should be order unity constants; we choose $Q_0 = 1.5$ and $f_{\rm d} = 1.5$ to match the typical Q and $R_{\rm d}/R_{\rm GI}$ realized in the 3D simulation.

For completeness, we also highlight the applicability of the above 1D semi-analytic model. The main physical assumptions of the model are that (i) the magnetic field is radially decoupled from the gas and remains dynamically unimportant in the disk (which also implies that accretion needs to be regulated by GI), and (ii) the accretion rate is approximately constant throughout the whole disk. These assumptions are likely valid for typical Class 0/I systems, as our 3D simulation suggests. However, after the dispersal of the envelope and pseudodisk (Class II or later), these assumptions are probably no longer valid due to the lack of fast accretion driven by the infalling pseudodisk. Additionally, these assumptions would become invalid at very small radii where the temperature is so high ($\gtrsim$2000 K) that thermal ionization becomes important and the field becomes well-coupled to the gas again. (Thermal ionization and gas opacity are not currently included in our simulation.) Our 1D model should therefore not be used to predict disk properties in this region. We note that, according to our 1D+pseudodisk model (which we present in the next subsection), the radius where the temperature reaches $\sim$2000 K is always $\lesssim$2 au. Due to the small size of this region, we do not expect the angular momentum contained within this radius or the angular-momentum transport and removal (by magnetorotational instability, magnetic braking, and/or outflow launching) that may occur within this radius to impact significantly our prediction for the disk profile farther out.

We comment that semi-analytic models for (partially or fully) gravitationally self-regulated disks have been commonly adopted in studies of protostellar or protoplanetary disk evolution (e.g., Lin & Pringle, 1987;

Krumholz & Burkert, 2010; Rice et al., 2010; Zhu et al., 2010; Kimura et al., 2021). The main contribution of our work is not developing this particular model, but justifying its assumptions in the context of magnetized Class 0/I disks that form self-consistently in 3D simulations of protostellar core collapse (Sections 3.3–3.5) and showing how such a model can be coupled to a pseudodisk simulation to give reliable predictions of disk evolution (Sections 4.2 and 4.3).

4.2 1D+pseudodisk model: evolving the 1D model with a pseudodisk simulation

In the previous section, we discussed how to estimate the disk profile using a 1D model when $M_{\rm tot}$, $L_{\rm tot}$, $\dot{M}_{\rm tot}$ and the magnetic field profile in the pseudodisk are known. Obtaining these inputs for the 1D model requires knowledge of pseudodisk evolution. Here we show that one should be able to predict accurately pseudodisk evolution using computationally inexpensive, large-inner-boundary simulations that do not directly resolve the disk (Section 4.2.1) and discuss how to predict disk evolution by coupling the 1D model to such a pseudodisk simulation (Section 4.2.2).

4.2.1 Hierarchical evolution and simulating the pseudodisk

Inspired by the observation in our exploratory work (XK21a) that 3D and (axisymmetric) 2D simulations produce nearly identical pseudodisk properties even though the latter does not produce correct disk evolution, we conjecture that the evolution of our protostellar system is largely hierarchical, meaning that whatever is happening in the protostellar disk has little impact on the evolution at larger scales (e.g., in the pseudodisk).

One can argue that such hierarchical evolution is generally expected for young protostellar systems as follows. First, the inner pseudodisk is falling in at near free-fall, with a velocity that generally exceeds both the Alfvén speed and the sound speed in the pseudodisk, making it impossible for perturbations to travel outward from the disk into the pseudodisk. Therefore, if the disk were to affect the pseudodisk, it would have to do so indirectly by affecting the magnetic field above the pseudodisk (the gas there is dynamically unimportant). Now consider the magnetic field above the pseudodisk; the field there is nearly force-free, and the disk affects this field mainly by setting the inner boundary condition of this force-free region through the magnetic field in the flux tube threading the disk. The importance of this inner boundary condition can be characterized roughly by the ratio $\Phi_B(R_{\rm d})/\Phi_B(R)$, where Φ_B is the magnetic flux through a circle with radius R at the midplane; if $\Phi_B(R_{\rm d})/\Phi_B(R) \ll 1$, this inner boundary condition should not be important at R. In the pseudodisk, the slope of the magnetic field B_z is shallower than $\propto R^{-2}$ (cf. Fig. 3.3); therefore

$\Phi_B(R)$ is dominated by the contribution around R, giving a small $\Phi_B(R_{\rm d})/\Phi_B(R)$ except very close to the disk. Altogether then, there is no obvious way for the disk to affect pseudodisk evolution significantly. We note in passing that this argument only says that most mass and angular momentum at large scale (both are concentrated in the pseudodisk near the midplane) are not affected by small (disk) scale dynamics. The same is not true for volume; large volumes in the polar region can be affected by the outflow launched at small scale.

For such a hierarchical evolution, it is possible to obtain correct pseudodisk evolution at significantly reduced computational cost, since resolving disk evolution at small scales is no longer necessary. Consider a simulation setup identical to our 3D simulation except that the inner boundary size is larger ($\gtrsim$ the circularization radius of infalling material) and we no longer forbid angular momentum from being advected through the inner boundary. We can also use a 2D (axisymmetric) domain, as the initial condition and pseudodisk evolution are both axisymmetric. In this case, a rotationally supported disk will be unable to form because the inner boundary is so large that all angular momentum just leaves the domain instead of being accumulated at small radii to allow disk formation and growth. While this type of large-inner-boundary simulation apparently does not directly produce the correct disk evolution, we do expect it to predict the pseudodisk properties relatively accurately if the physical picture of hierarchical evolution holds.

4.2.2 Coupling the 1D model to pseudodisk simulation

Now we outline how the pseudodisk evolution obtained with the aforementioned 2D large-inner-boundary simulation can be coupled to the 1D disk model to predict disk evolution.

Suppose we know $M_{\rm tot}$, $L_{\rm tot}$, and $R_{\rm d}$ at some time t; to evolve the system we want to get the value of these variables at $t+\Delta t$. For given disk size $R_{\rm d}$, the total mass of the protostar-disk system is

$$M_{\rm tot} \approx M(r{<}R_{\rm d}) \approx M^{\rm pd}(r{<}R_{\rm d}). \tag{4.8}$$

Here the superscript "pd" denotes the result obtained from the 2D pseudodisk simulation, which we expect to be a good approximation at $r \gtrsim R_{\rm d}$. [We use the mass within a spherical region $M(r{<}R_{\rm d})$ rather than a cylindrical region in order to exclude contributions from the low-density envelope at $z \gg R_{\rm d}$.] Using this relation, we can estimate $M_{\rm tot}(t+\Delta t)$ using $M^{\rm pd}(r{<}R)$ at $t+\Delta t$ and $R_{\rm d}$ at t. Then, for the angular-momentum evolution, we know that

$$\frac{{\rm d}L_{\rm tot}}{{\rm d}M_{\rm tot}} \approx \frac{F_L(R_{\rm d})}{F_M(R_{\rm d})} \approx \frac{F_L^{\rm pd}(R_{\rm d})}{F_M^{\rm pd}(R_{\rm d})}, \tag{4.9}$$

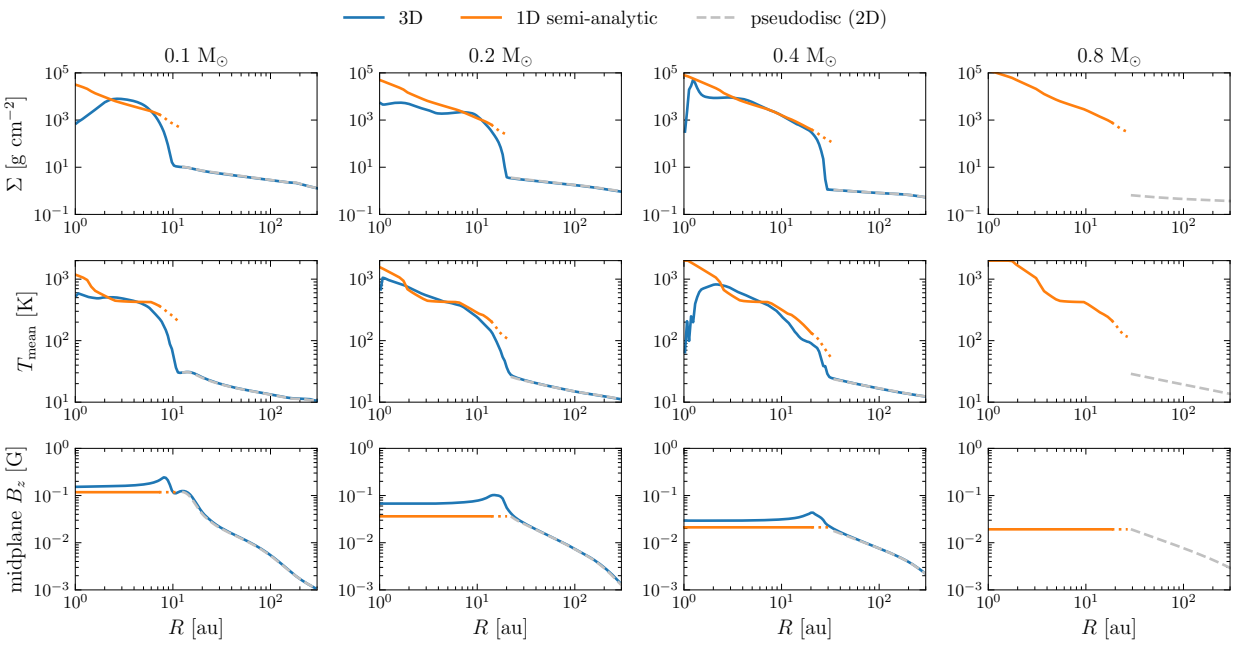

Figure 4.1: Comparison of disk profile between the 1D+pseudodisk model (orange) and our 3D simulation (blue) at a few different $M_{\rm tot}$. Solid lines correspond to $R \leq R_{\rm GI}$ and dotted lines mark $R_{\rm GI} \leq R \leq R_{\rm d}$ (where the assumption of $Q_{\rm K} = Q_0$ should no longer hold; we show this region only to mark disk size $R_{\rm d}$). The pseudodisk profile from our 2D, large-inner-boundary pseudodisk simulation (grey dashed) is also plotted at $R \geq R_{\rm d}$ for reference. The pseudodisk simulation agrees very well with the 3D simulation in this region.

where F_L and F_M are mass and angular-momentum fluxes (evaluated on the sphere with $r = R_{\rm d}$), respectively, due to advection. Note that we do not include the contribution from the magnetic stress to $F_L(R_{\rm d})$, which corresponds to the assumption that magnetic braking in the disk is negligibly weak. This is true for our simulation; more generally, one can verify this assumption by estimating the strength of braking with the model described in Section 3.5.2. Using this relation, we can update $L_{\rm tot}$ to $t + \Delta t$ as well. Next we estimate $\dot{M}_{\rm tot}(t + \Delta t)$ using $F_M^{\rm pd}(t + \Delta t)$. Finally, using the 1D model, we obtain $R_{\rm d}(t + \Delta t)$ from $M_{\rm tot}$, $L_{\rm tot}$, and $\dot{M}_{\rm tot}$ at $t + \Delta t$.

Using the procedure above (with initial condition $M_{\rm tot}$, $L_{\rm tot}$, $R_{\rm d} \to 0$), we can calculate the time evolution of the disk properties from pseudodisk simulation data.

4.3 Comparison between the 1D+pseudodisk model and the 3D model

Figs 4.1 and 4.2 compare the prediction of the 1D+pseudodisk model outlined in the two previous subsections (with the pseudodisk simulation performed using a 5-au inner-boundary size) with our 3D simulation. The

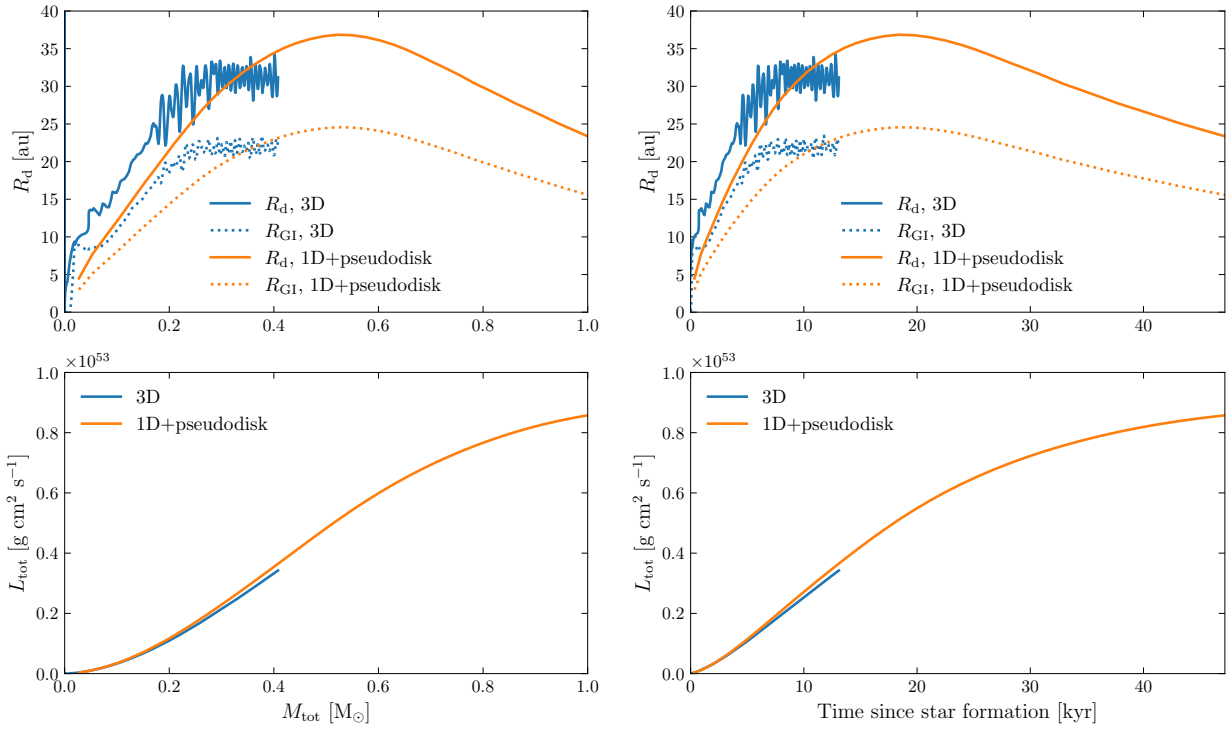

Figure 4.2: Left: Comparison of the evolution of disk size (top panel) and angular-momentum budget of the protostar-disk system (bottom panel) between the 1D+pseudodisk model (orange) and our 3D simulation (blue). In the top panel, the outer boundary of the gravitationally unstable region $R_{\rm GI}$ (defined by the location where $Q_{\rm K} = 2$ for our 3D simulation) is also plotted for reference. The 1D+pseudodisk model reproduces the evolution relatively well, and predicts disk shrinkage at late times. Right: same as the left panels but shows evolution against time. The 1D+pseudodisk model covers $\sim 4\times$ longer time after star formation with less than one day on 24 cores. For comparison, the 3D simulation took approximately four weeks on 96 cores.

comparison of disk profiles in Fig. 4.1 shows that the evolution is indeed largely hierarchical, with the 2D pseudodisk simulation accurately predicting all properties of the pseudodisk. The 1D model for the disk also shows reasonable agreement with the 3D simulation. We note that the increased difference in the innermost region ($\lesssim$2 au) is likely because the disk properties are being affected by the inner boundary in the 3D simulation. Also, the 1D model often shows some radial substructure in temperature, which is associated with a local minimum in the temperature-dependent opacity (cf. Fig. 5.1); such a feature is not visible in 3D mainly because in that case the disk is not isothermal in the vertical direction. This discrepancy between the 1D and 3D models can be resolved by taking into account the vertical temperature profile of the disk when computing the optical depth, as we do in Chapter 5. The comparison of disk-size evolution and angular-momentum budget in Fig. 4.2 also shows relatively good agreement. At early times, the 1D+pseudodisk model returns a disk size slightly smaller than observed in the 3D simulation, mainly because the assumption of a constant $Q_{\rm K} = 1.5$ would overestimate the column density at small radii (which in 3D is gravitationally stable for $M_{\rm tot}$ around 0.2 $\rm M_\odot$; see Fig. 3.11) and reduce the disk size when the total angular momentum is fixed. But later on, both the 3D simulation and the 1D+pseudodisk model show the disk size saturating at $\sim$30 au.

The 1D+pseudodisk model, which is evolved for a much longer time than is our 3D simulation, predicts the disk size to shrink eventually due to increasing magnetic braking in the pseudodisk (see Section 3.3.1). This trend is consistent with the recent survey of Class 0/I disks by Tobin et al. (2020) (see also Segura-Cox et al. 2018 and Andersen et al. 2019), which suggests a slight decrease in mean disk radius over time (from 45 au in Class 0 to 37 au in Class I). Our prediction of a gravitationally unstable disk of size $\sim$30 au is also consistent with a recent comparison between high-resolution multi-wavelength observations of a Class 0 disk and simulation-based mock observations, which suggests that the disk should be hot (at a few 100K) and gravitationally unstable (Zamponi et al., 2021). In Chapter 5 we will perform a more detailed comparison between our model and observations using data from a recent survey of protostellar disk (Tobin et al., 2020).

Overall, we find the 1D+pseudodisk model capable of predicting disk properties relatively accurately. The disk profile predicted from the 1D model serves as a good order-of-magnitude estimate in most cases, and the error in the prediction for disk size is no more than $\sim$50%. Given the high computational cost of performing long-term 3D simulations at sufficiently high resolution (comparable to our 3D simulation, or at least high enough to ensure numerical convergence), the 1D+pseudodisk model might serve as a good alternative to study the long-term disk evolution during the Class 0/I phase, especially for large parameter surveys.

We conclude this chapter by discussing some potential limitations of our 1D+pseudodisk model. Given our initial condition of an isolated pre-stellar core, our current simulations might not be able to capture

correctly the end of the main accretion phase and the dispersal of the envelope (and pseudodisk) around the protostar-disk system. Further studies of envelope (and pseudodisk) evolution during the Class 0/I phase, preferably in a more global (cluster- or cloud-scale) setup, would be necessary to study quantitatively the transition from Class I to Class II and predict the initial conditions of Class II disks. Moreover, the assumption of hierarchical evolution might no longer be applicable at these later times. This is mainly because the decrease in pseudodisk density could both increase the Alfvén speed and decrease the infall speed (as the pseudodisk becomes increasingly magnetically supported), allowing perturbations originating from the disk to propagate outwards in the pseudodisk. In summary, we expect our model to perform best in younger systems, and it might become less accurate at later times, when the envelope (pseudodisk) disperses and the system approaches Class II.

Chapter 5

Theory meets observation: testing our model on Orion protostellar disks

This chapter is adapted from material published in:

Xu, W. (2022). Testing a New Model of Embedded Protostellar Disks against Observations: The Majority of Orion Class 0/I Disks Are Likely Warm, Massive, and Gravitationally Unstable. ApJ, 934, 156.

In the previous chapter we introduced our semi-analytic disk model and discussed how it can be used to predict long-term disk evolution. This chapter features another application of our model, which is estimating disk properties by fitting our model to multi-wavelength dust-continuum observations.

There are two main motivations for fitting observations with our model. First, from an observational perspective, it is challenging to obtain direct estimates of disk properties (especially disk mass) from dust continuum observation (see our review in Section 1.3), because the observed flux depends not only on disk mass but also on disk temperature, dust opacity, and optical depth. Including multiple wavelengths barely helps with this problem, as the spectral index between different wavelengths also depends on multiple factors including optical depth, dust opacity index (which then depends on dust size distribution), and vertical temperature profile. It is therefore necessary to use a disk model with appropriate physical assumptions to lift these degeneracies. A common choice is to ignore these uncertainties by blindly assuming that the disk is optically thin and prescribing some arbitrary dust temperature (often $\sim$30 K) and opacity. This leaves large uncertainties in the result; in particular, when the disk is optically thick, this approach can significantly underestimate the disk mass. One can also avoid making these arbitrary assumptions and fit observation with parametrized disk models (often coupled with radiative transfer) with more degrees of freedom (e.g., Sheehan & Eisner, 2017; Sheehan et al., 2022). A major limitation of this approach, however, is that the model may be under-constrained by data. A sufficiently general model contains many ($\gtrsim$10) free parameters while currently available data sets often do not contain enough information to constrain these parameters reliably or test the physical assumptions involved in the model. Our model might be able to circumvent these limitations by using the physical constraints obtained in Chapter 3 (and summarized in Chapter 4) to avoid arbitrary assumptions while keeping the number of free parameters low.

Our second motivation comes from a theoretical perspective. Although we have demonstrated that our semi-analytic model can reproduce simulation results relatively accurately, it does not guarantee that it can correctly describe protostellar disks in nature, mainly because we cannot guarantee that our idealized initial condition is a reasonably good approximation to reality. Comparing our model with observation serves as a direct and quantitative test for whether our physical picture is applicable to reality.

With these motivations, we organize this chapter as follows. We begin by describing our model, with our model setup in Section 5.1, and details of how we fit the model to data in Section 5.2. We carefully test our model against observations in the VANDAM Orion survey (Tobin et al., 2020, hereafter T20) in Section 5.3, and discuss the estimates of Orion protostellar disk properties coming from our model with a focus on the physical implications of our results in Section 5.4. We compare our model with others in the literature to explain why we find much higher disk mass and demonstrate that our model has better predictive power in Section 5.5. We conclude with some additional discussion in Section 5.6.

5.1 Model setup

5.1.1 Physical assumptions

The physical assumptions of our model are generally the same as the 1D semi-analytic model discussed in Chapter 4. For the surface density profile of the disk, we assume a gravitationally self-regulated disk with constant, marginally unstable Q throughout the disk (except within the inner disk truncation radius; see Section 5.1.4),

$$Q = \frac{\kappa \bar{c}_s}{\pi \mathrm{G} \Sigma} \approx 1\text{–}2. \tag{5.1}$$

Here $\bar{c}_s$ the density-weighted average sound speed.

For the thermal budget of the disk, we assume that the disk is heated by gravito-viscous accretion and cooled by radiative cooling (cf. Section 3.4), so the effective temperature $T_{\rm eff}$ is given by

$$2\sigma T_{\rm eff}^4 \sim \frac{-g_R \dot{M}(R)}{2\pi R} \sim \frac{-g_R \dot{M}}{2\pi R} = \frac{1}{2\pi} \Omega^2 \dot{M}. \tag{5.2}$$

Here g_R is the radial gravity at the midplane, $\dot{M}(R)$ is the mass flux (accretion rate) through the disk at R, $\dot{M}$ is the rate of mass infall from the envelope onto the disk, and Ω is the orbital frequency. For simplicity, we assume $g_R = -\mathrm{G}M(<R)/R^2$ and $\Omega = \sqrt{-g_R/R}$. We also assume that $\dot{M}(R) \sim \dot{M}$, which is valid when the infall rate and the disk properties all evolve slowly at a timescale comparable to the lifetime of Class 0/I.

Assuming that the heating rate per mass at a given radius is approximately constant and that vertical convection is negligible (cf. Section 3.4.3), the vertical temperature profile can be well-approximated by a simple analytic form (Hubeny, 1990, Eq 3.11):

$$T^4(\tau_{\rm R}) \approx \frac{3}{4} T_{\rm eff}^4 \left[\tau_{\rm R} \left(1 - \frac{\tau_R}{2\tau_{\rm R,mid}} \right) + \frac{1}{\sqrt{3}} + \frac{1}{3\tau_{\rm P,mid}} \right]. \tag{5.3}$$

Here $\tau_{\rm R}$ is the Rosseland optical depth at a given location and $\tau_{\rm R/P,mid}$ is the Rosseland/Planck optical depth at the midplane. This approximation assumes that the opacities vary slowly in depth and the heating rate per mass is approximately constant.

There are only three minor differences between this model and the 1D model presented in Chapter 4. First, we no longer distinguish between the outer boundary of the gravitationally unstable region $R_{\rm GI}$ and disk size $R_{\rm d}$, because we do not have a good model to describe the transition in between. Second, this model self-consistently solves for the vertical temperature profile, which is necessary for computing observed flux at optically thick wavelengths. Finally, we no longer consider heating by the ambient temperature T_0, in order to reduce the number of free parameters in our model.

Comparison between previous studies and limitations of our assumptions

There are two main differences between our model and the assumptions in previous observational estimates of protostellar disk properties. The first is that we do not assume the emission to be optically thin at the observed wavelength. While typical protostellar disks are likely optically thick at (and even above) mm wavelength (as suggested by their low spectral indices; Li et al., 2017; Galván-Madrid et al., 2018), most previous mass estimates assumed optically thin emission for lack of a better method to infer the mass of obscured dust at $\tau \gtrsim 1$. Here we tackle this problem by using the physical constraints discussed above to relate the properties of the disk surface (at $\tau \lesssim 1$) to surface density and midplane temperature.

Another important difference concerns the assumed dust heating mechanism. When translating dust thermal emission to dust mass, most existing studies (e.g., T20; Sheehan & Eisner 2017; Sheehan et al. 2022) assume that dust temperature is set mainly by protostellar irradiation, while we assume that the dust temperature is set mainly by the gas in the disk, which is heated internally due to accretion during the main accretion phase.

The main limitations of our model also concern our temperature estimates. First, the envelope might heat the outer disk by scattering and reemitting the radiation from the protostar-disk system (D'Alessio et al., 1997). It is difficult to quantify the importance of this mechanism without introducing many free parameters as it would be sensitive to the properties of the envelope (Natta, 1993). Second, for a small subset of our sample the temperature in the outermost part of the disk can get below ~ 10 K and become comparable to the background temperature of interstellar irradiation; in this case this ambient irradiation could become the dominant source of heating. Generally, these limitations would lead us to underestimate the temperature in the outer disk in some cases.

5.1.2 Generating the disk profile

The physical assumptions given in the previous subsection provide enough constraints to generate the radial surface density profile and the radial and vertical temperature profile given three inputs: the mass of the protostar $M_\star$ (as the inner boundary condition), the disk size R_d (as the outer boundary), and the mass infall rate from the envelope $\dot{M}$. Scripts for solving the disk profile are available online at https://github.com/wxu26/GIdisk2obs; here we sketch out a method of solution.

First we discuss how we solve the surface density and vertical temperature profile at a given radius for a given set of Ω, κ, and $\dot{M}$. Here we treat "$\sim$" and "$\approx$" above as "$=$" when solving the disk profile and assume a constant $Q = 1.5$ in our fiducial model. We begin by mapping $(\tau_\mathrm{R,mid}, T_\mathrm{eff})$ to the vertical temperature profile and then to $(\Sigma, \bar{c}_s)$. For a given set of $(\tau_\mathrm{R,mid}, T_\mathrm{eff})$, we can solve $\tau_\mathrm{P,mid}$ by plugging Eq. 5.3 into the

constraint

$$\tau_{\rm P,mid} = \int_0^{\tau_{\rm R,mid}} \frac{\kappa_{\rm P}(T)}{\kappa_{\rm R}(T)} {\rm d}\tau_{\rm R}. \tag{5.4}$$

The Rosseland and Planck mean opacities $\kappa_{\rm P}, \kappa_{\rm R}$ are given in Section 5.1.4. Knowing $\tau_{\rm P,mid}$, we can then compute the vertical temperature profile $T(\tau_{\rm R})$ with Eq. 5.3 and use $T(\tau_{\rm R})$ to compute Σ (using the relation ${\rm d}\Sigma = \kappa_{\rm R} {\rm d}\tau_{\rm R}$) and $\bar{c}_s$. We then (numerically) convert this map into a map from $(\Sigma/\bar{c}_s, T_{\rm eff})$ to other local disk properties $(\Sigma, \tau_{\rm P,mid}, \tau_{\rm R,mid})$. Since Eq. 5.1 and 5.2 directly determine $\Sigma/\bar{c}_s$ and $T_{\rm eff}$, we can use this map to determine the local disk properties.

To generate a disk profile, we just need to update the radial profile of (Ω, κ) iteratively using the Σ profile calculated with current estimates of (Ω, κ) until the results converge. Here $M_\star$ and $R_{\rm d}$ are used as boundary conditions for this process.

5.1.3 Generating mock observation

Using the disk temperature profile $T(R, \tau_{\rm R})$ generated from our model, we can produce mock observations at a given wavelength (frequency).

We first compute the intensity $I_\nu(R)$ at disk surface using the temperature profile $T(R, \tau_{\rm R})$. Since the disk is often optically thick and the scattering opacity is generally higher than absorption opacity at the observed wavelengths, we need to include scattering in our calculation (Zhu et al., 2019). For simplicity, we assume that the disk is geometrically thin; this reduces solving I_ν at a given R to a 1D problem. The intensity is given by

$$I_\nu = \int_0^{\tau_{\nu,\rm tot}} e^{-\mu\tau_\nu} S_\nu \mu {\rm d}\tau_\nu. \tag{5.5}$$

Here τ_ν is the optical depth from disk surface (along vertical direction), $\tau_{\nu,\rm tot}$ is the total optical depth, $\mu = \cos I$ with I being the inclination of the disk, and S_ν is the source function. We map $\tau_{\rm R}$, which is used to specify the temperature profile, to τ_ν using the relation

$$\kappa_{\rm R}^{-1} {\rm d}\tau_R = (\kappa_{\nu,\rm abs} + \kappa_{\nu,\rm sca}^{\rm eff})^{-1} {\rm d}\tau_\nu. \tag{5.6}$$

Here $\kappa_{\nu,\rm abs}$ is the absorption opacity and $\kappa_{\nu,\rm sca}^{\rm eff} = (1 - g_\nu)\kappa_{\nu,\rm sca}$ is the effective scattering opacity, where g_ν is the forward-scattering parameter and the factor $(1 - g_\nu)$ accounts for anisotropic scattering (Ishimaru, 1978). The source function S_ν is given by

$$S_\nu = (1 - \omega_\nu) B_\nu(T) + \omega_\nu J_\nu(\tau_\nu), \tag{5.7}$$

where $B_\nu(T)$ is the Planck function, $J_\nu = \frac{1}{4\pi}\int I_\nu \mathrm{d}\Omega$ is the isotropic intensity, and $\omega_\nu = \kappa^{\rm eff}_{\nu,\rm sca}/(\kappa_{\nu,\rm abs}+\kappa^{\rm eff}_{\nu,\rm sca})$ is the single-scattering albedo. Now we only need to solve for J_ν. Under the Eddington approximation, the second moment of the radiative transfer equation becomes

$$\frac{1}{3}\frac{\mathrm{d}^2 J_\nu}{\mathrm{d}\tau_\nu^2} = (1-\omega_\nu)(J_\nu - B_\nu), \tag{5.8}$$

and we solve it (numerically) with the boundary condition of no incoming radiation field, which under the two-stream approximation is given by $\mathrm{d}J_\nu/\mathrm{d}\tau_\nu = \pm\sqrt{3}J_\nu$ at top/bottom surface (Miyake & Nakagawa, 1993). In the limit of high optical depth, including scattering generally reduces I_ν by a factor of $\mathcal{O}(\sqrt{1-\omega_\nu})$ (Zhu et al., 2019).

We then use the $I_\nu(R)$ profile to generate a mock observation image by orienting the disk using the position angle (PA) and inclination estimates from T20, which are based on the deconvolved shape of the best-fit gaussian profile of the 0.87 mm image, and convolve the image with a 2D gaussian beam whose widths and orientation are identical to the synthesized beam of the corresponding observation. This produces the intensity profile I_ν in Jy per beam.

5.1.4 Dust opacity and disk truncation

Generating disk profiles and generating mock observations from disk profiles both require knowledge of the opacity of the disk, which is dominated by dust opacity. Here we assume a constant dust-to-gas mass ratio of 0.01 and a power-law grain-size distribution with $\mathrm{d}n/\mathrm{d}a \propto a^{-3.5}$ (Mathis et al., 1977) having a maximum grain size $a_{\rm max} = 1$ mm and a minimum grain size $a_{\rm min} \to 0$. (As long as the minimum grain size $a_{\rm min} \ll a_{\rm max}$, the opacities are insensitive to the exact choice of $a_{\rm min}$.) The assumption of a constant dust-to-gas ratio should be reasonable, as the dust grains are well coupled to the gas (Appendix G).

Our choice of $a_{\rm max}$ assumes that the dust-size distribution is more similar to that in protoplanetary disks than that in the ISM, which would have $a_{\rm max} \lesssim \mu$m. This is because grain growth is expected to proceed quickly in the disk before approaching the equilibrium between coagulation and fragmentation (Birnstiel et al., 2011). Still, the exact value of $a_{\rm max}$ remains highly uncertain, as $a_{\rm max}$ can vary significantly across different disks and vary radially within a disk. To address this uncertainty, while we use a fiducial value of $a_{\rm max} = 1$ mm, we also produce models with $a_{\rm max} = 10$ μm, 100 μm, and 1 cm and compare the results across these models.

For this chapter, we adopt a dust opacity model that is more flexible than the one adopted in our simulation (Fig. 2.4) in order to account for grain sublimation and different grain-size distributions appropriately. The dust opacities are computed using the DSHARP opacity package (Birnstiel et al., 2018). The

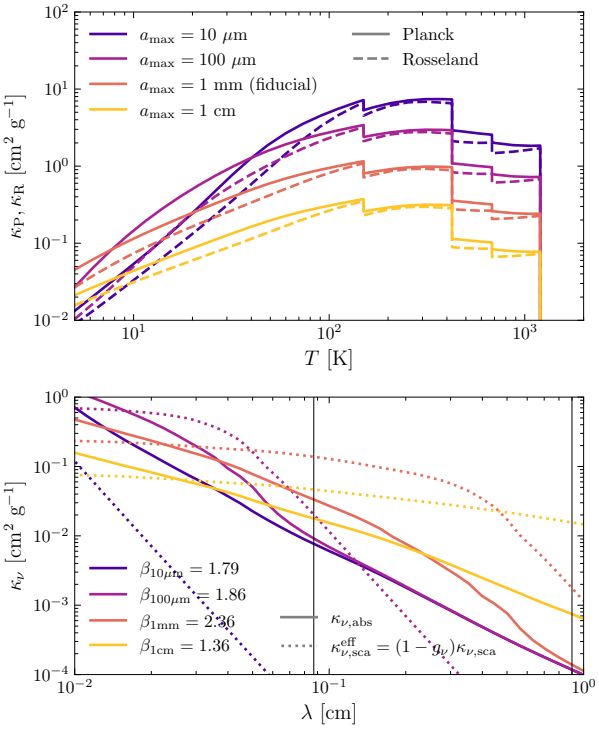

Figure 5.1: Top panel: Planck and Rosseland mean opacities $\kappa_{\rm P}, \kappa_{\rm R}$ for dust models with different maximum grain size $a_{\rm max}$. (For $a_{\rm max} < 10$ μm, the results are nearly identical to $a_{\rm max} = 10$ μm.) Bottom panel: absorption opacity $\kappa_{\nu,{\rm abs}}$ and effective scattering opacity $\kappa^{\rm eff}_{\nu,{\rm sca}}$ before any dust sublimation ($T < 150$ K). We also label the dust opacity indices β between 0.87 mm and 9 mm (vertical black lines) for $\kappa_{\nu,{\rm abs}}$.

Table 5.1: Summary of fiducial model parameters

Free parameters	
$M_\star, R_{\rm d}$	
Fixed parameters	
Toomre Q	1.5
Max grain size $a_{\rm max}$	1 mm
$\dot{M}/M_\star$	10^{-5} yr^{-1}
dust-to-gas ratio	0.01
grain porosity	0

assumptions for grain porosity and composition also follow that of Birnstiel et al. (2018), with zero-porosity grains composed of 20% water ice, 40% refractory organics, 7% troilite, and 33% silicates by mass. Since the temperatures in the inner part of our model profiles are often several hundred K and above, we need to include the effect of dust sublimation. We adopt the sublimation temperatures from Pollack et al. (1994) and remove water ice, refractory organics, troilite, and silicates from our dust mixture (while adjusting the dust-to-gas ratio accordingly) at 150, 425, 680, and 1200 K, respectively. Here for simplicity we have dropped the density dependence of water ice and silicate sublimation temperatures. Beyond 1200 K, we assume that $\kappa_{\rm P}, \kappa_{\rm R} \to 0$ as the remaining gas opacity there is much lower than the dust opacity below 1200 K. In Fig. 5.1 we summarize the opacities for dust models with different $a_{\rm max}$.

Once disk material exceeds 1200 K, it can barely cool by radiation due to the drop in opacity and thermal equilibrium can only be achieved at a much higher temperature when gas opacity becomes sufficiently high.[1] This results in a steep increase of temperature at $R < R_{\rm 1200K}$, where $R_{\rm 1200K}$ is the radius where midplane temperature first reaches 1200 K. The high temperature at $R < R_{\rm 1200K}$ would lead to well-coupled magnetic field because the thermal ionization of potassium increases disk ionization exponentially above $\sim$ 1000 K (Umebayashi, 1983). In this regime, our model is no longer applicable, and the disk is likely to be regulated by MRI and magnetized outflow, which efficiently transport angular momentum (with effective $\alpha \gtrsim 0.01$) to keep the surface density at $R \lesssim R_{\rm 1200K}$ low (cf. Gammie, 1996). For simplicity, we excise this region from our calculation and assume that the surface density and flux density are zero at $R < R_{\rm 1200K}$.

5.1.5 Choice of model parameters

We conclude this section by summarizing all parameters of our model and discussing the choice of free (and fixed) parameters. The disk profile can be solved with only three parameters: the mass of the star $M_\star$, the disk size $R_{\rm d}$, and the rate of mass infall (accretion) from the envelope (pseudodisk) to the disk $\dot{M}$ (Section 5.1.2). One intuitive choice is to leave all three parameters as free parameters; this, however, will

[1] Here we simply assume that the opacity beyond 1200 K is effectively zero. The behavior of the disk can be more complex when the finite dust opacity is taken into consideration; this will be investigated in our future studies.

be problematic for unresolved disks, where observation only has two degrees of freedom corresponding to the integrated flux at the two observed wavelengths. (There is a little additional information as a finite disk size will always cause the profile to deviate from being exactly Gaussian; but such information is negligible when the disk diameter is significantly smaller than the beam size.) Therefore, in order to fit all observed systems, we want our model to have at most two free parameters. This is achieved by assuming a constant $\dot{M}/M_\star = 10^{-5}$ yr^{-1}. The choice of this ratio is motivated by the typical lifetime of the main accretion phase (Class 0 and Class I), and is also similar to our simulation results. The mass dependence is meant to capture (to some extent) the possible difference in accretion rates between low-mass and high-mass systems, but not the variation of accretion rate in a single system as the protostar gains mass. This is of course a very crude approximation; it is unclear whether the duration of the main accretion phase has any mass dependence, and it does not correctly capture the temporal variation of accretion rate within the main accretion phase (where $\dot{M}$ should remain approximately constant or decrease as $M_\star$ and the age of the system increase). We address the large systematic uncertainty in our assumed $\dot{M}$ by quantifying how a different choice of $\dot{M}/M_\star$ impacts the result in Section 5.2.3 and 5.3.1.

We also comment that there are several other ways for estimating $\dot{M}$, but they each have their own limitations. First, one can simply choose a fixed $\dot{M}$, but it would not be reasonable to assume that all systems – which span several orders of magnitude in mass – have identical accretion rates. Besides, we find that the best-fit model for some low-mass systems would have $M_\star \to 0$ if we assume a fixed $\dot{M} \sim 10^{-5}$ M$_\odot$ yr^{-1}. Another possibility is to estimate $\dot{M}$ using the luminosity of the protostar $L_\star$. However, $L_\star$ depends on the accretion rate onto the star $\dot{M}_\star$ and can be highly variable (potentially due to the modulation of the disk), such that an instantaneous estimate of $\dot{M}_\star$ is not necessarily a good approximation of the mass infall rate $\dot{M}$ (cf. Offner & McKee, 2011; Zakri et al., 2022). Therefore, our choice of a constant $\dot{M}/M_\star$ is probably already the best we can do before observations reach better resolution (allowing us to fit $\dot{M}$ as a free parameter) or provide direct estimates of $\dot{M}$ (e.g., through envelope dynamics, which is already available for a small number of systems, e.g., Kristensen et al. 2012; Pineda et al. 2012).

In addition to our two free parameters $M_\star, R_\mathrm{d}$ and our fixed $\dot{M}/M_\star$, there are a few additional fixed parameters to the model, including an assumed constant Toomre Q of 1.5 (following the idea of gravitational self-regulation) and parameters for the fiducial dust model (Section 5.1.4). We summarize the fiducial model parameter choices in Table 5.1. In Sections 5.2.3, 5.3.1, and Appendix H we also discuss how different choices of these fixed parameters impact the agreement between model and observation and the estimated disk properties.

5.2 Fitting the model to observation

5.2.1 Sample selection

We use observations from the VANDAM Orion survey (T20; Tobin, 2019a,b), a large survey of protostellar systems in the Orion molecular clouds with ALMA (0.87 mm) and VLA (9 mm) at ∼40 au resolution. Our sample consists of all systems that are detected at both wavelengths and have positive deconvolved major and minor axes at 0.87 mm (which are used for estimating inclination). This gives a total of 163 systems; among them, 98 are Class 0, 40 are Class I, 21 are Flat Spectrum, and 4 are unclassified.

5.2.2 Fitting

For each system, we vary the two free parameters of the model, $M_\star$ and R_d, to minimize the error between observed and model images (flux density). We fit the model with images as opposed to visibilities (in frequency space) because it is easier to characterize the systematic uncertainty of the model (due to the oversimplifications we made) in real space rather than in frequency space.

We evaluate the error between the observed flux density I_obs and the model flux density I_model using

$$\chi^2 \equiv \min\left\{\frac{|I_\mathrm{obs}-I_\mathrm{model}|^2}{2\sigma_\mathrm{obs}^2}, \frac{\log(I_\mathrm{obs}/I_\mathrm{model})^2}{2\sigma_{\log\ \mathrm{model}}^2}\right\}. \tag{5.9}$$

Here σ_obs is the observation uncertainty, estimated with the RMS flux density of each field of observation, and $\sigma_{\log\ \mathrm{model}}$ captures the systematic uncertainty of the model due to the many oversimplified assumptions in our model. The errors in disk properties due to these oversimplifications are generally of order unity (cf. XK21b Fig. 16); therefore we choose $\sigma_{\log\ \mathrm{model}} = \log(2)/2$, such that $\pm1\sigma$ covers a factor of 2 difference in $I_\mathrm{obs}/I_\mathrm{model}$. χ^2 roughly corresponds to the minus log likelihood (per beam) of observing the given deviation between model and observation.

In order to find the best-fit model, we minimize

$$l \equiv \int \chi_A^2 S_{B,A}^{-1}\mathrm{d}S + \int \chi_V^2 S_{B,V}^{-1}\mathrm{d}S. \tag{5.10}$$

Here $S_{B,A/V} = \frac{\pi}{4\ln 2} b_{A/V,\mathrm{maj}} b_{A/V,\mathrm{min}}$ is the beam area (converted to au^2 to match dS), with b being the FWHM. Roughly speaking, $\int ... S_{B,A/V}^{-1}\mathrm{d}S$ represents a summation over beams. The integral covers a square region around the protostar with width 800 au or 4× the disk radius estimate in T20, whichever is larger. l can be interpreted as the minus log likelihood of observing the given error.

We minimize l under the constraint that the model flux density at 0.87 mm (before blurring) needs to

be $\geq\sigma_{\rm obs}$ at the disk edge. This is because the fit becomes much less reliable when the flux is below the detection limit, as the model tends to fit the envelope emission around the disk (which is not included in our model) by incorrectly increasing the disk size. When the fit is affected by this constraint, the best-fit disk size should be interpreted as an estimate for the lower limit of the true disk size. For our fiducial model, 31 systems are in this regime.

5.2.3 Quantifying systematic uncertainties

In order to avoid overfitting, we have reduced the number of free parameters in our model by making a series of oversimplifying and/or arbitrary assumptions. It is important to evaluate whether our estimated disk properties are sensitive to these assumptions and quantify the systematic uncertainties in our results. This is done by fitting our sample with different assumed model parameters, and using the measured dependence to estimate the systematic uncertainties in our results. The details of this process are documented in Appendix H. In summary, assuming one order-of-magnitude uncertainty in $a_{\rm max}$ and $\dot{M}/M_\star$ and an order-unity uncertainty in Q, the uncertainties in key disk properties are generally of order unity and at most a factor of ~ 3 (see Table H.2).

5.3 Testing the model against observation

5.3.1 Agreement between model and observation

In order to test our model against observation, we first check whether the model is consistent with a large fraction of systems in our sample. Note that we do not expect the model to be consistent with all systems, as the physical assumptions of our model corresponds to only one of several possible scenarios of disk formation (see Section 5.6.3).

We evaluate the agreement between the best-fit model and observation using the mean χ^2 inside each disk,

$$\chi^2_{\rm mean} = \frac{\chi^2_A S_A S^{-1}_{B,A} + \chi^2_V S_V S^{-1}_{B,V}}{S_A S^{-1}_{B,A} + S_V S^{-1}_{B,V}}. \tag{5.11}$$

Here $S_{A/V}$ is the area of the region where the ALMA/VLA mock observation is above the detection limit, and $\chi^2_{A/V}$ is the mean χ^2 for the ALMA/VLA mock observation within this region. In other words, $\chi^2_{\rm mean}$ is an average of $\chi^2_{A/V}$ weighted by the (detectable) disk area in unit of beams. Fig. 5.2 provides a few examples of systems at different $\chi^2_{\rm mean}$. For $\chi^2_{\rm mean} \lesssim 1$, the model agrees well with observation. Fig. 5.3 plots the distribution of $\chi^2_{\rm mean}$ for our fiducial model; 57% of our sample can be fit reasonably well by the model ($\chi^2_{\rm mean} \leq 1$).

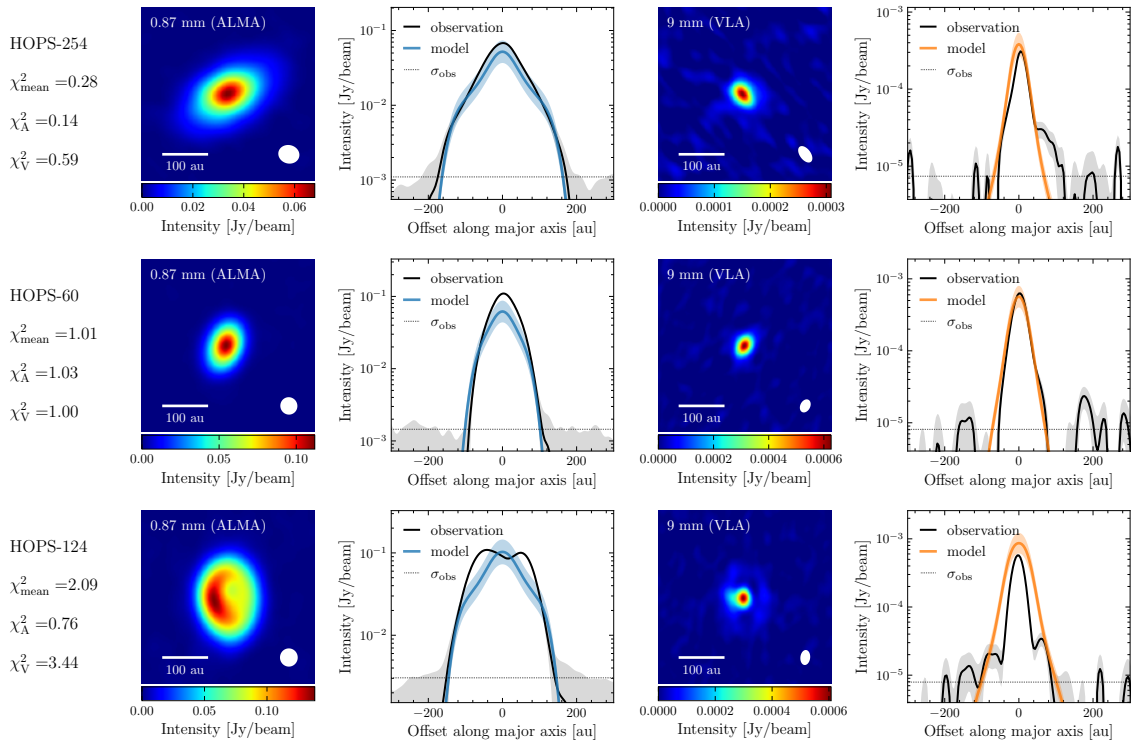

Figure 5.2: Examples of well-resolved systems with different $\chi^2_{\rm mean}$. Each row shows the observations at both wavelengths and compares the best-fit model with observation by taking cuts along the disk major axis (positive offset is in the direction of increasing RA). Shaded areas show 1σ observation and model uncertainties. The first two panels show examples where the model agrees well with the observations ($\chi^2_{\rm mean} \lesssim 1$); 57% of our sample are in this regime. The last panel shows an example where the model cannot fit the observation very well. For this particular system this is due to the presence of substructure (which at current resolution is uncommon in the sample). In general, the lack of a good fit could be due to a variety of reasons, such as disk being inconsistent with our model, envelope contamination, and binarity.

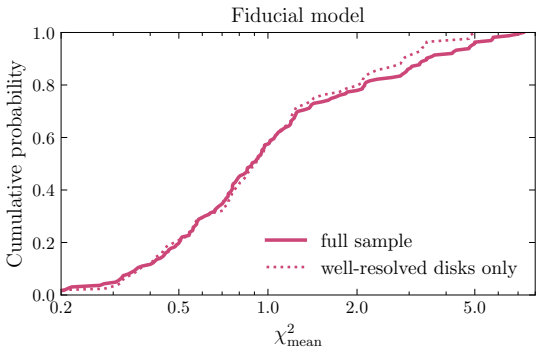

Figure 5.3: Cumulative distribution of model error $\chi^2_{\rm mean}$ for our fiducial model. 57% of our sample can be fit well with our model ($\chi^2_{\rm mean} \leq 1$). The result is similar if we limit the sample to only well-resolved disks (T20 disk size estimate >50 au).

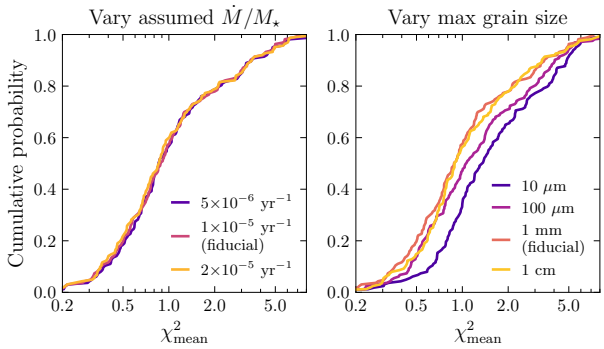

Figure 5.4: Cumulative distribution of model error $\chi^2_{\rm mean}$ when we vary the assumed $\dot{M}/M_\star$ (left panel) and the assumed dust size distribution (right panel). The agreement remains overall similar except for very small grain size ($a_{\rm max} = 10\mu$m).

Fig. 5.4 shows the distribution of $\chi^2_{\rm mean}$ if we vary the assumed dust-size distribution and accretion rate, which are both subject to relatively large systematic uncertainties. The agreement between model and observation is relatively insensitive to different choices of these parameters, while showing a slight preference for a maximum grain size $a_{\rm max} \sim$ mm over larger (1 cm) or smaller ($\lesssim$100 μm) grains.

To give a more direct impression on how well the model fits observation, we also compare flux density (Fig. 5.5) and apparent disk size (Fig. 5.6) between model and observation. Our model is consistent with observations at both wavelengths, and there is no apparent systematic difference in any of these metrics. Especially, our model naturally reproduces the trend that the apparent disk size shrinks towards longer wavelength, which is also visible in Fig. 5.2.

5.3.2 Is the good agreement coincidental?

Good agreement between observation and model prediction alone cannot be a strong evidence for arguing that the model is a good description of real disks. It remains possible that the good agreement is just an overfit, where the model is too flexible (or the observables provide too few constraints) and predicts an unrealistic disk profile which happens to produce the right observables. While one cannot completely rule out this possibility, we argue that an overfit is unlikely as follows.

First, we note that our model has only two degrees of freedom (two free parameters, $M_\star$ and $R_{\rm d}$), while the observations often offer more constraints than that. For disks that are well resolved by ALMA, the model needs to fit not only the integrated flux at both wavelengths but also the disk size and the radial profile of the flux density. In Fig. 5.3 we see that the agreement between the model and well-resolved systems remains good, suggesting that the agreement between the model and observation is likely not because the model has too many free parameters.

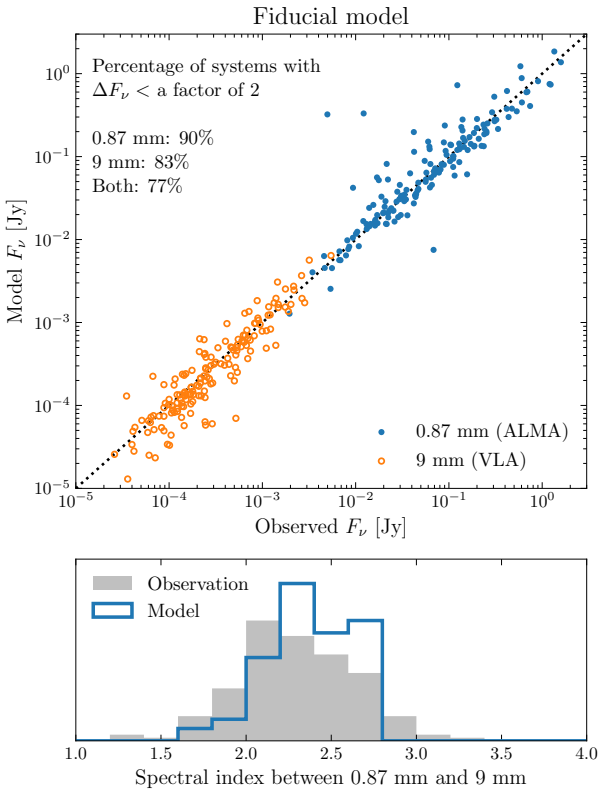

Figure 5.5: Top panel: Comparison between flux densities from our model and observation. The observed flux densities are quoted from T20. Bottom panel: the distribution of spectral indices. For most of the systems, our model is able to fit the flux at both wavelengths well.

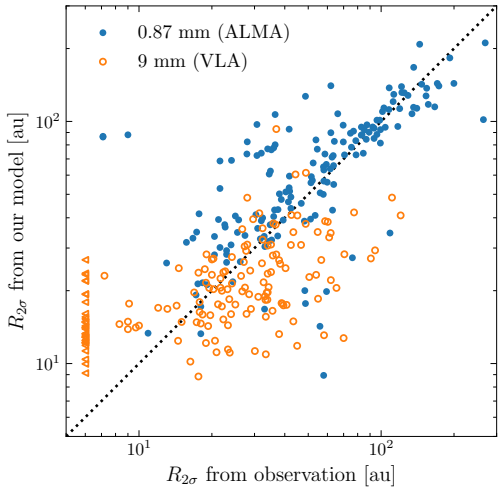

Figure 5.6: Comparison between apparent disk sizes from our model and observation. Here the apparent disk size $R_{2\sigma}$ is 2σ of the deconvolved gaussian fit to the observed image or mock observation. Our model broadly agrees with observation at both wavelengths, and reproduces the trend that $R_{2\sigma}$ is systematically smaller at longer wavelength.

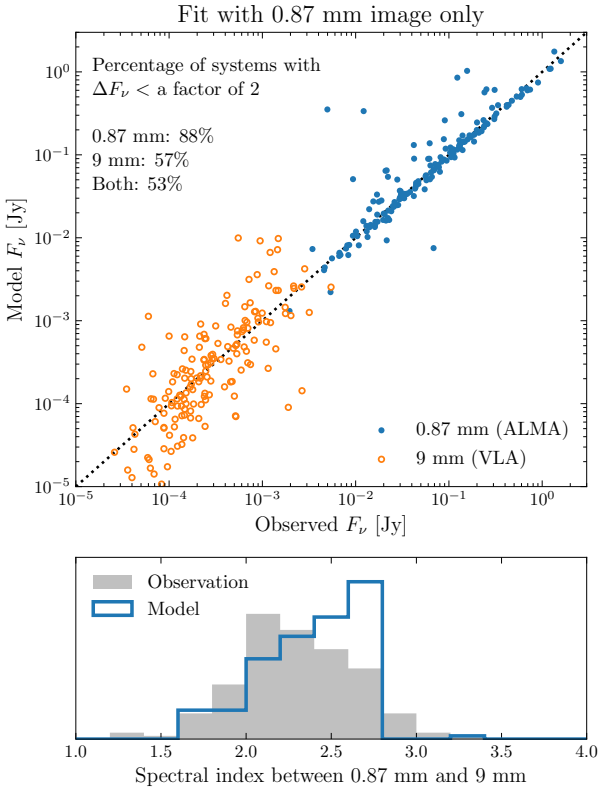

Figure 5.7: Same as Fig. 5.5 but shows results for our model fitted using only the 0.87 mm image. There is more scattering for the 9 mm flux but no systematic difference between model prediction and observation. This demonstrates the predictive power of our model, as most of the 9 mm flux come from dust that is not visible ($\tau \gg 1$) at 0.87 mm.

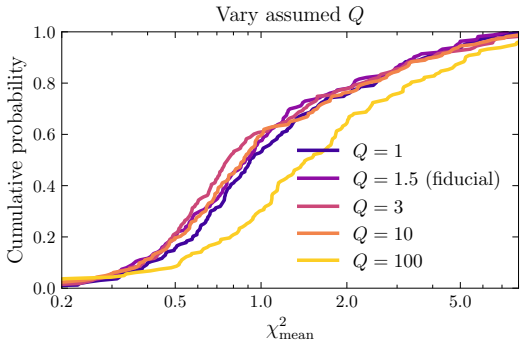

Figure 5.8: Cumulative distribution of model error $\chi^2_{\rm mean}$ when we vary the assumed Toomre Q. Observation strongly prefers $Q \lesssim 10$ over $Q \sim 100$.

Second, we test the predictability of our model by fitting it only with single-wavelength 0.87-mm (ALMA) observations and use it to predict the 9-mm flux. This is not a trivial task; the emission is generally optically thick at 0.87 mm, and most of the 9-mm emission come from dust that is not visible at 0.87 mm. Therefore, correctly predicting the 9-mm flux requires a good estimate of the total dust mass using the emission at the surface of the disk and the assumed physical constraints in our model. In Fig. 5.7 we see that the model predicts the 9-mm flux without systematic error and reproduces the typical spectral index relatively well.

5.3.3 Does observation favor gravitationally self-regulated disks?

A central assumption of our model is that the disk is gravitationally self-regulated with a marginally unstable Toomre Q. However, as we commented earlier, this assumption should not be taken for granted (cf. Section 5.6.3). Here we try to use observation to constrain (at the population level) the typical Toomre Q parameter of our sample.

In Fig. 5.8 we compare the distribution of $\chi^2_{\rm mean}$ for several different assumed values of Q. The agreement between observation and model is similarly good for $Q = 1 - 10$, but deteriorates as we further increase Q. Fig. 5.9 compares the distribution of disk spectral index for different Q values; it shows a preference for $Q = 1.5$ over larger Q values, with the caveat that the difference between $Q = 1.5$ and $Q = 3$ remains small.

In summary, these evidences show that order-unity Q is preferred at a population level. We can understand this preference by noticing that, at a higher Q and for the same (optically thick) 0.87-mm emission, the model would under-predict the surface density and 9-mm flux, making it difficult to reproduce the observation at both wavelengths.

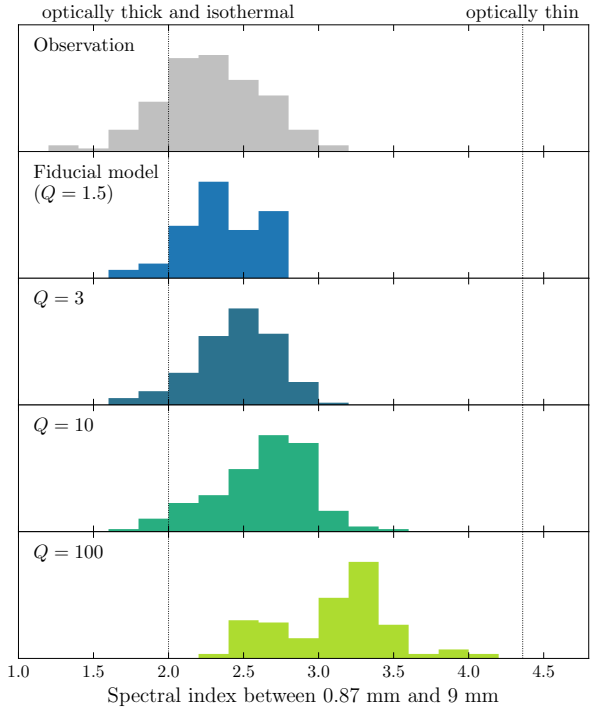

Figure 5.9: Distribution of disk spectral indices for observed systems (first panel), our fiducial model (second panel), and models assuming larger Q values (last three panels). At large Q, the model is no longer able to reproduce the observed spectral index distribution, even though observations at both wavelengths are used to fit the model.

	$M_{\rm d}$ [$M_\odot$]				$R_{\rm d}$ [au]			$T_{\rm mean}$ [K]		
	Mean	25%	Med	75%	25%	Med	75%	25%	Med	75%
All	0.68	0.23	0.52	0.85	45.3	77.7	144.4	228.0	272.4	311.6
Class 0	0.72	0.26	0.62	0.93	48.8	92.8	174.1	221.4	268.7	299.8
Class I	0.73	0.30	0.45	0.87	36.7	54.4	119.1	268.1	305.5	331.4
Flat Spectrum	0.39	0.17	0.24	0.40	44.0	54.6	78.9	238.6	253.3	265.7
T20 (0.87 mm)	0.024	0.005	0.013	0.031	32.5	51.8	82.9	44.0	58.0	71.9

Table 5.2: Distributions of key disk properties. Here we only include systems with $\chi^2_{\rm mean} \leq 2$. The last row shows results from T20 (based on 0.87 mm observation) for the same sample.

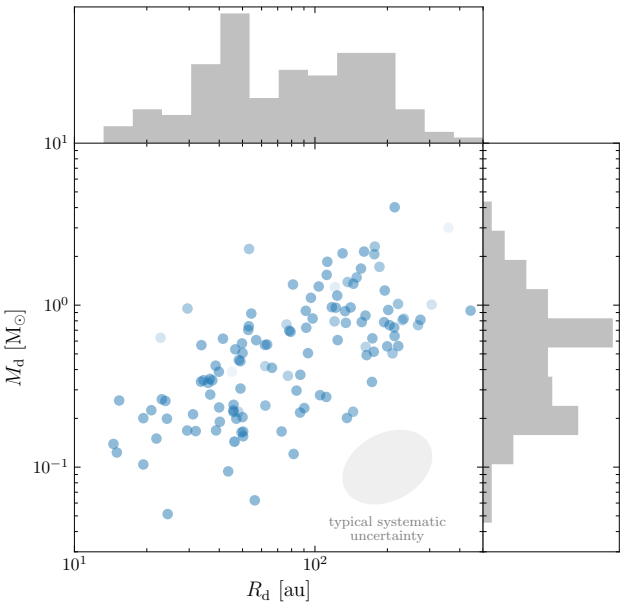

Figure 5.10: Distribution of disk size R_d and mass M_d for our fiducial model. We only include systems with $\chi^2_\mathrm{mean} \leq 2$; for $\chi^2_\mathrm{mean} > 1$, the opacity of the marker decreases as χ^2_mean increases. We also plot the 1σ uncertainty due to systematic uncertainties in the assumed model parameters for reference (cf. Appendix H). There is a positive correlation between disk size and mass.

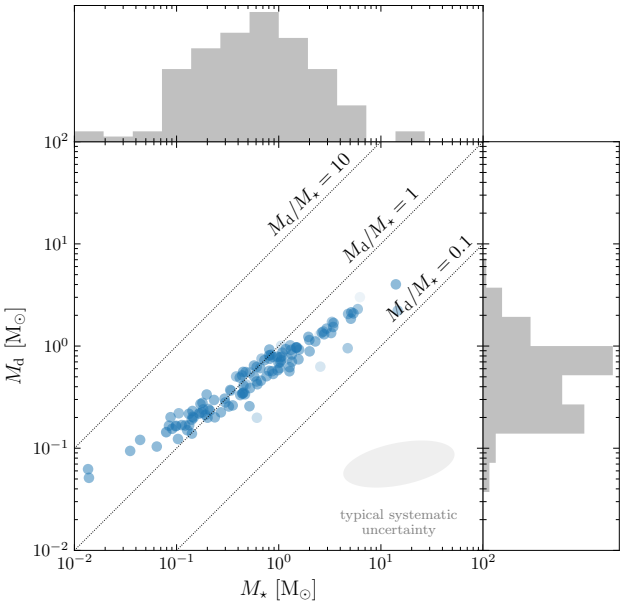

Figure 5.11: Same as Fig. 5.10, but for the distribution of stellar mass $M_\star$ and disk mass M_d. Most systems have order-unity disk-to-star mass ratio. There appears to be a strong correlation between M_d and $M_\star$, with slope ≈ 0.6; this is a generic property of gravitationally self-regulated disks (Section 5.4.1; cf. Appendix I).

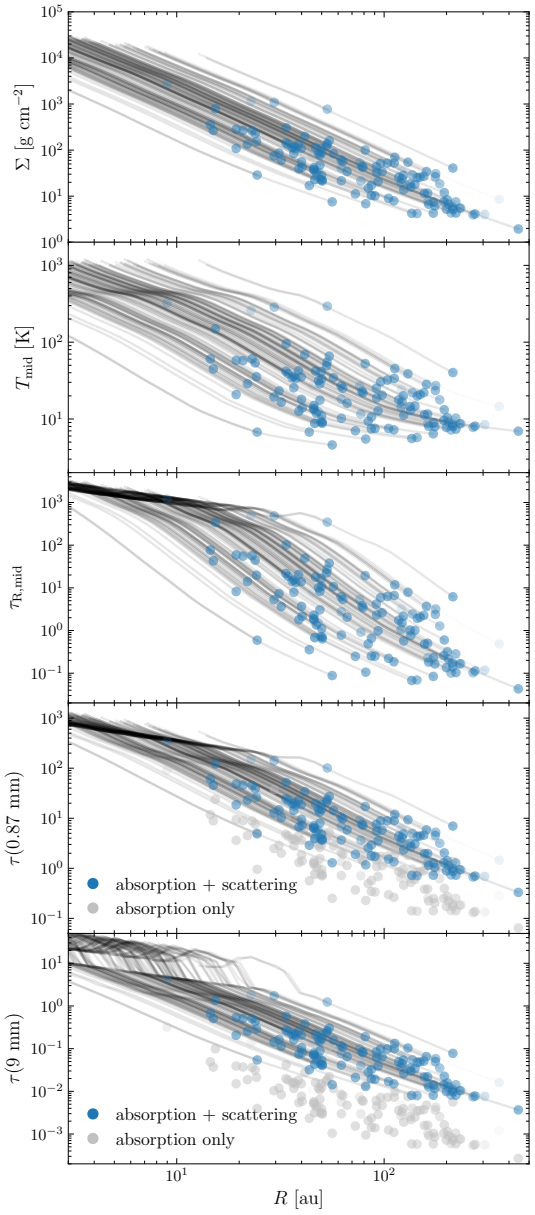

Figure 5.12: Radial profiles of disks in our sample, estimated with our fiducial model. From top to bottom, we show surface density Σ, midplane temperature $T_{\rm mid}$, miplane Rosseland mean optical depth $\tau_{\rm R,mid}$, and optical depth at the two observed wavelengths. We mark the disk edge with a blue dot in each curve. For the last two panels we also show the optical depth at disk edge for absorption opacity only. We only include systems with $\chi^2_{\rm mean} \leq 2$; for $\chi^2_{\rm mean} > 1$, the opacity of the marker decreases as $\chi^2_{\rm mean}$ increases.

5.4 Disk properties and physical implications

In this section we discuss the new estimates of the properties of Orion protostellar disks obtained with our fiducial model and discuss the physical implication of our results on disk evolution and fragmentation.

5.4.1 Summary of disk properties and comparison with previous studies

We begin with an overview of the estimated properties of our disks, which are summarized in Table 5.2 and Figs. 5.10 - 5.12. Here we only focus on systems with $\chi^2_{\rm mean} \leq 2$, as for systems with large $\chi^2_{\rm mean}$ our model may not be able to provide reliable estimates.

Disk mass and surface density: Our sample has a mean disk mass of 0.68 $M_\odot$ and median disk mass 0.52 $M_\odot$, which is often comparable to or more massive than the protostar mass; there also appears to be a correlation between disk mass and stellar mass, with slope ≈ 0.6 (Fig. 5.11). The surface density profile of our disks approximately scale as $\Sigma \propto R^{-2}$ (Fig. 5.12), therefore a significant portion of disk mass is concentrated in the inner part of the disk. These properties are all generic features of a gravitationally self-regulated disk, and agree with previous semi-analytic calculations (e.g., Lin & Pringle 1990; Rafikov 2009; also see Appendix I). Also note that the tight correlation in Fig. 5.11 is partly because we have assumed (for simplicity) constant $\dot{M} \propto M_\star$. In reality, the distribution would likely be more scattered.

Our mass estimates are significantly higher than previous estimates for the same data set (T20, Sheehan et al. 2022). This difference is mainly due to different model assumptions; we make a more detailed comparison between these models and argue that our model features better predictive power in Section 5.5.

Our disks are also somewhat more massive than the disks in the large-scale hydrodynamics disk population synthesis by Bate (2018), which gives $M_{\rm d}/M_\star \sim 0.1 - 1$. This difference could be related to insufficient resolution as commented in the resolution study in Bate (2018). XK21b also demonstrated that $M_{\rm d}/M_\star$ can be $\mathcal{O}(1)$ even when magnetic fields have been included, as long as there is no excessive numerical dissipation (which might be common among earlier simulations due to their relatively low resolution in the innermost several 10 au).

Disk size: The disk size in our sample has a median of 77.7 au and shows a wide distribution, with a factor of $\sim$3 difference between the first and third quartile. Our disk sizes are often larger than those quoted in T20, but this is mainly because T20 reports the apparent disk size $R_{2\sigma}$; there is no longer a systematic difference if we compare $R_{2\sigma}$ of our model with the T20 results (Fig. 5.6). We also see a positive correlation between disk mass and disk size; a similar correlation has been reported for older disks (Tripathi et al., 2017).

Disk temperature: The midplane temperature in our disks ranges from $\lesssim$10 K to 1200 K, and the

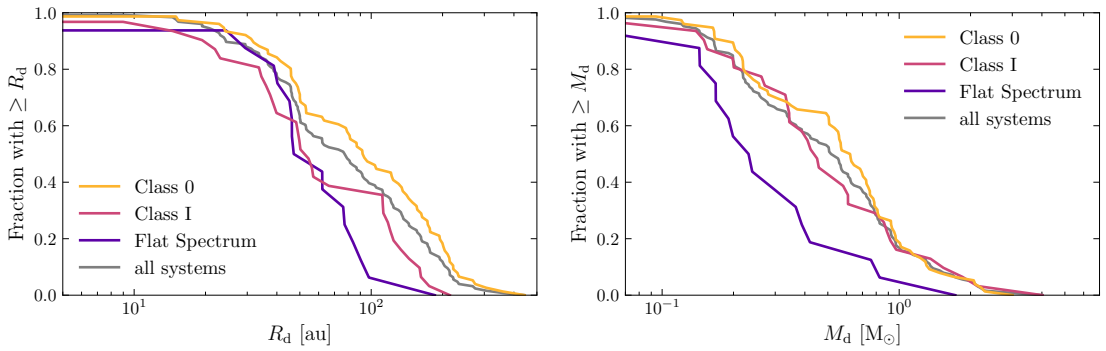

Figure 5.13: Cumulative distribution of disk size (top panel) and mass (bottom panel) for different evolutionary stages. Both disk size and mass decrease towards later evolutionary stages. These trends are discussed in Section 5.4.2.

mean temperature $T_{\rm mean}$ is generally several 100 K. Such high temperatures are in part because the disk is often optically thick ($\tau_{\rm R,mid} \gtrsim 1$) and cannot cool efficiently. It is also related to the steep surface-density scaling, which makes the hotter inner disk dominate the density-weighted averaging. The disk temperature decreases in radius with a relatively steep power-law slope ≈ -1, and $T_{\rm mid}$ generally varies by more than an order of magnitude across the disk. (As a result, $T_{\rm mean}$ should not be interpreted as a single characteristic temperature of the disk.) Also note that the vertical temperature profile (Eq. 5.3) implies that the observed dust (from the disk surface) can be significantly cooler than $T_{\rm mid}$, especially at 0.87 mm.

5.4.2 Implication on disk evolution

Now we discuss the evolution of disk size and mass by comparing their statistics across different evolutionary stages and against Class II disks in the literature. We also attempt to link the observed trends with theories of Class 0/I disk evolution.

Disk mass evolution: The estimated masses of Class 0 and Class I disks in our sample are largely similar, while those of Flat Spectrum disks are lower by more than a factor of 2.[2] Physically, the decrease of disk mass towards later evolutionary stages is expected because as the envelope disperses, the accretion rate could drop exponentially (Fischer et al., 2017) and that leads to less accretion heating, lower disk temperature, and thus a lower disk mass required for gravitational self-regulation. In other words, as the accretion rate drops, a gravitationally self-regulated disk tends to decrease its mass to maintain marginal instability (this could be achieved by keeping the accretion rate from the disk to the star slightly above the accretion rate from the envelope to the disk).

[2]This relation could be biased by our assumption of a constant $\dot{M}/M_\star$, as in reality $\dot{M}/M_\star$ should systematically decrease toward later stages. But given the weak correlation between assumed $\dot{M}/M_\star$ and estimated $M_{\rm d}$ (Table H.1), this qualitaive trend of $M_{\rm d}$ evolution should remain the same.

One could also estimate the mass of a gravitationally self-regulated disk at the end of envelope dispersal as an initial condition of Class II evolution. At that point accretion heating is low enough that the disk temperature should become comparable to the ambient temperature of ~ 10 K, and that produces a disk mass of order $0.1 M_\odot$ for a solar-mass star with a ~ 100 au disk (cf. Xu & Kunz 2021a Section 5.3). This is still significantly higher than early observational estimates of Class II disk masses in young ($\lesssim 3$ Myr) star-forming regions (Taurus, Ophiuchus, Lupus, ONC), which are of order $3 \times 10^{-3} M_\odot$ (Andrews et al. 2013; Ansdell et al. 2016; Tripathi et al. 2017; Eisner et al. 2018; see a summary in T20 Figs. 14, 15). But the simple models applied in these studies could systematically underestimate the disk mass by 1-2 orders of magnitude according to recent studies adopting different (and probably more robust) methods of mass estimation (Booth et al., 2019; Powell et al., 2019; Anderson et al., 2022). Accounting for this bias, the typical disk mass at the beginning of Class II would be broadly consistent with gravitationally self-regulated evolution during Class 0/I.

Disk size evolution: The size of disks in our sample decreases after Class 0, consistent with the trend observed in T20. While there is no firm conclusion on what causes this trend, one possible explanation for the shrinking disk size is enhanced magnetic braking in the inner envelope at later times (XK21b). During the collapse of a magnetized core, ambipolar diffusion decouples magnetic flux from the gas before the gas is accreted by the protostar-disk system. Most of the magnetic flux that initially belongs to the material in the protostellar-disk system piles up around it in a growing magnetically dominated region (a smoother version of the "magnetic wall", cf. Li & McKee 1996; Tassis & Mouschovias 2005). At later times, infalling material needs to pass through this magnetically dominated region and would lose most of its angular momentum before reaching the disk. In this case, while accretion is still increasing the total mass of the protostar-disk system ($M_\star + M_\mathrm{d}$), it barely increases the angular-momentum budget of the disk. This could lead to a decrease in disk size, as demonstrated in the simulation of XK21b.

This trend of shrinking disk size should end as the envelope disperses and system transitions into Class II, and the disk size can start growing again if the disk remains mainly regulated by angular-momentum transport (e.g., gravitational or magneto-rotational instability) as opposed to angular-momentum removal (wind, magnetic braking).[3] Unfortunately, it would be difficult to test whether disk size increases during the transition into Class II with observation. While Class II disk-size estimates are available for a number of young ($\lesssim$3 Myr) star-forming regions (Eisner et al., 2018), the typical disk size can differ by a factor of a few across different star-forming regions, making it difficult to make a meaningful comparison between these populations and our sample of Orion protostellar disks.

[3]Here "disk size" refers to the gas disk size. The dust disk size generally decreases, because unlike Class 0/I disks (Appendix G), Class II disks show a significant level of radial drift of dust grains (cf. Ansdell et al., 2018) due to their longer evolution timescale and weaker dust-gas coupling.

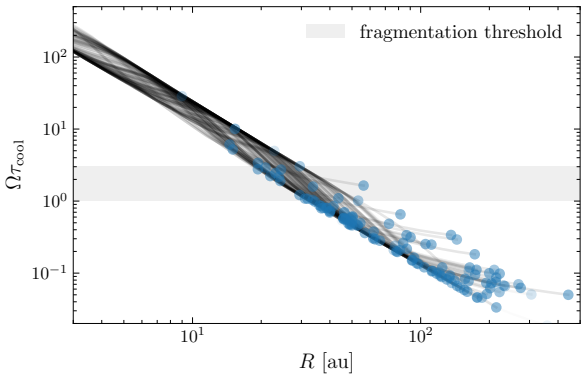

Figure 5.14: $\Omega\tau_{\rm cool}$ profiles of our best-fit models. Most systems are likely prone to fragmentation beyond 20-50 au, with the caveat that our model may not be applicable to fragmenting disks. See discussion in Section 5.4.3.

Summary: Comparing disk properties across different stages of protostellar evolution and with statistics of Class II disks in the literature, we find that the evolution is affected by several different (and often competing) mechanisms, resulting in non-monotonic evolution where the evolution of disk size and mass switches from growth to decay before finishing Class 0. Especially, the cooling of the disk during the dispersal of the envelope might play an important role in bridging protostellar disks to young protoplanetary disks, which are significantly less massive.

5.4.3 Fragmentation

One important possible outcome of GI is fragmentation, which can lead to the formation of (stellar and substellar) companions and are thought to be related to the outbursts which are common among protostellar systems (Vorobyov & Basu, 2006). Here we check whether the disks in our sample might be prone to fragmentation.

Fragmentation occurs when the overdensities produced in GI-driven perturbations cool rapidly enough to collapse under their own gravity before being disrupted by orbital shear and other perturbations. Traditionally, the condition of fragmentation is $\Omega\tau_{\rm cool} < 1-3$ (Gammie, 2001). Here the cooling timescale $\tau_{\rm cool}$ is the ratio between disk internal energy U and cooling rate $2\sigma T_{\rm eff}^4$, both of which can be computed from our model disk profile.

In Fig. 5.14 we plot the radial profile of $\Omega\tau_{\rm cool}$ for our best-fit models. $\Omega\tau_{\rm cool}$ decreases in radius, and the disks are commonly prone to fragmentation beyond 20-50 au. This is broadly consistent with the analytic prediction of Clarke (2009) and the observation of a bimodal distribution of companion separation in protostellar systems (Tobin et al., 2016, 2022) where the peak at smaller separation ($\sim$100 au) could be due to fragmentation. Note that such observation does not provide a very good constraint on whether

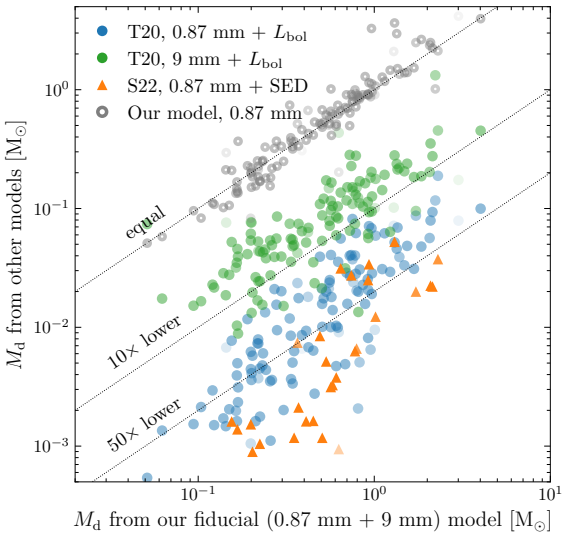

Figure 5.15: Comparison between disk mass estimates from our fiducial model (fitted with images at both wavelengths) and models based on images at a single wavelength. These include the optically thin, isothermal models from T20 for each observed wavelength (blue and green dots), the parametrized disk+envelope model coupled with radiative transfer from Sheehan et al. (2022) (orange triangles), and our model fitted with 0.87 mm images only (grey circles). The difference can be attributed to the fact that single-wavelength observation leaves a degeneracy between mass (or optical depth) and temperature, which need to be resolved by the (highly different) physical assumptions of each model (Section 5.5).

fragmentation commonly occurs as the visibility and survival rate of fragments are still not well understood.

It is worth noting that our physical picture of a gravitationally self-regulated disk cannot be directly applied to a fragmenting disk. Still, it is possible that a similar kind of self-regulation exists whereby the disk self-regulates by switching between gravitationally stable and unstable (fragmenting) states (cf. Vorobyov et al., 2013, 2019). In this case, whenever the surface density becomes high enough for disk to be unstable and fragment, this part of the disk can get rid of some mass by forming this fragment which will migrate away, and that (together with the associated angular-momentum transport and heating) can push the disk back towards a gravitationally stable state. This fragmentation self-regulation keeps the disk around a marginally stable state, with occasional fragmentation. Qualitatively, this is similar to the constraint of gravitational self-regulation used in our model. One caveat is that this argument has not been studied in sufficient detail in theory or simulation to allow reliable predictions, and there is still a lot of uncertainties regarding the interplay between fragmentation and disk evolution.

5.5 Comparison with previous studies

Our disk model produces disk masses that are significantly higher than those in T20 and Sheehan et al. (2022), as shown in Fig. 5.15. In this section we discuss why it is possible to get such drastically different

estimates from the same observational constraints, and point out that the prediction of our model is more consistent with the constraints from multi-wavelength observation.

5.5.1 Summary of models

We begin by summarizing the model used in these three studies. T20 estimated the disk mass by assuming isothermal, optically thin dust; hence the disk mass is directly porportional to the observed flux. The dust temperature was assigned following an empirical scaling law, $T_{\rm dust} = 43\,(L_{\rm bol}/{\rm L}_\odot)^{1/4}$ K. Since this estimate requires only the flux at a single wavelength, T20 obtained separate disk mass estimates for 0.87 mm and 9 mm; the two estimates differ by a factor of ~ 5.[4]

Sheehan et al. (2022) performed radiative transfer calculations on a generic, 17-parameter disk+envelope model to fit the SEDs (up to 0.87 mm) and 0.87 mm visibilities for a subset of the VANDAM Orion sample. The 9-mm data has not been used for fitting or validation. The model assumes that the grains are mainly heated externally and ignores internal accretion heating. The resulting disks are optically thin at 0.87 mm with masses slightly lower than the T20 results.

Our model fits the 0.87-mm and 9-mm images simultaneously and assumes that the disk is marginally gravitationally unstable and internally heated; external heating by protostellar irradiation is ignored (Section 5.1.1). Our model has only two free parameters to ensure that the data have more degrees of freedom than the model, and fiducial values are assigned to other relevant parameters (Section 5.1.5). We generally find the disks to be massive and optically thick. We also tried to fit our model with only 0.87-mm images and obtained similar results (Fig 5.7; also see grey circles in Fig 5.15).

5.5.2 Why can models get different results from the same observation?

Fundamentally, this is because the models in T20 and Sheehan et al. (2022) are both based on single-wavelength images, yet the disk properties are under-constrained by single-wavelength observation. Roughly speaking, the luminosity of the disk is given by

$$L_\nu \sim \kappa_{\nu,{\rm abs}} M_\nu^{\tau<1} B(T_\nu^{\tau<1}) \tag{5.12}$$

where $M_\nu^{\tau<1}$ is the mass of the visible ($\tau_\nu < 1$) portion of the disk (or $\tau_\nu\sqrt{1-\omega_\nu}$ if scattering is strong; cf. Zhu et al. 2019) and $T_\nu^{\tau<1}$ the typical temperature in this region. If we only know L_ν (from F_ν), there is

[4] T20 also adopted a much higher 9-mm opacity, which comes from assuming a mm-cm dust opacity index of ≈ 1; such small dust opacity index requires the grains to be highly porous (Woitke et al., 2016) or have size much larger than this wavelength (i.e. $\gtrsim$ a few cm; cf. Draine 2006). The difference between T20's estimates at 0.87 mm and 9 mm would be more significant (and their 9 mm estimate would be closer to our mass estimate) if T20 adopted our dust model.

still a degeneracy between mass and temperature.

Estimating disk properties requires lifting this degeneracy (or constraining how much mass is invisible, when the disk is optically thick). Each model achieves this by using its own physical assumptions (which could also involve additional observational information such as $L_{\rm bol}$ and SED) to provide additional constraints. In other words, the estimated disk properties would depend on both observational constraints and model assumptions. Therefore, it is not too surprising that models using significantly different physical assumptions produce different estimates of disk properties but all fit the observed data well.

5.5.3 Which model should I trust?

From a theoretical perspective, each of the three models have their limitations. For example, in terms of the dust heating mechanism, T20 and Sheehan et al. (2022) only consider protostellar heating while we only consider accretion heating; in reality both mechanisms might be important. As a result, it would be difficult to confidently tell *a priori* which of the models is a better approximate to reality. Meanwhile, the three models are all consistent with the observations they are based on; so we would need more stringent tests.

One ideal choice would be to test the model's predictive power, i.e. whether it can predict observables that may not be directly derived from the data fed to the model. This could help to rule out incorrect or overfitted models. The multi-wavelength observations of the VANDAM Orion survey offers an excellent opportunity to perform such test; we can check whether the model can use the 0.87-mm data to predict the 9-mm data (or the spectral index between 0.87 mm and 9 mm). The models in T20 and (Sheehan et al., 2022) both fail at this test; they produce optically thin disks with dust spectral index $\gtrsim 1$, which results in a spectral index $\gtrsim 3$ (much higher than the observed mean spectral indices of $\sim$2.2) and a systematic under-prediction of the 9-mm flux. Meanwhile, our model demonstrates good predictive power and successfully predicts the 9-mm emission when fitted with only the 0.87-mm data (Fig. 5.7).

In summary, while it is difficult to judge models based on how reasonable their physical assumptions are, current constraints from multi-wavelength observations prefer our model over existing models in T20 and Sheehan et al. (2022).

5.6 Discussion

5.6.1 Visibility of GI-induced disk substructures

In this subsection we discuss why our argument that the majority of protostellar disks may be gravitationally unstable is not invalidated by the fact that dust-continuum observations of most protostellar disks show no

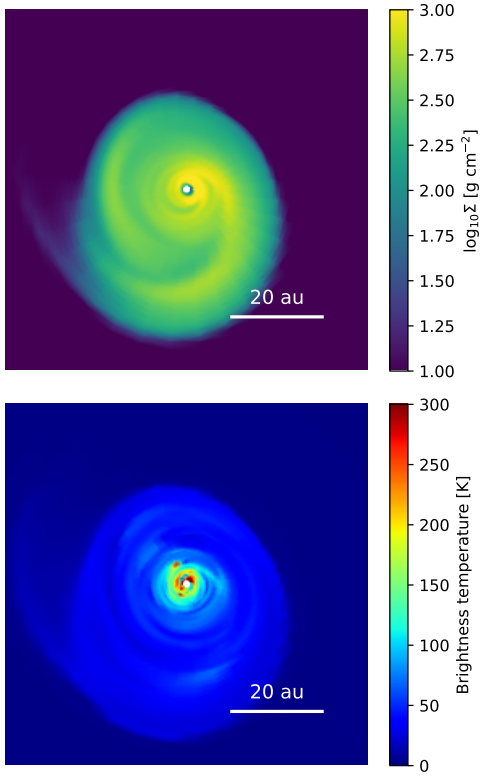

Figure 5.16: Column density (top panel) and brightness temperature at 0.87 mm (bottom panel; this is proportional to the flux density at this wavelength) from the snapshot of a simulation of protostellar disk formation. While the disk contains prominent large-scale spirals, they are barely visible at sub-mm wavelength due to high optical depth.

spiral structure. The most straightforward explanation is the lack of resolution, as the typical width of the spirals would be $\sim H$ and our current observation is nowhere close to resolving that. In addition to resolution, the detectability of spirals are also going to be limited by disk optical depth and/or detection sensitivity; here we discuss these issues in order to better plan and interpret future observations.

At sub-mm wavelengths, the high optical depth of the disk would decrease the visibility of gravitationally excited spiral structures. To provide an example, we re-ran our protostellar disk formation simulation and computed the brightness temperature (which would be directly proportional to flux density) at 0.87 mm in Fig. 5.16. Here our setup is identical except that our azimuthal (ϕ) domain covers the whole 2π. The disk is gravitationally unstable and shows prominent spirals with order-unity amplitude in the column-density profile. However, since only the surface of the disk is visible, the observed flux is not directly proportional to the column density. As demonstrated in Fig. 5.16, the large-scale spirals are barely visible and we only see less coherent, small-scale perturbations at the surface of the disk, which are likely associated with shocks in the cascade of turbulent perturbations driven by the large-scale spirals. Meanwhile, gravitationally unstable protoplanetary (Class II) disks generally suffer less from this problem, as they are cooler and have lower column density (e.g., Rowther et al., 2020).

The problem of high optical depth is less severe for wavelengths $\gtrsim$ cm, as typical protostellar disks will be optically thin at these longer wavelengths (Fig. 5.12). However, at such long wavelengths, detection sensitivity becomes the bottleneck. For example, for the 9-mm observations in T20, only a small (and often unresolved) central portion of the disk is above the detection limit. In the future, observations with higher sensitivity and resolution at $\sim$ cm wavelength might be a direct probe for spiral structures in protostellar disks.

5.6.2 Is the 9-mm flux contaminated by free-free emission?

When fitting our model to observation we have assumed that the fluxes at both observed wavelengths are dominated by dust thermal emission. While this assumption is reasonable at 0.87 mm, the 9-mm flux may be subject to contamination from free-free emission coming from ionized gas in the outflow. Tychoniec et al. (2018) studied radio emission in Perseus and concluded that $\sim$60% of the 9-mm flux there could be due to free-free emission, based on an extrapolation of (free-free dominated) fluxes at 4.1 cm and 6.4 cm. Here we evaluate whether our sample of Orion protostars also suffer from a high level of free-free contamination at 9 mm.

We begin with a rough estimate using results from Tychoniec et al. (2018), which fit an empirical relation between $L_{\rm bol}$ and 4.1-cm luminosity, and find that the typical spectral index of free-free emission (between

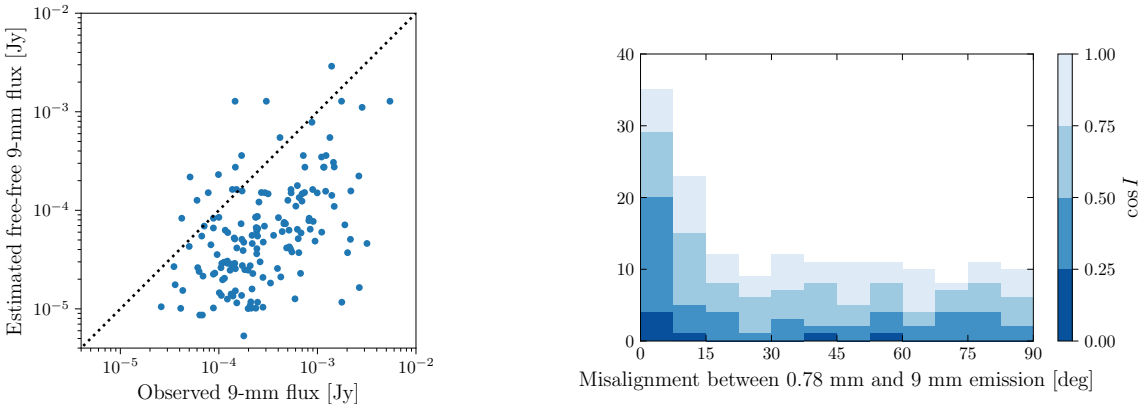

Figure 5.17: Top: contribution from free-free emission to 9-mm flux estimated with empirical scaling relations in Tychoniec et al. (2018). Bottom: the distribution of misalignment between orientations of 0.87-mm and 9-mm emission from PA estimates in T20, showing a peak for aligned emission and a large base of unresolved systems. Both results suggest that free-free emission in the outflow is unlikely to dominate the observed 9-mm flux.

4.1 cm and 6.4 cm) is 0.3–0.4. Using this empirical relation and assuming a free-free spectral index of 0.4 between 9 mm and 4.1 cm, we can estimate the free-free luminosity at 9 mm from $L_{\rm bol}$. This estimated 9-mm free-free emission is generally lower than the observed 9-mm emission, with median $F_{\rm 9mm}^{\rm free-free}/F_{\rm 9mm} \sim 0.2$ (Fig. 5.17, top panel). Therefore, the observed 9-mm flux should still be dominated by dust thermal emission.

However, this conclusion is subject to some uncertainty given the spread in the estimated 9-mm flux. It is also unclear whether the empirical scaling laws in Perseus would also be applicable to Orion. To address these caveats, we check another diagnostic which is less quantitative but more robust: the misalignment between the orientation (PA) of 0.87-mm and 9-mm emissions. When the source is resolved, the 9-mm emission should be aligned with the 0.87-mm emission if it is dominated by dust in the disk, and approximately perpendicular if it is dominated by free-free emission in the outflow. In the bottom panel of Fig. 5.17 we plot the distribution of the estimated misalignment; the distribution peaks at 0° and does not show any peak around 90°, consistent with disk-dominated 9-mm emission. One caveat, however, is that most systems in our sample are unresolved or barely resolved at 9 mm; for these systems there are no accurate estimates of orientation, which results in a large "base" of random estimated misalignment in the distribution, and we cannot directly confirm whether they have disk-dominated emission. This problem can be somewhat alleviated by looking at subsamples of low-inclination systems; for example, for systems with $\cos I \leq 0.5$ (0.25), the fraction of aligned systems (ΔPA $< 15°$) increases to 47% (71%).

In summary, current data prefer disk-dominated emission at 9-mm, yet future observations at longer wavelengths and/or higher resolution are needed to constrain more reliably the level of free-free contamination.

5.6.3 Uncertainties in the physical picture and other scenarios of disk formation

The physical picture of protostellar disk formation can be diverse. While in this chapter we demonstrate that the observed Class 0/I disks are consistent with gravitational self-regulation at the population level, this does not rule out other scenarios of disk formation. Here we discuss the main (theoretical) uncertainties in our assumed physical picture (gravitational self-regulation), and how they result in other possible scenarios of disk formation.

One major uncertainty in our assumption of gravitational self-regulation is whether the disk is in a quasi-steady state in which accretion from the envelope onto the disk is approximately balanced by angular-momentum transport within the disk. For example, if the initial condition of the pre-stellar core and subsequent envelope evolution are highly turbulent, the specific angular momentum of the material being accreted onto the disk might undergo large variations, and the disk evolution might be driven mainly by such variations as opposed to angular-momentum transport within the disk. In the most extreme case, the whole disk might just be formed by a small jet of high-angular-momentum material; this is one possible explanation for the small number of systems in the VANDAM Orion survey that show prominent rings (Sheehan et al., 2020). From a theoretical point of view, it is unclear whether the envelope should be highly turbulent and the accretion onto the disk highly variable. Some molecular-cloud-scale simulations support this possibility (e.g., Kuffmeier et al., 2017; Kuznetsova et al., 2020), with the caveat that the adoption of ideal MHD in these studies could overestimate the level of turbulence at and below core/envelope scale.

Another uncertainty in our argument is whether ambipolar diffusion inside the disk is strong enough to render magnetically driven angular-momentum transport/removal negligible. This could depend on both initial conditions and disk chemistry, which are subject to large uncertainties. Especially, the abundance of small grains plays an important role in determining the strength of non-ideal MHD effects (Zhao et al., 2018a). If the magnetic field is better coupled to the gas, disk evolution (and accretion onto the protostar) could be driven mainly by magnetic braking and wind launching and the disk could be gravitationally stable and less turbulent. As a side note, it is also possible that this kind of magnetic regulation co-exists with gravitational self-regulation in some disks (either in the same region as in Fromang et al. 2004 and Riols et al. 2021, or in different parts of the disk as in Zhu et al. 2010), especially during later times (e.g., Flat Spectrum) when the accretion rate drops.

Chapter 6

Summary and outlook

6.1 Summary

6.1.1 Simulating disk formation

We use a radiative 3D non-ideal MHD simulation that includes ambipolar diffusion and Ohmic dissipation (whose strengths are computed from an equilibrium chemical network) and a novel model of radiative cooling to follow the evolution of a pre-stellar core until ∼10 kyr after protostar formation (Chapter 2). We observe the formation of a massive, gravitationally unstable protostellar disk that grows to ∼30 au in radius (Section 3.2).

Our simulation realistically models most physical ingredients that may impact disk formation, to a level comparable to state-of-the-art simulations in the literature (Sections 2.2 and 2.4.3). Additionally, our simulation adopts a more realistic initial condition (Sections 2.1 and 2.4.2) and achieves a better balance between numerical convergence and long-term integration (Sections 2.3 and 2.4.4) than most previous studies. We demonstrate that our simulation achieves relatively good numerical convergence despite its modest resolution (which is required for long-term integration), and argue that at current level of resolution convergence requires an azimuthally symmetric grid and special treatment of the inner boundary condition to avoid unphysical angular momentum loss (Section 2.4.4; also see Appendices B and D). For the spherical-polar grid we adopted, correcting the geometric error of EMF in the polar region is also important (Appendix C.7).

6.1.2 The physical picture of disk formation and evolution

With the help of our simulation results, we obtain the following insights regarding the physical picture of Class 0/I disk evolution:

- The mass and angular-momentum budget of the disk is determined mainly by accretion through the pseudodisk (with angular momentum budget affected by magnetic braking in the pseudodisk during the infall), and that mass and angular-momentum removal from the protostar-disk system by outflow and disk magnetic braking are negligible (Section 3.3.1). The redistribution of angular momentum inside the disk is facilitated mainly by GI and the spiral waves it excites (Section 3.3.2); this causes the disk size to exceed greatly the circularization radius of pseudodisk material (Sections 3.2.2). We also argue that the gravitational self-regulation of accretion and the redistribution of disk angular-momentum should eventually leave most of the disk marginally unstable, with a Toomre Q of order unity (Sections 3.3.2 and 3.3.3).

- The temperature profile of the disk is determined primarily by an equilibrium between heating and radiative cooling. While heating can come from a number of different sources, the total heating per radius can be well approximated (to within a factor of a few) by the accretion energy release $-g_R\dot{M}_{\rm tot}$ (Section 3.4.1). Unlike (flared) Class II disks, the disk generally have aspect ratio H/R increasing towards smaller radii, therefore (direct) protostellar irradiation is no an important heating source for most of the disk. The relation between the cooling rate and disk temperature is well approximated by assuming a non-turbulent disk cooled by radiation (Section 3.4.2). The convection driven by the gravitationally driven spirals, although present, is not strong enough to significantly increase the efficiency of cooling. We provide an analytic estimate of the strength of spiral-driven convection and show that it should remain unimportant for all Class 0/I disks (Section 3.4.3).

- Inside the disk, strong non-ideal magnetic diffusion (mainly ambipolar) and a slow accretion timescale cause the magnetic field to decouple fully in the radial direction, producing a nearly uniform magnetic field inside the disk whose strength approximately matches the field strength in the pseudodisk near disk edge (Section 3.5.1). While the magnetic field remains relatively well coupled to the gas in the azimuthal direction, disk magnetic braking remains weak and barely affects disk evolution. We propose an analytic estimate of the disk magnetic braking rate, which is consistent with the simulation results (Section 3.5.2). Overall, the magnetic field is dynamically unimportant inside the disk. On the other hand, magnetic braking in the pseudodisk becomes increasingly important at later times, which causes the disk angular momentum budget to saturate (Section 3.3.1).

6.1.3 Modeling protostellar disk with a semi-analytic model

The physical insights discussed above lead us to a surprisingly simple quantitative physical picture of Class 0/I disk evolution, and allows the construction of a 1D semi-analytic model (Section 4.1) based on self-

regulated GI, thermal balance, and magnetic decoupling. The model predicts the radial profiles of disk column density, mean disk temperature, and disk magnetic-field strength, provided inputs such as the total mass and angular momentum of the protostar-disk system.

One application of this semi-analytic model is predicting the long-term evolution of the disk. For this purpose, the inputs to the model can be obtained from a physics-rich yet inexpensive simulation of pseudodisk evolution, which uses a larger inner boundary and does not resolve the disk (Section 4.2). We argue that such a 1D+pseudodisk model can produce reasonable predictions because the evolution of the system is largely hierarchical: the small-scale (disk) dynamics do not greatly affect the large-scale (pseudodisk) evolution (Section 4.2.1). We verify this idea of hierarchical evolution and show that our 1D+pseudodisk model agrees reasonably well with the 3D simulation, and the long-term evolution of disk size predicted using the 1D+pseudodisk model (which saturates at $\sim$30 au and eventually shrinks) is consistent with recent observations (Section 4.3). Such a hierarchical modelling of disk formation and evolution circumvents the numerical difficulty of tracing the long-term evolution (e.g., throughout Class 0/I) with well-resolved (or at least numerically convergent) direct simulation.

6.1.4 Observed properties of protostellar disks

Our 1D semi-analytic model can be extended into a 2D model (which includes the vertical temperature profile) that can be used for generating dust-continuum mock observations (Section 5.1). Compared to existing protostellar disk models in the literature, the physical assumptions on gravitational self-regulation and disk thermal budget in our model reduce the number of free parameters without making arbitrary assumptions. Additionally, our model can produce reliable disk mass estimates even when the disk is optically thick. We fit this model to observational data to test our physical assumptions and to infer disk properties from multi-wavelength dust-continuum observations (Section 5.2).

We find that our model fits relatively well to the majority of the protostellar disks in the VANDAM Orion survey (T20; Section 5.3). Moreover, the observations prefer our fiducial model with marginally unstable Toomre Q compared to models with larger Q values, suggesting that the assumption of gravitational self-regulation is likely valid at the population level. (This, however, does not rule out other scenarios of disk formation; cf. Section 5.6.3)

Using our model, we produce new estimates of Orion protostellar disk properties (Section 5.4). Our main findings include:

- Disks are significantly more massive than previously expected, with typical disk-to-star mass ratio $M_{\rm d}/M_\star = \mathcal{O}(1)$. The high optical depth at $\lesssim$ mm wavelengths could have caused a systematic under-

estimation of disk mass in previous studies (e.g., T20).

- Both disk mass and disk size decrease towards later stages of protostellar evolution. These trends might be associated with the decrease in accretion rate and the increased magnetic braking in the inner pseudodisk at later times. In general, the evolution of disk properties throughout its lifetime is determined by a competition of several different mechanisms and can be non-monotonic.

- Our model suggests that most disks in our sample are likely prone to fragmentation beyond 20-50 au, with the caveat that the applicability of our model to fragmenting disks cannot be guaranteed.

One limitation of our model is that its estimates are subject to large uncertainties, mainly due to the order-magnitude systematic uncertainties in assumed model parameters (Section 5.2.3). While the qualitative trends we reported, including the $\mathcal{O}(1)$ disk-to-star mass ratio, are robust against these uncertainties, the estimates for key disk properties are subject to uncertainties of a factor of a few. In the future, these uncertainties may be reduced if observations provide additional constraints on protostar mass, accretion rate, or grain size distribution.

6.2 Open questions and possible future works

In this section we discuss some open questions in the physical picture of protostellar disk formation and evolution that are yet to be addressed by this thesis, and outline several possible future works. We begin by discussing open questions related to the conclusions of this thesis in Sections 6.2.1 – 6.2.4. These open questions are grouped by their relevant spatial scales, as problems at the same spatial scale can often be investigated within the same simulation (or a set of simulations with similar setups). We then discuss possible future applications of our semi-analytic disk model on producing theoretical predictions and observationsal constrains on grain evolution (Section 6.2.5), which could have important implications for planet formation.

The goal of this section is not to list all open problems related to protostellar disks, but to highlight those that appear to be steps away from being addressed (in terms of code development and/or analytic calculation) given the physical picture and numerical setup presented in this thesis. The number of possible future works discussed here is likely more than what can be accomplished within a few years, and they will be prioritized (and deprioritized) based on the availability of collaboration, the progress on code development, and the emergence of new ideas on analytic modeling of key physical mechanisms.

6.2.1 Cloud/core scale ($\gtrsim 0.1$ pc): whether cores are isolated, and how the main accretion phase ends

One major limitation of the type of simulation presented in this thesis is that the initial condition is an isolated pre-stellar core while in reality the core may and may not be sufficiently isolated for this to be a good approximation. Recent simulations (Kuffmeier et al., 2017; Kuznetsova et al., 2020) find that in ideal MHD the accretion of mass and angular momentum onto pre-stellar and protostellar cores can be highly episodic, suggesting that the evolution of a core (and the protostar-disk system formed within it) should not be studied in isolation to the turbulent, filimentary molecular cloud at larger scales. A possible future project is to include non-ideal MHD effects (mainly ambipolar diffusion) in this type of simulation. The simple analytic estimate in Eq. (2.4) suggests that non-ideal MHD effects (mainly ambipolar diffusion) can become important at and below core scale (~ 0.1 pc), and it would be necessary to including non-ideal MHD effects in this type of simulation to test whether ambipolar diffusion creates quiescent, isolated cores (cf. Barranco & Goodman, 1998) and to decide whether the type of pre-stellar core and pseudodisk infall discussed in this thesis is relevant to real protostellar systems. We also comment that there is relatively little technical difficulty in performing this kind of simulation (other than the relatively high computational cost associated with resolving core evolution in a non-ideal MHD simulation), and similar simulations have been performed, for example, by Chen & Ostriker (2014). The focus of this study would be to perform careful analyses on the simulation from the perspective of core mass and angular-momentum accretion history to decide whether the interaction between the core and the environment is negligible from the beginning of dynamical contraction to the end of the main accretion phase, and if not, how such interaction should be modeled in the context of disk formation.

A related question that could be (partially) addressed by this kind of simulation is how the main accretion phase ends. For the type of initial condition adopted in this thesis, the envelope around the central part of the pre-stellar core will continue to fall in until most mass in the domain (much more than the core mass) has been accreted. There are a number of possibilities of how this can be prevented, such as support provided by turbulence or magnetic field, competitive accretion, and feedback from nearby massive stars. The effect of some of these mechanisms can be quantified in the cloud-scale simulation discussed above. From the perspective of protostellar disk evolution, this allows one to determine when and how the accretion rates of mass and angular momentum should drop at later times, which are important for modeling the transition from the gravitationally self-regulated main accretion phase to the Class II phase where the disk can have different radial profiles and transport mechanisms.

6.2.2 Envelope/pseudodisk scale ($100\ \mathrm{au} \lesssim r \lesssim 0.1\ \mathrm{pc}$): quantitative and predictive modeling of magnetic braking and the Hall effect

Magnetic braking in the pseudodisk directly affects the angular momentum budget of the protostellar disk. Supposing the assumption of an isolated, quiescent core is valid, at least to the extent that the type of quiescent pseudodisk in our simulation is a reasonable approximation to reality, it might be possible to build a predictive model of pseudodisk magnetic braking. This would be similar to our modeling of disk magnetic braking in Section 3.5.2, although our model cannot be directly generalized to the pseudodisk as our estimate of vertical differential rotation is valid only in the rotationally supported disk. If such a model can be developed, it could be used together with a piecewise self-similar models of other pseudodisk properties (e.g., Dapp et al. 2012) and our 1D semi-analytic disk model to predict disk evolution directly from the initial conditions of the pre-stellar core. Together with the cloud-scale simulation discussed above, this allows a first-principle population synthesis of protostellar disks. Similar population synthesis has been performed with direct simulation Bate (e.g., 2018), but the reliability of direct simulations are limited by incomplete physical ingredients insufficient resolution.

A related topic of interest is the role of the Hall effect in shaping the angular-momentum budget of the disk. This has been studied in a qualitative fashion in a number of previous simulations (cf. Section 2.2.1), yet we do not have a quantitative and predictive model for how much rotation can be produced by the Hall effect. One interesting limit to look at could be when magnetic braking in the inner pseudodisk is strong (which should be the case at sufficiently late time, cf. Fig. 3.10). In that case the rotation will eventually become the azimuthal Hall drift velocity, which is $\sim (\eta_{\rm H}/\eta_{\rm A}) v_{\rm K}$ given that the pseudodisk is in near free fall and the field is radially marginally decoupled. That implies circularization when $\eta_{\rm H}$ becomes $\gtrsim \eta_{\rm A}$, which could occur at several 10 au – 100 au (cf. Fig. 3.4). Of course, this is just a very hand-waving argument; we might be able to refine and test it once the Hall effect is included in our simulations.

6.2.3 Outer disk (few $\mathrm{au} \lesssim r \lesssim 100\ \mathrm{au}$): magnetic braking, outflow, and fragmentation

In this subsection we discuss two open questions regarding the dynamics of the outer disk which, although important, are probably not going to be prioritized by us in the near future due to technical challenges in performing reliable simulations.

In this thesis we conclude that the protostellar disk is gravitationally self-regulated because other angular-momentum transport/removal mechanisms, mainly magnetic braking and outflow launching, are inefficient, and we simply ignore the contribution of these mechanisms when predicting disk properties and evolution.

However, a more careful parameter survey is still required to determine whether this remains true for most of the parameter space of pre-stellar core initial condition and dust grain size distribution. Another related question is whether one can quantitatively predict how much outflow is being launched and how outflow propagates to larger distances. Such predictions can be directly compared with observations of outflow velocity and morphology. A major challenge here is to perform reliable simulations of outflow launching and propagation that can be used to inspire and test such models, as the simulation results might be sensitive to resolution, choice of density floor, and ionization of the outflow by protostellar irradiation (which may also require non-equilibrium chemistry).

Another caveat in our model is that it assumes a non-fragmenting disk, while our disk property estimates for observed protostellar disks suggest that the outer part of the disk might be commonly subject to fragmentation (Section 5.4.3). In order to improve our predictive model of protostellar disk evolution, one needs to understand how fragmentation regulates disk evolution by transporting mass and angular momentum. This is a topic that has not received much attention (despite many previous studies on disk fragmentation) mainly because properly simulating fragmentation requires high resolution (and small timestep) while studying the feedback of fragmentation on disk evolution requires long-term ($\sim$ disk evolution timescale and $\gg$ orbital timescale) integration; satisfying both requirements is highly challenging, even in 2D (e.g., Vorobyov et al., 2019).

6.2.4 Inner disk ($\lesssim$ few au): disk truncation, magnetic flux problem, and variability

Finally, we consider the smallest scale in our problem, the inner disk within a few au. The main question to investigate here is how and where the disk transitions from being gravitationally regulated to being dominated by other angular-momentum transport/removal mechanism. Understanding this question allows us to improve our semi-analytic disk model for applications discussed Chapter 5 and Section 6.2.5. Furthermore, it is directly related to the magnetic-flux problem of the protostar and might help explain some types of variability observed in protostellar systems.

The revival of magnetic transport: what happens beyond grain sublimation?

An intuitive idea is that the gravitationally self-regulated disk should become thermally ionized at sufficiently small radii, which then causes the disk column density to drop due to efficient angular-momentum transport by magnetic mechanisms such as MRI and magnetic braking. (This is reminiscent of the edge of the MRI "dead zone" in Class II disks.) However, estimating where such transition occurs could be tricky (and this

transition might not be stable; cf. Zhu et al. 2010). In Section 5.1.4 we assume that after the sublimation of dust grains ($T_{\rm subl} \approx 1200$ K) cooling becomes inefficient and the disk becomes thermally ionized. However, in reality the finite gas opacity might instead make the disk cool more efficiently (by making it less optically thick while still having $\tau_{\rm R,mid} \gtrsim 1$) once the temperature goes beyond $T_{\rm subl}$, thereby locking the disk (midplane) temperature around $T_{\rm subl}$ until much smaller radii. What further complicates the problem is the contribution from protostellar irradiation. We have demonstrated that, before grain sublimation, H/R decreases in radius thus the disk is self-shielded. That is no longer the case once the disk reaches $T_{\rm subl}$; now the midplane temperature is approximately fixed and $H/R \propto R^{1/2}$. Thus the disk might have a hotter, thermally ionized surface, and that may lead to stronger magnetic transport.

We also comment that the transition from decoupled to coupled magnetic field also affects magnetic flux budget of the protostar. It is a classic problem in star formation that the star contains many orders of magnitude less magnetic flux than the pre-stellar core from which it was born (Babcock & Cowling, 1953). Qualitatively, this can be resolved at least partially by ambipolar diffusion, which in the disk (and the inner pseudodisk) decouples the field from the gas to reduce the mass-to-flux ratio (e.g., Desch & Mouschovias, 2001; Kunz & Mouschovias, 2010; Dapp et al., 2012). However, this argument alone cannot predict the stellar magnetic flux budget quantitatively, as the mass-to-flux ratio in the disk is non-uniform and increases towards smaller radii as long as the field remains decoupled. Therefore, in order to make a more quantitative prediction for how much flux is accreted to the star (or set an upper limit for it, suppose there might be other magnetic flux transport mechanisms at smaller radii), one needs to know when the field becomes "recoupled" to the gas.

Tidal truncation by the star-disk "binary"

Another possibility is that all the complexities discuss above are not relevant at all because the disk is tidally truncated at a larger radii. A gravitationally self-regulated disk is expected to be massive and can host asymmetric density perturbations with order-unity amplitude (e.g., odd-m spirals), thus the center of mass of the disk and the protostar-disk system do not have to coincide with the protostar. This might cause the star and the disk center of mass to orbit around each other (although not exactly like two point masses due to the finite extent of the disk) and creates a tidal potential that resembles that of a binary. This could truncate the disk at a scale comparable to the "binary" separation and excite spiral waves at the inner disk that could replace GI to become the main angular-momentum transport mechanism. In our fiducial simulation this mechanism is artificially suppressed by our choice of a domain with $m = 2$ symmetry; but in Appendix D.2 we discuss a simulation which covers full 2π in ϕ that seems to be affected by this effective "binary" potential (with the caveat that the treatment of accretor in this simulation is not fully

self-consistent).

Variability of a gravitationally self-regulated disk

Another interesting problem regarding the inner disk is how much variability should be present in the inner part of a gravitationally self-regulated disk. More precisely, we want to investigate whether (and how) the disk oscillates between being stable and unstable. This is motivated by the observation that for a marginally unstable disk, the (time-averaged) effective α should decrease towards smaller radii; eventually, the effective α might be so small that the disk has to transport angular momentum in a "quantized" fashion, where short period of gravitational instability is followed by much longer periods of stability. If this conjecture is true, it might be used to explain some types of protostar variability (cf. Fischer et al. 2022).

Investigating these questions with simulations

Our current simulation framework can be used for studying (some or all of) these questions with relative few modifications. The basic methodology would be to take a snapshot from a global simulation and restart with smaller inner boundary (and, optionally, higher resolution) and simulate for a period of time until a quasi-steady state is reached at small radii. In terms of physical ingredients and numerical treatments, modifications on chemistry and opacity to support higher temperatures discussed in Chapter 2 has already been implemented; we could also add a simple model of ray-tracing along radial rays to capture thermal ionization by protostellar irradiation at small radii. It would probably be more challenging to properly study the tidal truncation by an effective "binary" potential, as that requires moving the accretor away from the origin (or moving the whole domain along with the accretor). Another caveat is that this type of simulation might not have sufficient resolution to capture transport by MRI or other small-scale instabilities quantitatively if they do occur. Still, this should not undermine the value of such studies, given that the inner edge of a gravitationally unstable disk is relatively unexplored territory.

6.2.5 Application of semi-analytic disk model to planet formation: theoretical predictions and observational constraints on grain evolution

Grain growth in Class 0/I is becoming an increasingly important problem in planet formation. Although the timescale of the Class 0/I phase is just a fraction of the Class II phase, it is possible that a significant amount of grain growth (and even planetesimal formation) occurs in Class 0/I, given the higher disk mass and the very different disk environment (which may and may not be more suitable for grain growth and planetesimal formation). Existing studies that attempt to predict grain growth (and grain-gas decoupling) in protostellar

disks with direct simulation such as Vorobyov et al. (2019) and Lebreuilly et al. (2020) are limited by low resolution and oversimplified dust evolution model. These limitations mainly come from the expense of directly simulating disk evolution, which can be circumvented using our semi-analytic disk model. We can use our model's prediction of disk properties (including the level of gravitoturbulence) as the input to a more realistic dust evolution model (e.g., `DustPy` by Stammler & Birnstiel 2022) to predict the long-term evolution of grains. This will allow us to determine whether a nontrivial fraction of dust grains can be converted to large grains (Stokes number $\gg 1$) and planetesimals that can survive into Class II even though most of the mass (gas, and the well-coupled small grains) in the Class 0/I disk will be accreted by the protostar before the system enters Class II. In other words, we can determine whether Class II disks may host a population of large grains and planetesimals inherited from Class 0/I. (Such a population would be hard to identify observationlly, given than they barely contribute to opacities at sub-mm and mm wavelengths where disks are commonly observed.)

Our semi-analytic model can also be used to provide new observational constraints on the typical grain size in protostellar disks. As we argued in this thesis, protostellar disks are often optically thick at sub-mm wavelengths, and this makes it difficult to infer the dust opacity index (which can constrain the grain size distribution) directly from the observed spectral index. However, the model presented in Chapter 5 might help resolve this problem. In principle, if we include the grain size as a free parameter in our model, it can be used for extracting grain size constraints from multi-wavelength observations. For example, in an ongoing collaboration with H. B. Liu et al., we fit resolved observations of TMC-1A at four different wavelengths to our model and preliminary results suggest that the grain size can be constrained to within a factor of 2.

Overall, our semi-analytic model is a convenient tool for generating protostellar disk properties which can be used to both theoretical and observational studies. Just like the commonly adopted α-disk model, it is probably not going to capture all the details of the disk accurately, yet its simplicity and predictive (or generative) power makes it a good baseline. We hope that this model will benefit studies of planet formation in Class 0/I similar to how α-disk models have benefited such studies in protoplanetary disks.

Appendix A

Chemical network and reaction rates

A.1 Neutral collision timescales

Following Desch & Mouschovias (2001, and references therein), the collison timescales between charged species and neutrals are given by

$$\tau_{sn} = \frac{1}{a_{s\mathrm{He}}} \frac{m_s + m_{\mathrm{H_2}}}{m_{\mathrm{H_2}}} \frac{1}{n_{\mathrm{H_2}} \langle \sigma_{\mathrm{coll}} w \rangle_{s\mathrm{H_2}}}. \tag{A.1}$$

Here $a_{s\mathrm{He}}$ accounts for collisions with helium atoms, and is equal to 1.14 for ions, 1.16 for electrons, and 1.28 for grains. $n_{\mathrm{H_2}} = 0.85 n_{\mathrm{n}}$ is the H_2 number density (which comes from assuming a mass composition of 75% H_2 and 25% He). The rate constant $\langle \sigma_{\mathrm{coll}} w \rangle_{s\mathrm{H_2}}$ is $1.69 \times 10^{-9} \mathrm{cm^3 s^{-1}}$ for ions, $1.3 \times 10^{-9} \mathrm{cm^3 s^{-1}}$ for electrons, and $\pi a^2 / (8 k_{\mathrm{B}} T / \pi m_{\mathrm{H_2}})^{1/2}$ for grains with radius a.

A.2 Species and reactions at low temperature

Our chemical network is largely similar to that in Kunz & Mouschovias (2009). It includes the following species: electrons, one gas-phase atomic metal A (for Na, Mg, K, etc.), one gas-phase molecule M, and a family of neutral and singly charged grains g with different radii (a). We choose $m_{\mathrm{A}} = 23.5 m_{\mathrm{p}}$ (a value between Na and Mg), $n_{\mathrm{A}} = 2.05 \times 10^{-6} n_{\mathrm{n}}$, and $m_{\mathrm{M}} = 29 m_{\mathrm{p}}$ (corresponding to HCO).

The reactions we consider in this thesis are listed in the top (low-temperature) part of Table A.1. For simplicity we do not include grain coagulation or fragmentation; the only reaction between grains are charge

exchanges. For this thesis we ignore CR shielding and the rate coefficient for CR ionization is simply

$$\zeta = \zeta_{\rm CR} = 1 \times 10^{-17} s^{-1}, \tag{A.2}$$

but see Appendix A.4 for the more general expression of ζ. The rate coefficients for electrons and ions attachment onto grains are given by

$$\alpha_{s\mathrm{g}^0} = \pi a^2 \left(\frac{8k_{\rm B}T}{\pi m_s}\right)^{1/2} \left[1 + \left(\frac{\pi e^2}{2ak_{\rm B}T}\right)^{1/2}\right] \mathcal{P}_s, \tag{A.3}$$

$$\alpha_{s\mathrm{g}^\pm} = \pi a^2 \left(\frac{8k_{\rm B}T}{\pi m_s}\right)^{1/2} \left[1 + \left(\frac{e^2}{ak_{\rm B}T}\right)\right] \left[1 + \left(\frac{2}{2 + ak_{\rm B}T/e^2}\right)^{1/2}\right] \mathcal{P}_s, \tag{A.4}$$

where s can be i or e, with the grain charge $\alpha_{s\mathrm{g}^\pm}$ chosen accordingly. $\mathcal{P}_s$ denotes the sticking probability in collision, and we take $\mathcal{P}_{\rm i} = 1, \mathcal{P}_{\rm e} = 0.6$. Similarly, the grain-grain rate coefficients are

$$\alpha_{\mathrm{g}^+_{a_1}\mathrm{g}^-_{a_2}} = \pi a_{\rm sum}^2 \left(\frac{8k_{\rm B}T}{\pi m_{\rm red}}\right)^{1/2} \left[1 + \left(\frac{e^2}{a_{\rm sum}k_{\rm B}T}\right)\right] \left[1 + \left(\frac{2}{2 + a_{\rm sum}k_{\rm B}T/e^2}\right)^{1/2}\right], \tag{A.5}$$

$$\alpha_{\mathrm{g}^\pm_{a_1}\mathrm{g}^0_{a_2}} = \pi a_{\rm sum}^2 \left(\frac{8k_{\rm B}T}{\pi m_{\rm red}}\right)^{1/2} \left[1 + \left(\frac{\pi e^2}{2a_{\rm sum}k_{\rm B}T}\right)^{1/2}\right] \mathcal{P}_{12}, \tag{A.6}$$

where $a_{\rm sum} = a_1 + a_2$, $m_{\rm red} = m_1 m_2/(m_1 + m_2)$, and the charge exchange probability $\mathcal{P}_{12} = a_2^2/(a_1^2 + a_2^2)$. This expression of $\mathcal{P}_{12}$ assumes that after a collision, the probability for a charge to end on a grain is proportional to its surface area. (Kunz & Mouschovias 2009 and Desch & Mouschovias 2001 both had typos in the expression for $\mathcal{P}_{12}$.)

A.3 Reactions at high temperature

When the simulation in this thesis was developed, we believed that the disk would not be too hot (i.e. at most a few 100 K). That turns out to be wrong, and if one wishes to model chemistry realistically in the innermost $\sim$ au (few au in some cases) it might be necessary to include a few reactions that becomes important at higher temperatures. The first of these is thermal ionization of potassium; although not very abundant (compared to, for example, Na), its lower ionization threshold makes it the dominant source of thermal ionization below ~ 3000 K (cf. Kunz & Mouschovias 2009, before Eq. 15), with a ionization rate

$$\frac{{\rm d}n_{\mathrm{K}^+}}{{\rm d}t} = 4.1 \times 10^{-15} n_{\mathrm{H}_2} n_{\mathrm{K}^0} \left(\frac{T}{1000\ {\rm K}}\right)^{1/2} \exp\left(-\frac{5.04 \times 10^4\ {\rm K}}{T}\right)\ {\rm cm^{-3}s^{-1}}. \tag{A.7}$$

Effective reaction	Rate coefficient	Comment
Low temperature reactions (currently included)		
$H_2 \rightarrow (...) \rightarrow M^+ + e^-$	ζ	CR ionization
$M^+ + e^- \rightarrow (M_1^0 + M_2^0)$	$2 \times 10^{-7}\ T_{300}^{0.75}$	dissociative recombination
$A^+ + e^- \rightarrow A^0$	$2.8 \times 10^{-12}\ T_{300}^{0.86}$	radiative recombination
$A^0 + M^+ \rightarrow A^+\ (+ M^0)$	2.5×10^{-9}	charge transfer
$e^- + g^0 \rightarrow g^-$	α_{eg^0}	electron attachment onto grains
$e^- + g^+ \rightarrow g^0$	α_{eg^+}	electron attachment onto grains
$i^+ + g^0 \rightarrow i^0 + g^+$	α_{ig^0}	ion attachment onto grains
$i^+ + g^- \rightarrow i^0 + g^0$	α_{ig^-}	ion attachment onto grains
$g_{a_1}^+ + g_{a_2}^- \rightarrow g_{a_1}^0 + g_{a_2}^0$	$\alpha_{g_{a_1}^+ g_{a_2}^-}$	grain-grain collision
$g_{a_1}^\pm + g_{a_2}^0 \rightarrow g_{a_1}^0 + g_{a_2}^\pm$	$\alpha_{g_{a_1}^\pm g_{a_2}^0}$	grain-grain collision
High temperature reactions (to be included in future studies)		
$A^0 \rightarrow A^+ + e^-$	α_K	potassium thermal ionization
$g^0 \rightarrow g^+ + e^-$	α_{te0}	grain thermionic emission
$g^- \rightarrow g^0 + e^-$	α_{te-}	grain thermionic emission

Table A.1: List of effective reactions in our chemical network; the low-temperature portion corresponds to the reactions in Kunz & Mouschovias (2009) Table 1. A stands for atomic species that can be ionized, M for molecular species that can be ionized, and i for any ion (A or M). Neutral species not traced by the network are bracketed. In the first reaction, we have skipped a few intermediate reactions that happen quickly, and ignored neutral reactants and products for these intermediate reactions. (See Kunz & Mouschovias 2009 Table 1 for the intermediate steps for HCO^+ production; for a longer list of possible reactions that eventually lead to moelcular ion production, see Marchand et al. 2016 Table A1.) The reaction rate is rate coefficient multiplied by the number density of all LHS species. Numeric values for rate coefficients are given in cgs unit with $T_{300} = T/(300\ \mathrm{K})$. Other rate coefficients are defined in the text.

Since our chemical network does not trace potassium separately, the effective reaction corresponding to this is $A^0 \rightarrow A^+ + e^-$ with rate coefficient

$$\alpha_{\rm K} = 4.1 \times 10^{-15} n_{\rm H_2} \frac{n_{\rm K^0}}{n_{\rm A^0}} \left(\frac{T}{1000 \text{ K}} \right)^{1/2} \exp\left(-\frac{5.04 \times 10^4 \text{ K}}{T} \right) \text{ s}^{-1}. \tag{A.8}$$

Here we assume that the ionization level is low and take $n_{\rm K^0}/n_{\rm A^0} = n_{\rm K}/n_{\rm A} = 0.11$.[1]

At high temperature, it is also necessary to include the thermionic emission of electrons from grain surface. The rate coefficients of this reaction for neutral and positively charged grains are given by the Richardson's law,

$$\alpha_{\rm te0} = \lambda_{\rm R} \frac{4\pi m_e (k_{\rm B}T)^2}{h^3} \exp\left(-\frac{W + e^2/a}{k_{\rm B}T} \right), \tag{A.9}$$

$$\alpha_{\rm te-} = \lambda_{\rm R} \frac{4\pi m_e (k_{\rm B}T)^2}{h^3} \exp\left(-\frac{W}{k_{\rm B}T} \right). \tag{A.10}$$

Here $\lambda_{\rm R}$ is an order-unity parameter depending on the grain material, and W is the energy required to make an electron escape the solid. We take $\lambda_{\rm R} = 1/2$ and $W = 5$ eV following Desch & Turner (2015).

A.4 CR shielding and isotopic ionization

In this thesis we have ignored the effect of CR shielding; however, the high disk column density we obtained in simulation suggest that this is not a good approximation. (However, we do not expect this to affect our main finding regarding magnetic field decoupling in the disk; see discussion in Section 3.5.2.) For future studies, we will include the effect of CR shielding and use (Umebayashi & Nakano, 1980)

$$\zeta = \zeta_{\rm CR} \exp\left(-\frac{\Sigma_{\rm H_2}}{96 \text{ g cm}^{-2}} \right) + \zeta_{\rm iso}. \tag{A.11}$$

Here the H_2 column density $\Sigma_{\rm H_2}$ at a given location will be estimated with the same algorithm as our optical depth estimate, which is discussed in Section C.3. $\zeta_{\rm iso}$ is the ionization rate due to radioactive decay of isotopes, which dominates ionization at very high column density. We assume that $\zeta_{\rm iso}$ is mainly contributed by ^{26}Al, which produces $\zeta_{26} = 1.94 \times 10^{-19}$ s^{-1} (Kunz & Mouschovias, 2009). One caveat is that ^{26}Al has a half-life of 0.716 Myr, which might be shorter than the lifetime of pre-stellar core if the core is initially

[1] This corresponds to taking $n_{\rm K} = 2.7 \times 10^{-7}\, n_{\rm H_2}$, following Kunz & Mouschovias (2009). There is a slight inconsistency between our choices of $n_{\rm K}$ and $n_{\rm A}$; the former assumes that approximately all potassium are gas phase atoms, while the latter takes into consideration the depletion of metal by accretion on grains; such depletion can be about one order of magnitude in the ISM for Na, K, and Mg (Snow, 1975). Our choice might still be reasonable, because at temperatures when thermal ionization becomes important, the K atoms in dust grains have likely been released during grain sublimation. Moreover, the temperature at which thermal ionization becomes important is insensitive to the exact value of $n_{\rm K}$ given the exponential temperature dependence of the thermal ionization rate.

subcritical.

Appendix B

Angular momentum diffusion at the grid scale

In this appendix we use a simple test problem to demonstrate and quantify the diffusion of angular momentum at the grid scale. The initial condition of our test problem is a 2D (R, ϕ) pressure-less ring with a density profile

$$\rho = \rho_0 \exp\left[-\frac{(R - R_0)^2}{2\sigma_0^2}\right], \tag{B.1}$$

with $\rho_0 = R_0 = 1$ and $\sigma_0 = 0.1R_0$. The ring evolves under the gravity of a central point mass with $GM = 1$, and the initial rotation profile is Keplerian. Physically, the ring is in equilibrium and its profile should remain constant.

We ran this test problem on Cartesian grids with resolution $R_0/\mathrm{d}x$ ($\mathrm{d}x$ is the cell size) ranging from 8 to 64; snapshots of the runs are shown in Fig. B.1. For all runs, the ring becomes significantly deformed after 100 Ω^{-1}. Meanwhile, the total angular momentum in the domain remains approximately the same during all simulations, suggesting that numerical dissipation (or creation) of angular momentum is slow compared to the numerical diffusion of angular momentum. Changing the integrator or Riemann solver does not qualitatively change the results. For comparison, we also ran the problem on a low-resolution cylindrical grid (Fig. B.1, bottom row). Because of the symmetry of the grid, the change in density profile is negligible even when the resolution is lower than all Cartesian runs.

To quantify the rate of angular momentum diffusion on a Cartesian grid, we measure the width of the ring by calculating the mass-weighted standard deviation of radius, σ. The evolution of σ for each run is shown in Fig. B.1. Diffusion of angular momentum tends to increase σ. Empirically, the timescale for the

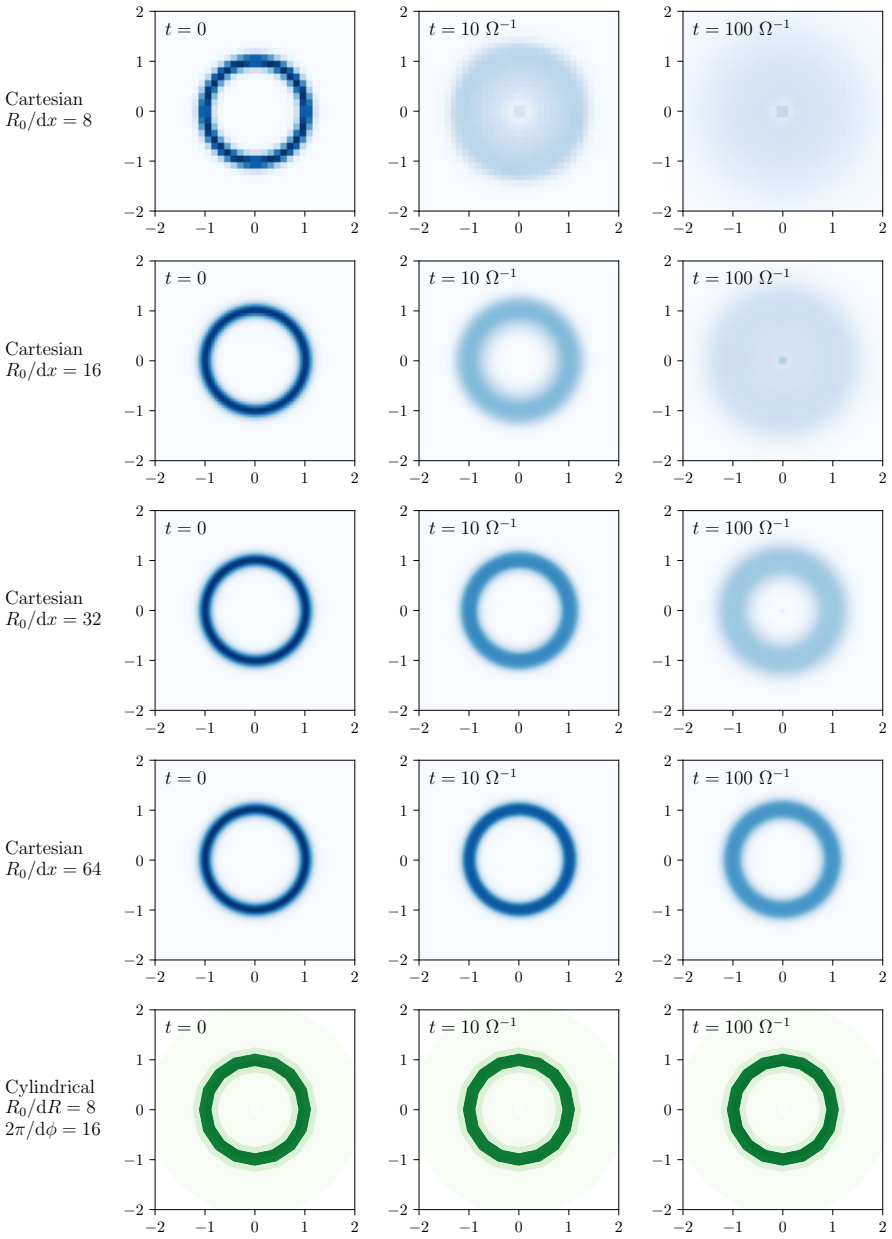

Figure B.1: Snapshots for evolving a 2D pressure-less ring with Keplerian rotation using different resolutions and grid geometries. All panels use the same color scale, except that the hue is different for different grid geometries. Cartesian grids are highly diffusive when resolution is low. Meanwhile, diffusion is negligible on a cylindrical grid even at very low resolution.

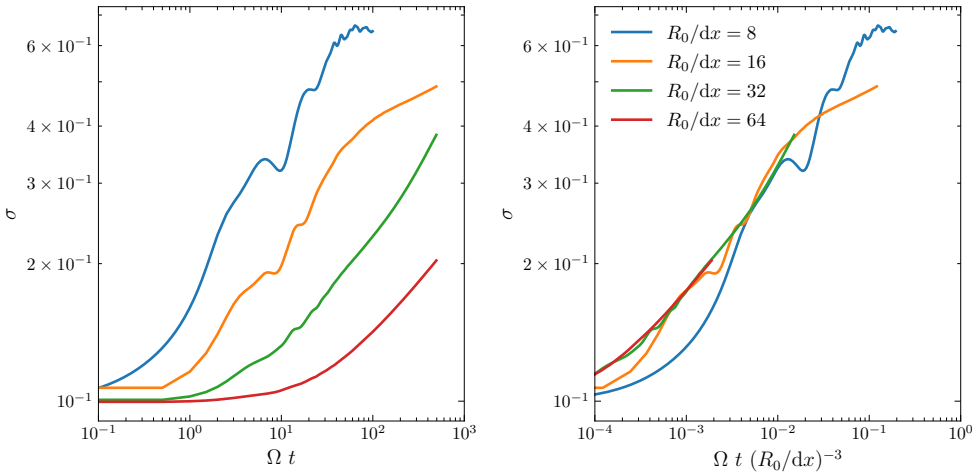

Figure B.2: Evolution of the ring width σ for Cartesian runs with different resolutions.

system to be dominated by this numerical diffusion, which is marked by σ reaching $\mathcal{O}(1)$, is of order

$$t_{\mathrm{diff}} \sim \Omega^{-1} \left(\frac{\mathrm{d}x}{R_0} \right)^3 . \tag{B.2}$$

Testing with different choices of σ_0, time integrator, and Riemann solver shows that this result is not sensitive to these choices as long as $\sigma_0 \ll 1$.

Appendix C

Details of numerical setup

C.1 A test of our self gravity solver

To test our implementation of self gravity, we simulate the collapse of an isothermal spherical overdensity whose centre is not at the coordinate origin. We set $4\pi G = 1$ and isothermal sound speed $c_s = 1$. The overdensity is centred at $(r, \theta, \phi) = (4, 1, 0.7)$, with initial density profile given in Figure C.1. At a resolution of $\mathrm{d}r/r \approx \mathrm{d}\theta \approx 0.05$ (comparable to our fiducial resolution) and $\mathrm{d}\phi \approx 0.1$, the evolution agrees well with analytic results, and the overdensity remains spherical during the collapse without visible artifacts (Figure C.1).

C.2 Handling discontinuity in interpolation tables

`Athena++` only supports interpolation tables with uniform spacing. Currently, our abundance profiles are smooth and using a uniform grid in $\log T$ works well. However, when we include dust sublimation for future studies, the species abundances will have sudden jumps at sublimation temperature T_{subl}, and a uniform $\log T$ grid will lead to large error unless the resolution is very high (which would significantly increase computational cost). In order to handle this, we first map $\log T$ to an auxiliary variable x and use x as the input variable for the interpolation table. The mapping is constructed such that integer x are the gridpoints for the interpolation table, and for each sublimation temperature, $T_{\mathrm{subl}} \pm \epsilon$ is mapped to $x = i$ and $x = i + 1$ for some integer i. This allows the code to properly treat discontinuities.

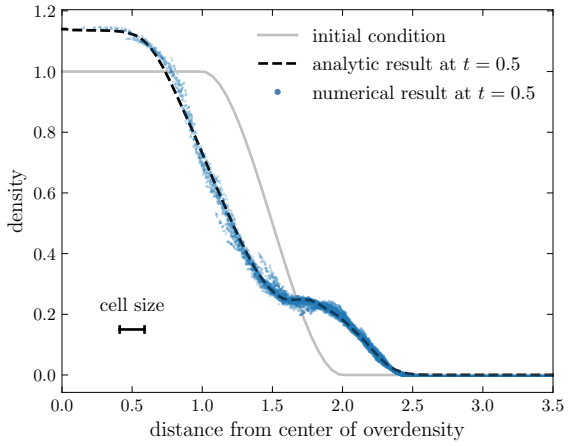

Figure C.1: Test problem for our self gravity algorithm. We simulate the collapse of a spherical overdensity whose centre is not at the coordinate origin. The numerical result shows good agreement with analytic prediction. The initial condition and typical cell size in the overdense region are also shown for reference.

C.3 Optical depth estimate

We estimate the optical depth at a given point in the simulation domain as follows. For a given cell, we define a collection of L-shaped rays that first go along the polar (θ) direction towards the pole by ≥ 0 cells, and then go out in the radial (r) direction to the outer radial boundary. We use the minimum optical depth (for Planck mean opacity) among these L-shaped rays as the optical depth estimate of the cell.

In practice, in order to find the L-shaped ray with the minimum optical depth τ_L, we first calculate the optical depth at a given cell on the radial ray going from it to infinity τ_r, and then perform the following iteration:

$$\begin{aligned} \tau_L^{i,0} &= \tau_r^{i,0}, \\ \tau_L^{i,j} &= \min\left\{\tau_r^{i,j}, \tau_L^{i,j-1} + \tfrac{1}{2} r_i \Delta\theta_{j-1} \kappa_P^{i,j-1} + \tfrac{1}{2} r_i \Delta\theta_j \kappa_P^{i,j}\right\}. \end{aligned} \tag{C.1}$$

The i, j are indices for the r and θ grid, where $j = 0$ corresponds to the cell at the pole; $\Delta\theta$ denotes the grid spacing in θ. The algorithm above can be efficiently parallelized, since in the first step (calculate τ_r) each column of radial cells can be processed independently and in the second step (calculate τ_L) each column of polar cells can be processed independently. We perform domain remaps similar to those used in our self-gravity solver to let each thread access whole columns of radial/polar cells in each of these steps.

In Fig. C.2 we compare the τ estimate using τ_L with the actual τ calculated using ray tracing with 160 randomly oriented rays from each cell. The error in $\exp(-\tau)$ and $\exp(-1/\tau)$, which are used in our cooling model to scale the optically thin cooling term and the diffusion term, are both small in most of the domain, suggesting that our estimate of τ is sufficiently accurate for the radiation model we adopt.

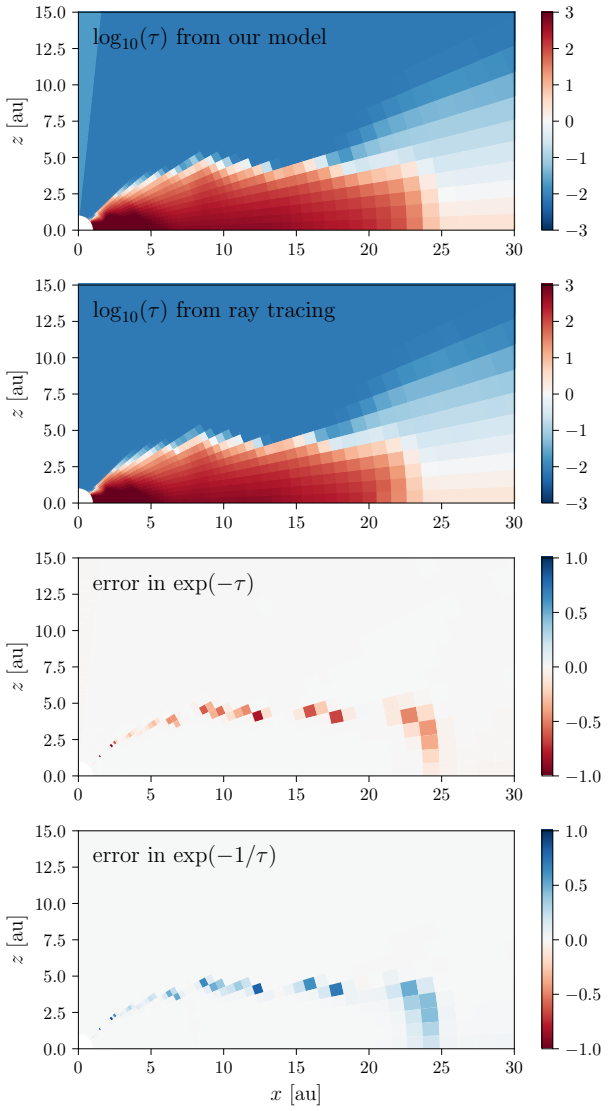

Figure C.2: Comparison between optical depth τ estimated using our method and calculated via ray tracing. Here we use a simulation snapshot at $M_{\rm tot} = 0.2$ M$_\odot$ at $\phi = 0$. The ray-tracing result shows the minimum optical depth among 160 randomly oriented rays for each cell. The absolute error in $\exp(-\tau)$ and $\exp(-1/\tau)$, which are used in our cooling model to scale the optically thin cooling term and the diffusion term, are also shown in the lower panels. The error remains $\ll 1$ for most of the domain.

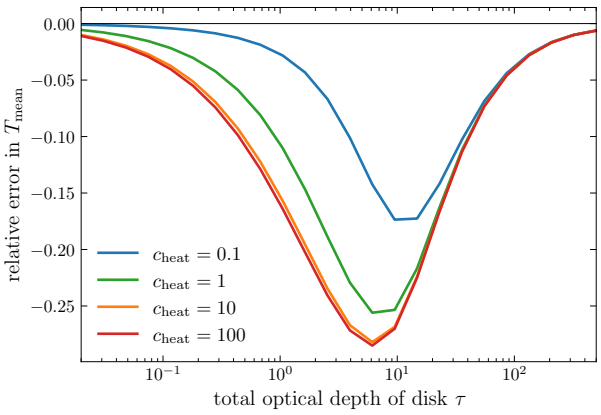

Figure C.3: The error in mean temperature (when compared to full radiative transfer) for our test problem, across a wide range of heating rates $c_{\rm heat}$ and disc optical depths τ. The error remains small ($\lesssim$25%), and converges to zero at $\tau \ll 1$ and $\tau \gg 1$.

C.4 A test of our radiative cooling model

To test the accuracy of our radiation model, we use the following test problem. Consider the equilibrium temperature profile of a heated, geometrically thin disc, in which the typical length scale in the disc plane is effectively infinity when compared to the disc thickness, making this becomes a 1D problem in the vertical direction. Let the disc have a density profile $\rho = \rho_0 \exp(-z^2/H^2)$ and a constant, frequency-independent κ. We solve for the temperature profile and the mean temperature when the whole disc is in thermal equilibrium with a uniform heating per mass Q/ρ. We characterize the heating rate by a dimensionless parameter

$$c_{\rm heat} \equiv Q/(4\rho\kappa\sigma T_0^4), \tag{C.2}$$

where T_0 is the ambient temperature at infinity.

Given that this problem is 1D, it can be solved semi-analytically with full radiative transfer without difficulty. In Figs C.3 and C.4, we compare the result of full radiative transfer with our radiation model. Fig. C.3 shows the error in the mean disc temperature predicted by our model for a wide range of different heating rates and disc optical depths. We see that the error converges to zero at $\tau \ll 1$ and $\tau \gg 1$, and is maximal at $\tau \sim 10$. The error increases for increasing $c_{\rm heat}$, but saturates for $c_{\rm heat} \gtrsim 10$ (which can be considered as the limit of ambient temperature $T_0 \to 0$). The error in mean temperature is always $\lesssim$25%, suggesting that our model can give quite accurate predictions for the mean disc temperature.

Fig. C.4 provides a more detailed comparison of the vertical temperature profile at a given $c_{\rm heat} = 10$ and a few different τ. Note that the $\tau = 10$ (green) curves are in the regime where the error of our model is maximal (see Fig. C.3). The temperature profile our model produces matches well with the full

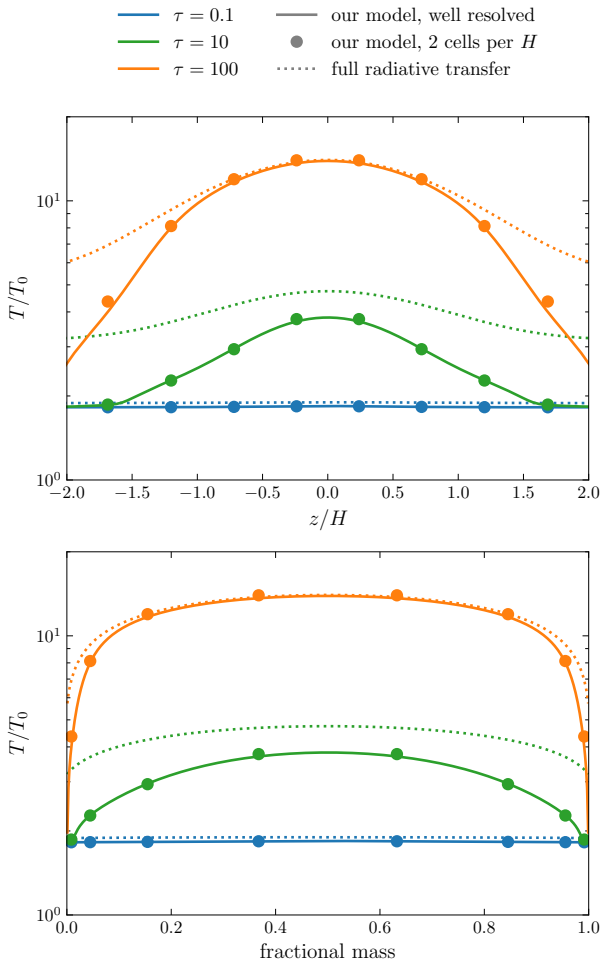

Figure C.4: Top panel: Comparison of vertical disc temperature profile in thermal equilibrium for our test problem. All calculations have $c_{\rm heat} = 10$, and different colours mark different disc optical depths τ. Different line styles correspond to our model with high resolution (solid), our model with low resolution (2 cells per scale height H, circles), and the exact result from full radiative transfer (blue). Bottom panel: Same as top panel, except the horizontal axis now shows the fractional mass below a given height. For most of disc mass, our model gives small error.

radiative transfer result near the midplane (which contains most of the mass), especially when τ is $\gg$1 or $\ll$1. The error in the atmosphere is often finite, and our model tends to underestimate the temperature there. This is mainly because our model ignores the re-absorption of the cooling radiation from the midplane by the optically thin atmosphere. However, we do not expect this finite error to have a large impact on our simulation result. The temperature in the atmosphere affects the dynamics through pressure and chemistry (which eventually affects the non-ideal diffusivity). The low density in the atmosphere makes the pressure dynamically unimportant, even if the temperature is as high as the disc. Meanwhile, the non-ideal diffusivity there is always high (in fact usually above its numerical cap; see Appendix C.5), and some order-of-unity modification to this already high diffusivity should barely affect the evolution.

An additional advantage of our radiation model is that it performs very well at low resolution. As shown in Fig. C.4, at resolution as low as 2 cells per scale height (which is already lower than the typical resolution seen in our simulation), the prediction of the model still matches the well-resolved result very well.

C.5 Velocity and diffusivity caps

To avoid very short timesteps, we apply caps on the magnetic diffusivities (both Ohmic and ambipolar), the flow speed v, and the Alfvén speed $v_{\rm A}$. The cap on η is applied directly when calculating η at each timestep, and the caps on v and $v_{\rm A}$ are applied by increasing the density when the velocity magnitude is greater than the cap value (while holding momentum and magnetic field constant). In other words, the velocity cap is implemented as a density floor.

These caps are often applied to star-formation simulations, but sometimes their values are chosen somewhat arbitrarily. Here we try to give a more physical choice of cap values that in principle should not affect the dynamics of the protostar-disc system. In order to do this, we choose the caps so that the characteristic timescale associated with the diffusivity cap, $r^2/\eta_{\rm cap}$, is at most comparable to the free-fall timescale; and the characteristic timescale associated with the velocity cap, $r/v_{\rm cap}$, is much smaller than the free-fall timescale. (Note that both free-fall timescale and Keperian timescale are $\sim \sqrt{-g_r/r}$.) The rationale for the first criteria is that, when the diffusion timescale is smaller than the gravitational (free-fall, or Keplerian) timescale, the field is largely decoupled from the gas and the behaviour is not sensitive to the exact value of η. We also want to choose the cap values such that the CFL timestep corresponding to the caps are similar in the whole domain. This allows us to choose higher caps for larger r without affecting the timestep. Since the cell size are $\propto r$, we choose $\eta_{\rm cap} \propto r^2$ and $v_{\rm cap} \propto r$. Combining the two motivations above, we choose

the following caps for diffusivity and velocity:

$$\eta_{\rm cap} = f_\eta \tau_g^{-1} r^2, \quad v_{\rm cap} = f_v \tau_g^{-1} r. \tag{C.3}$$

Here τ_g^{-1} is defined as the global maximum of $\sqrt{-g_r/r}$ on the midplane, and f_η, f_v are constant factors. Our default choices are $f_\eta = 1$ and $f_v = 50$. To check whether these caps affect the dynamics, in XK21a we re-ran some 2D simulations with $f_v = 100$ and with $f_\eta = 0.5$, 2, and 4. All of these runs show nearly identical behaviour.

In XK21b (which is the simulation shown in this thesis) we introduce another diffusivity cap to tackle another numerical issue that could produce artifacts in low-density regions. Consider a region where the total energy is dominated by the magnetic energy and the magnetic diffusivity is very large. The low kinetic energy and pressure requires the magnetic field to be nearly force-free, i.e., $\boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{B}) \approx 0$. Meanwhile, large magnetic diffusivity also requires $\boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{B}) \approx 0$, since otherwise that would create a large drift velocity (and therefore large $\partial_t \boldsymbol{B}$; see Section 3.5.1). However, in `Athena++`, the interpolations in magnetic field used for calculating the Lorentz force and the non-ideal EMF are not exactly self-consistent; i.e., the small numerical interpolation errors in $\boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{B})$ are generally going to be different between the Lorentz force and non-ideal drift velocity. When $\boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{B})$ physically needs to be very small (below the typical interpolation error), this could lead to significant unphysical perturbations in the flow.

To avoid this problem, we now choose

$$\eta_{\rm cap} = \min\left(f_\eta \tau_g^{-1} r^2, \quad f'_\eta \sqrt{-g_r r^3}\right). \tag{C.4}$$

The second term (where we choose $f'_\eta = 10$) is determined by the local dynamical timescale and ensures that the diffusivity is still strong enough to force the field to be nearly force-free in strongly diffusive regions, but not so strong as to cause the unphysical perturbations discussed above. In practice, the second term only affects part of the outflow cone, and increasing f'_η barely affects the result.

C.6 Polar averaging

One major disadvantage of using spherical-polar coordinates for 3D simulations is that the cells become narrow wedges towards the pole, which may limit the timestep severely. This problem is sometimes tackled using mesh refinement near the midplane, which allows the ϕ resolution near the midplane to be higher than that around the pole. However, this choice is not suitable for our simulation since our self gravity solver is

not (yet) compatible with mesh refinement.

To circumvent this problem, we need to "de-refine" cells near the pole in the ϕ direction. This is achieved by performing a "polar averaging", where high-spatial-frequency flctuations in the ϕ direction is removed near the pole.

C.6.1 Polar averaging algorithm in XK21a

After each update of cell quantities, we perform a Fourier transform in the ϕ direction for cells near the pole and truncate the high-frequency terms of density, momentum, and magnetic field. (For magnetic field, we do this to B_r and B_θ, and then B_ϕ follows from $\boldsymbol{\nabla}\cdot\boldsymbol{B}=0$.) We also perform the same truncation for the EMF before using it to update the magnetic field. The number of terms kept, n_ϕ, is chosen such that the effective cell size in the ϕ direction,

$$\mathrm{d}\phi^{\mathrm{eff}} = r\sin\theta(\mathrm{d}\phi)N_\phi/n_\phi, \tag{C.5}$$

is no smaller than $r(\mathrm{d}\theta)_{\min}$. Here θ is the poloidal angle of the cell centre, and $\mathrm{d}\theta$ and $\mathrm{d}\phi$ are the mesh spacings in the poloidal and azimuthal directions; $(\mathrm{d}\theta)_{\min}$ is the minimum $\mathrm{d}\theta$, which is located in the midplane. We then use this effective cell size to replace the actual cell size when computing the CFL condition for the next timestep. This averaging procedure is necessary only when we need $n_\phi < N_\phi$ to get $\mathrm{d}\phi^{\mathrm{eff}} \geq r(\mathrm{d}\theta)_{\min}$. For example, in our simulation, the averaging is only applied to the first two cells in the θ direction near the pole.

C.6.2 Polar averaging algorithm in XK21b and this thesis

The polar-averaging scheme we used in XK21a is equivalent to convolving with a filter that takes the shape of a wavelet. This algorithm works well for barotropic simulations, but can be problematic when applied to non-barotropic simulations. When using this scheme on a non-barotropic system, we are averaging total energy, density, momenta, and magnetic field, while internal energy is implicitly inferred from them. Because the filter for the convolution is not always non-negative, it is possible to obtain negative internal energy after the average. Since the code will force the pressure to be non-negative, this results in strong unphysical heating and can break the simulation.

The above problem can be resolved if we switch to a filter that is always non-negative. (One can mathematically prove that this guarantees the resulting internal energy to be positive.) Therefore, for simulations in this thesis we adopt a triangular filter instead. For example, if we are averaging a variable y

over k_0 cells, we have

$$y_k \text{ after averaging} = \sum_{s=-k0}^{k0} y_{k+s} \frac{k_0 - |s|}{k_0^2}. \tag{C.6}$$

One caveat is that this filter damps all modes with nonzero (spatial) frequency. Since the filter is applied at every timestep, any non-axisymmetric perturbation will be strongly damped in the polar-averaging region, which covers the first two cells around the pole. This is not too problematic for our current simulation since the polar region is expected to be near axisymmetry. However, if in the future we want to simulate a system with a non-axisymmetric polar region (e.g., with misalignment between the magnetic and rotation axes), we will need to improve this treatment.

C.7 Polar correction for EMF

C.7.1 Problem description and possible origin

During the first few 100 kyr of pre-stellar collapse, the evolution is slow and we expect the rotation in the inner part of the core to be approximately uniform (similar to the initial condition). However, as the top left panel of Fig. C.5 demonstrates, our simulation shows significant unphysical spin-up near the pole, where the rotation rate is well above the initial rotation. (This problem is less significant after protostar formation, as demonstrated in XK21a Appendix D.)

This problem is related to geometric truncation errors in the EMF calculation, which eventually lead to unphysical generation of a toroidal magnetic field (Fig. C.5 bottom left panel) and unphysical angular-momentum transport. The origin of this problem can be illustrated by considering a uniformly rotating system threaded by a vertical magnetic field. Physically, one expects the path integral of the poloidal EMF around a cell to be zero, so that the toroidal field remains zero. However, in a simulation with finite resolution, the EMF on each edge are subject to different geometric errors, and these errors in general do not cancel out. Additionally, in a spherical polar grid the amplitudes of such errors do not decrease quickly as resolution increases because the error for the first couple of cells around the pole is always $\sim\mathcal{O}(1)$.

C.7.2 Polar EMF correction algorithm in XK21a

To avoid the unphysical spin-up observed in Fig. C.5, we introduce a correction in our code to ensure that a uniformly rotating system remains uniformly rotating, with no unphysical generation of B_ϕ or artifact in Ω. This can be achieved, e.g., by modifying how the left and right interface values of v_ϕ are interpolated in the Riemann solver. (This correction is not applied to simulations shown in the main text.)

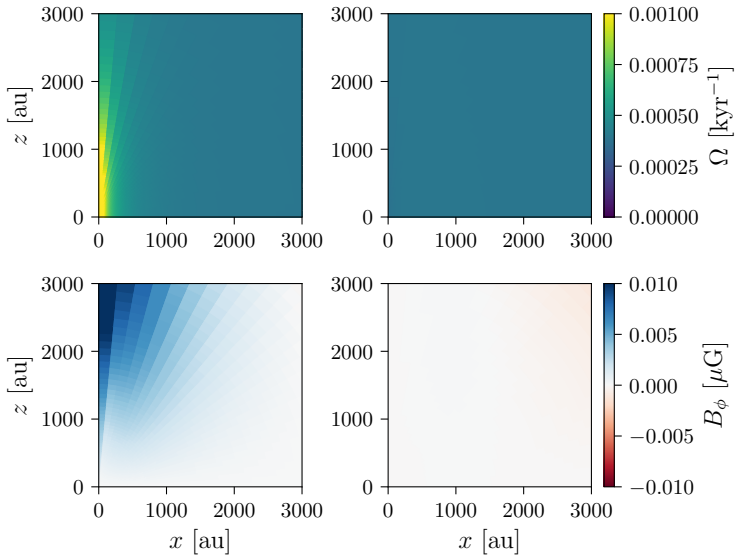

Figure C.5: Comparison between 2D simulations in XK21a with (right) and without (left) the polar correction for EMF at $t = 100$ kyr. Physically, the inner few 1000 au shown here should still be rotating approximately uniformly. Without the polar correction, positive B_ϕ is generated unphysically; this leads to spin-up of the polar region.

In `Athena++`, for a 2D axisymmetric spherical polar grid, the poloidal EMF on cell edges (the path integral of which is used to update the toroidal field) is calculated using a Riemann solver on the cell interface in r and θ directions. For each interface, the code first interpolates cell-centered quantities to get left and right states on the interface, and then uses these left and right interface states to compute fluxes on the interface, including the face-centered EMF. Note that, for the r and θ directions, the face-centered EMF is the same as the edge-centered EMF because of axisymmetry. The problem lies in how the azimuthal velocity v_ϕ is interpolated: the interpolation algorithm in `Athena++` directly interpolates v_ϕ and, for a uniformly rotating system, the interpolated state on the interface sometimes differs from the actual v_ϕ. This kind of error generally will not cancel out in the path integral of the EMF, making it possible to generate unphysical B_ϕ.

Our fix to the problem is fairly simple. Instead of directly interpolating v_ϕ, we first calculate cell-centered $\Omega = v_\phi/(r\sin\theta)$ and then interpolate Ω to the cell interface. We then compute the edge-averaged v_ϕ on the interface based on the interpolated Ω, and use that as input for the Riemann solver. (Here, 'edge-averaged' means averaging along an edge in the r direction for a θ interface and along the θ direction for an r interface.) This eliminates geometric error when the system is uniformly rotating. This correction works very well for uniformly rotating systems, and applying it to our 2D simulations removes the artifact near the pole (Fig. C.5, right panels) without introducing any new artifacts.

In 3D the correction above can no longer fully solve the problem. This is because the r, θ edge-centered

EMFs are not the same as the corresponding interface EMFs in the absence of axisymmetry. Instead, now an edge EMF depends on all four adjacent cells and interfaces and that introduces additional geometric errors that are harder to account for. Therefore, for simulation in this thesis we do not include any EMF correction.

C.7.3 Polar EMF correction algorithm for future studies

For future studies we plan to use a different (and probably more effective) version of polar EMF correction. Instead of modifying the input velocity of the Riemann solver (which then affects the face-centered EMFs), we directly modify the edge EMF computed by the code before the EMF is used for updating the magnetic field. More specifically, we scale it by the ratio between edge-averaged $R = r\sin\theta$ and edge-centered R. This fixes the geometric error to lowest order; the result is exact when the system is uniformly rotating with uniform vertical field. Initial tests of this treatment suggests that it performs well in both 2D and 3D.

Appendix D

Numerical convergence

D.1 Resolution

We first check the numerical convergence with respect to increased resolution. In Figs. D.1 and D.2, we compare the angular momentum budget and 1D profiles between simulations with different resolutions. Increasing r and ϕ resolution barely affect the simulation results. However, increasing θ resolution leads to a decrease in angular momentum budget, which then impacts the disk profile (especially disk size). The poor convergence with respect to θ resolution might be related to the problem discussed in Appendix C.7, where geometric error in the EMF calculation in polar region leads to unphysical spin-up of the inner part of the pre-stellar core. To test whether this is the case, we perform the same convergence test on our new simulation setup (Figs. D.3 and D.4), which includes a polar EMF correction that should significantly alleviate this problem (Appendix C.7.3). With this correction, increasing θ resolution barely affects simulation results.

Given the low ϕ resolution in our fiducial simulation, we also check whether the existence of large-scale spirals (as opposed to a small-scale "gravito-turbulence") is due to insufficient resolution. In Figs. D.5 and D.6 we check this by comparing midplane density snapshots and the average amplitude in modes with different m. Our fiducial resolution is enough to produce correct amplitudes for low-m modes, therefore the existence of global, large-scale spirals is not just a resolution effect. One caveat is that the exact shape of the spiral and how they propagate at the outer edge of the disk can be slightly different at higher resolution; generally, we find the spirals propagating slightly further out at higher resolution.

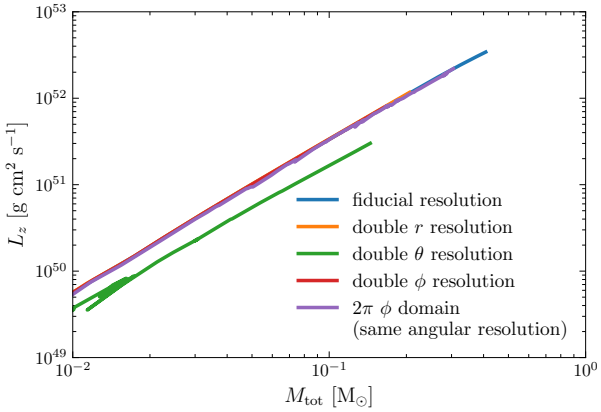

Figure D.1: Mass-angular-momentum relation for our simulation at different resolutions. All results are nearly identical, except for the run with double θ (green); this is related to the polar artifact discussed in Appendix C.7 and can be fixed by introducing a polar EMF correction (see Fig. D.3).

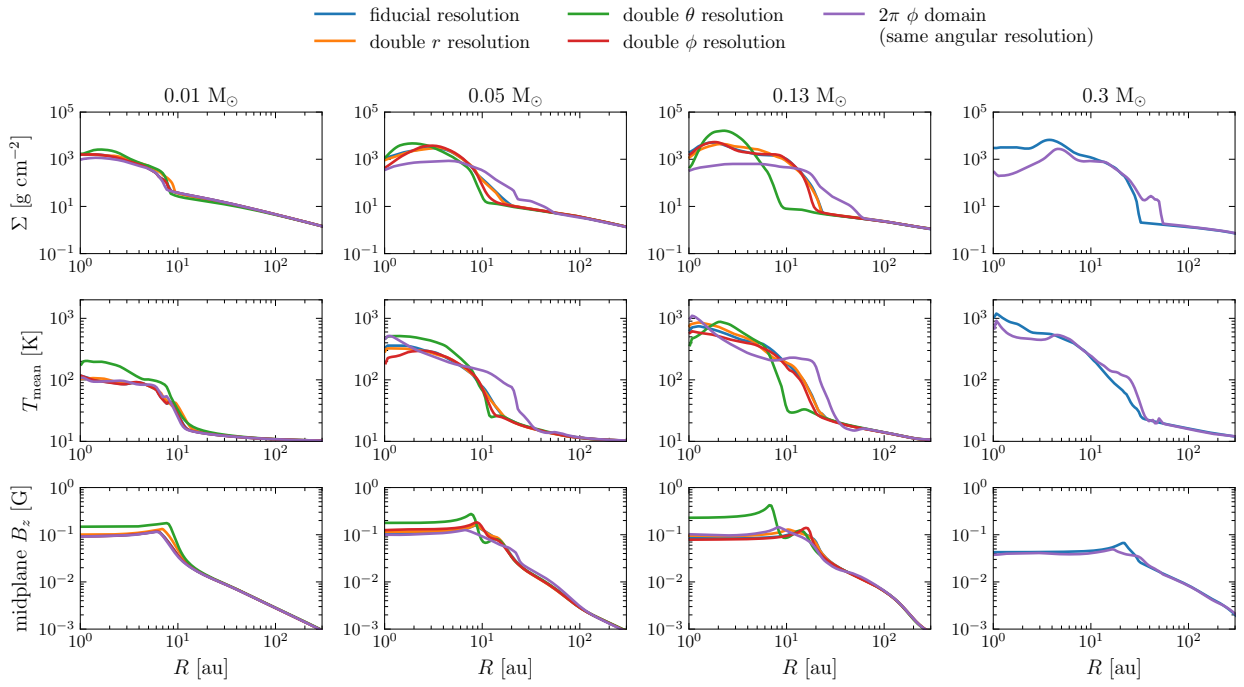

Figure D.2: 1D profiles of our simulation at different resolutions at several different epochs. Most runs show good agreement, with the exception of the double θ resolution run (green) and the 2π ϕ domain run (purple). The former is related to the polar artifact discussed in Appendix C.7 and can be fixed by introducing a polar EMF correction (see Fig. D.4). The latter is probably related to the interaction between the accretor (or inner boundary) with the disk, which is discussed in Appendix D.2.

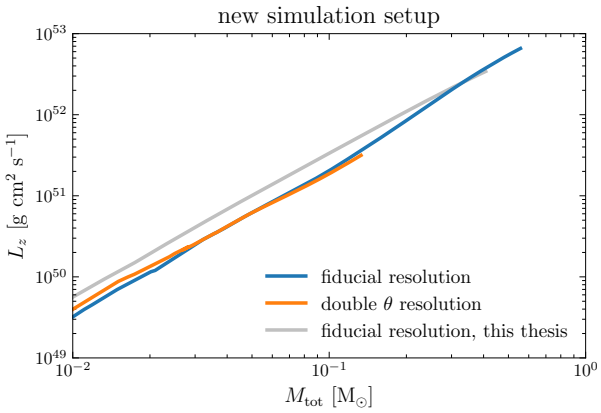

Figure D.3: Similar to Fig. D.1 for our new simulation setup to be used for future studies. The lack of convergence in θ resolution is resolved by introducing the polar EMF correction introduced in Appendix C.7.3.

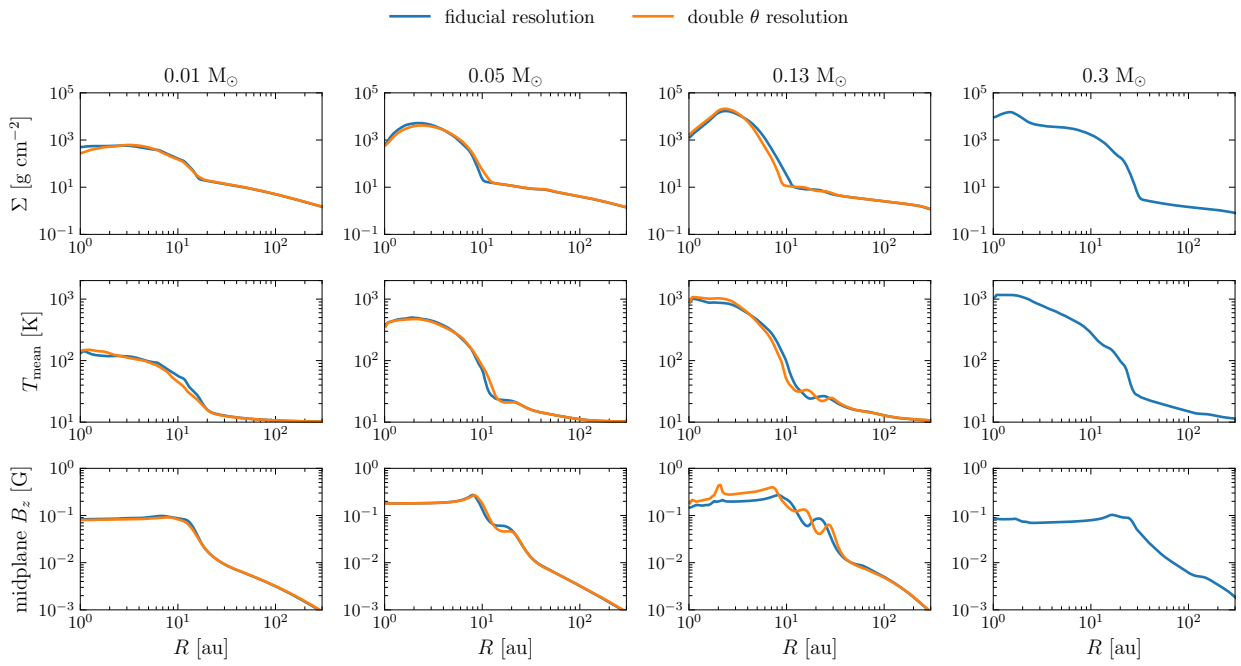

Figure D.4: Similar to Fig. D.2 but for simulations using our new setup. The lack of convergence in θ resolution is resolved by introducing the polar EMF correction introduced in Appendix C.7.3.

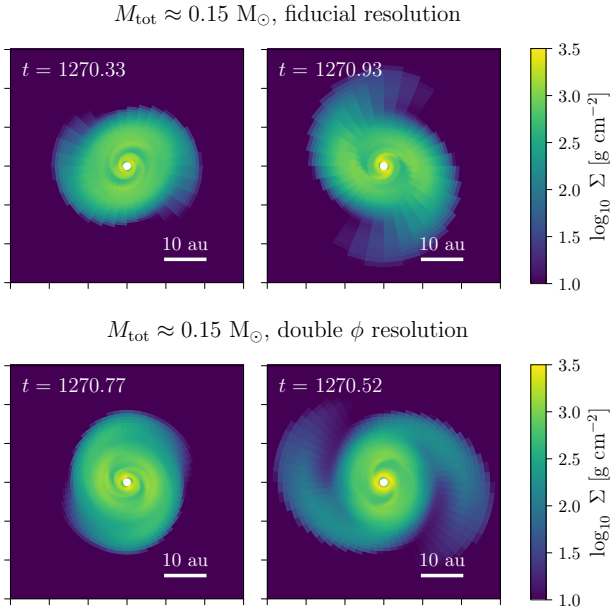

Figure D.5: Midplane density snapshots similar to Fig. 3.5 for fiducial (top) and double-azimuthal-resolution (bottom) runs around the same $M_{\rm tot}$.

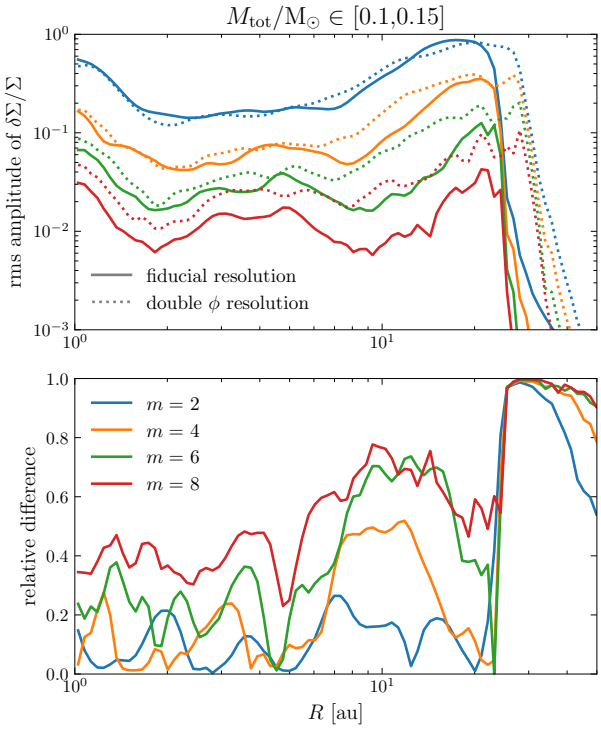

Figure D.6: RMS relative amplitude of column-density perturbations $\delta\Sigma/\Sigma$ with different azimuthal wavenumbers m (top panel) for the fiducial (solid) and double-azimuthal-resolution (dotted) runs. The relative difference (absolute value) is shown in the bottom panel for reference. The fiducial run accurately predicts the amplitude of the dominant $m = 2$ mode, but the error increases for higher m.

D.2 The effect of having a 2π ϕ domain

In addition to varying resolution, in Figs. D.1 and D.2 we also include a run with the same (angular) resolution but a ϕ domain covering full 2π. (This run is used for making the mock observation in Fig. 5.16, as it do not have the artificial $m = 2$ symmetry.) The angular momentum budget of the disk remains similar to the other runs, suggesting that this change does not affect the pseudodisk. This is expected as the pseudodisk is close to axisymmetry. However, the 1D profiles are different from the other runs in that the column density at small radii ($\lesssim 5$ au at later times) is much lower and the disk size is slightly larger. The latter could result from the $m = 1$ spiral propagating further, and additional angular momentum available to the outer disk when the inner disk is less massive.

The less massive inner disk is more challenging to explain. The low column density makes the inner disk gravitationally stable, suggests that angular momentum transport there is facilitated by mechanisms other than gravitational instability. One possibility is that the angular momentum transport at small radii is facilitated by the motion of the protostar. Once we relax the forced $m = 2$ symmetry, the disk center of mass is no longer pinned at the origin, nor does it have to coincide with the protostar. Given that the disk containing finite mass (often as massive as the protostar), the disk center of mass and the protostar can orbit around each other. This relative motion might then create perturbations similar to spiral waves driven by a binary in a circumbinary disk, which transports angular momentum and truncate the disk at a finite radii. Another possibility is that the motion of the disk with respect to the inner boundary simply makes the inner boundary carving a hole larger than $r_{\rm in}$, which implies that the truncation at a relatively large radius is mainly a numerical effect.

However, our current numerical setup does not allow a more detailed investigation on these possibilities. Since we always fix our accretor at the origin, our simulation only includes the motion of the disk center of mass around the protostar, but not the protostar's motion. Given this limitation, the results obtained from this simulation may not be accurate and we leave the investigation of disk tidal truncation by this binary-like protostellar motion to future studies.

D.3 Inner boundary size and boundary condition

Finally, we consider the effect of the inner boundary size $r_{\rm in}$ and inner boundary condition. This has been studied in XK21a with a set of barotropic 2D simulations with different inner boundary sizes and inner boundary conditions. In Fig. D.7 we show the M–L_z relation of a set of 2D barotropic simulations from XK21a with $r_{\rm in}$ ranging from 0.25 au to 2 au. We see a very good agreement, even when the largest $r_{\rm in}$ we

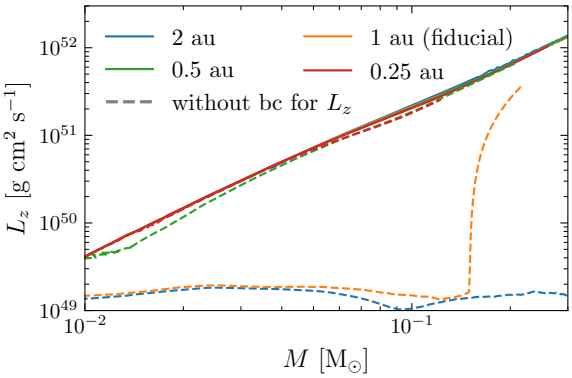

Figure D.7: Mass–angular-momentum relation of the protostar-disk system for 2D simulations in XK21a (similar to the simulation in this thesis but with a barotropic EoS) with different inner boundary sizes and boundary conditions. Simulations without the zero angular-momentum flux boundary condition are plotted as dashed lines.

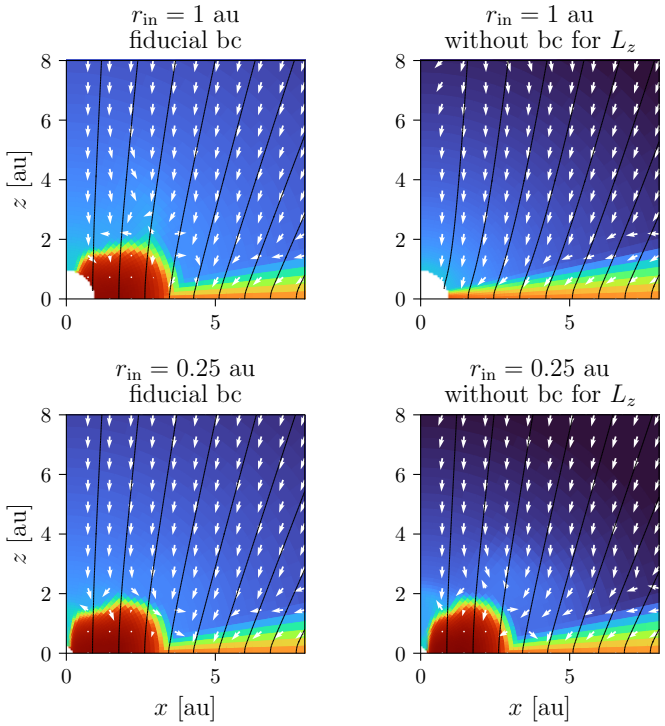

Figure D.8: Snapshots at $M_{10} = 0.1 M_\odot$ for the 2D simulations in XK21a shown in Fig. D.7. Color corresponds to log density. Simulations show good agreement if $r_{\rm in}$ is sufficiently small or the boundary condition on L_z flux is applied. When $r_{\rm in}$ is relatively large and the L_z flux boundary condition is not used (top right panel), disk formation is fully suppressed.

use is already comparable to the size of the first hydrostatic core/torus.

To test whether our inner boundary condition is correct and necessary, we run another set of 2D simulations with a different boundary condition where we do not force the angular-momentum flux through the inner boundary to be zero. The result for the two different boundary conditions are similar for $r_{\rm in} \lesssim 0.5$ au, suggesting that our boundary condition does not introduce any artifact that affects the angular-momentum budget of the system. For larger $r_{\rm in}$, however, simulations without the angular-momentum flux boundary condition show very poor numerical convergence and much lower L_z. And for $r_{\rm in} \geq 1$ au, removing the angular-momentum flux boundary condition makes the torus disappear completely (Fig. D.8). Therefore, having a boundary condition that limits the angular-momentum flux through the inner boundary is necessary, unless one can afford an inner boundary size as small as 0.5 au or less.

The necessity of having a small inner boundary or limiting the angular-momentum flux through the inner boundary has a physical origin. The mean specific angular momentum coming from the pseudodisk is quite low; for our fiducial simulation, the circularization radius is $\sim$1 au. (Note that $L_z \propto M^{1.5}$ gives approximately constant circularization radius.) In order to form a disk, mass has to be accumulated around this circularization radius until the torus becomes gravitationally unstable and starts transporting angular momentum. However, when the inner boundary is larger than this circularization radius and angular momentum can be freely advected through the inner boundary, all infalling angular momentum will be lost through the inner boundary and the disk cannot form.

Additionally, the two boundary conditions compared here correspond to two limiting scenarios of angular-momentum transport at $r < r_{\rm in}$. Our fiducial boundary condition with zero L_z flux effectively assumes that all angular momentum that gets within $r_{\rm in}$ eventually returns to $r > r_{\rm in}$; removing the L_z flux boundary condition effectively assumes that all angular momentum that gets past $r_{\rm in}$ is eventually removed from the protostar-disk system. The fact that these two boundary conditions give very similar evolution when $r_{\rm in} \lesssim 0.5$ au suggests that the evolution is not sensitive to the details of the angular-momentum transport at small radii ($\lesssim$0.5 au). One caveat is that these two boundary conditions both implicitly assume that nothing occurring inside of the inner boundary can affect angular-momentum transport in the active domain.

Appendix E

Detailed estimate of disc thermal budget

In this appendix we analyze the thermal budget of an accreting disc (subject to no external irradiation) to show that the cooling rate required for thermal equilibrium is generally of order $-g_R\dot{M}$. This result does not require very specific assumptions on how energy and angular momentum are transported in the disc, and is still applicable when the disc thickness and disc-to-star mass ratio are finite.

E.1 Definitions

For our problem, we can identify three independent dimensionless small parameters:

$$\epsilon_1 \equiv \frac{\dot{M}}{\pi R^2 \Sigma \Omega}, \tag{E.1}$$

$$\epsilon_2 \equiv (H/R)^2, \tag{E.2}$$

$$\epsilon_3 \equiv \frac{\pi R^2 \Sigma}{M_{\rm tot}}. \tag{E.3}$$

Here ϵ_1 corresponds to the rate of column density evolution (in units of Ω), ϵ_2 is the square of disc aspect ratio (and is comparable to the ratio between thermal and kinetic energy), and ϵ_3 is comparable to the disc-to-star mass ratio. Thin-disc models typically simplify the problem by assuming $\epsilon_2, \epsilon_3 \to 0$; such an assumption is not necessarily valid in young protostellar discs. To ensure the generality of our results, we do not assume any specific ordering between these parameters, except that $\epsilon_1, \epsilon_2 \ll 1$, $\epsilon_3 \lesssim 1$, and $\epsilon_3 \lesssim \epsilon_2^{1/2}$. The last relation comes from the assumption that gravitational self-regulation should be able to maintain

$Q \gtrsim 1$. For brevity we use the normalization $\Sigma, R, \Omega \sim 1$ when labeling the size of each term.

We expect each variable to be the sum of a slowly varying axisymmetric component, corresponding to the mean disc profile whose characteristic rate of evolution is $\lesssim \epsilon_1 \Omega$, and a rapidly varying (and generally non-axisymmetric) component, corresponding to turbulent fluctuations whose characteristic rate of variation is $\sim \Omega$. For any variable a, we separate these two components by dividing it into a mean component $\bar{a}$ and a fluctuating component a' as follows:

$$\bar{a}(t, R, z) \equiv \mathcal{F}(\langle a \rangle_\phi), \quad a' \equiv a - \bar{a}, \tag{E.4}$$

where $\mathcal{F}$ is a low-pass filter (in time) with a cutoff frequency between Ω and $\epsilon_1 \Omega$, and $\langle \dots \rangle_\phi$ denotes an average in ϕ. We assume that the mean and fluctuating components are well separated in frequency space. More specifically, we assume that there exists a filter $\mathcal{F}$ such that (i) $\bar{a}$ and its time derivative are dominated by contribution from frequencies $\lesssim \epsilon_1 \Omega$, and a' and its time derivative are dominated by contributions from frequencies $\gtrsim \Omega$; and (ii) the fluctuating component of one variable a' is uncorrelated with the mean component of any other variable $\bar{b}$, or some function of the mean component of multiple variables (e.g., $\bar{b}\bar{c}, \bar{b}/\bar{c}$).

For variables that are approximately constant inside the disc in vertical direction at a given radius, such as velocity and gravitational potential, we can further split the mean component $\bar{a}$ into a z-independent piece and a z-dependent piece,

$$a_0(t, R) \equiv \langle \overline{\rho a} \rangle_z / \langle \bar{\rho} \rangle_z, \quad a_1 \equiv \bar{a} - a_0. \tag{E.5}$$

Here $\langle \dots \rangle_z \equiv \int_{-\infty}^{\infty} \mathrm{d}z \, (\dots)$. Our earlier assumption that the mean and fluctuating components are well-separated in frequency implies that a_0, a_1 are both uncorrelated with any fluctuating variable b'.

E.2 Basic equations

The evolution of mass, momentum, and energy in the disc can be described by the following equations:

$$\partial_t \rho + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}) = 0, \tag{E.6}$$

$$\partial_t (\rho \boldsymbol{v}) + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v} \boldsymbol{v}) = \rho \boldsymbol{g}, \tag{E.7}$$

$$\partial_t (\rho u + \tfrac{1}{2} \rho v^2) + \Phi \partial_t \rho + \boldsymbol{\nabla} \cdot [(\rho u + \tfrac{1}{2} \rho v^2 + \rho \Phi + p) \boldsymbol{v}] = Q_{\mathrm{rad}}. \tag{E.8}$$

where Φ includes contributions from both the protostar and the disc. To get the last equation we have used that $\rho\boldsymbol{v}\cdot\boldsymbol{g} = -\boldsymbol{\nabla}\cdot(\rho\boldsymbol{v}\Phi)+\Phi\boldsymbol{\nabla}\cdot(\rho\boldsymbol{v}) = -\boldsymbol{\nabla}\cdot(\rho\boldsymbol{v}\Phi)-\Phi\partial_t\rho$. These equations ignore the effect of the magnetic field for simplicity; this is a reasonable approximation for our simulation, since the magnetic field in the disc is very weak and mostly straight due to magnetic diffusion. Meanwhile, for discs in which the magnetic field affects the transport of angular momentum and energy, one can show that the heating associated with these processes is at most $\sim -g_R\dot{M}$ per radius (e.g., Balbus et al., 1994).

Next, we use the following operator to average the equations:

$$\langle\,\ldots\rangle \equiv \frac{1}{\Delta t}\int_{t-\Delta t/2}^{t+\Delta t/2}\mathrm{d}t'\int_{-\infty}^{\infty}\mathrm{d}z\int_0^{2\pi}R\mathrm{d}\phi\,(\ldots). \tag{E.9}$$

This averaging corresponds to integrating over the cylindrical surface at R and averaging over a time window Δt, which we choose to be $\lesssim(\epsilon_1\Omega)^{-1}$ but $\gg\Omega^{-1}$. This averages out turbulent perturbations while retaining the long term evolution of the mean disc. The operator (E.9) has the following properties:

$$\begin{aligned}
&\langle\partial_t\,\ldots\rangle = \partial_t\langle\ldots\rangle,\\
&\langle\boldsymbol{\nabla}\cdot\boldsymbol{A}\rangle = \partial_R\langle A_R\rangle,\\
&\langle\boldsymbol{\nabla}\cdot\mathsf{T}\rangle\cdot\hat{\boldsymbol{\phi}} = R^{-1}\partial_R\langle R\mathsf{T}_{R\phi}\rangle.
\end{aligned}$$

Here $\boldsymbol{A}$ is any vector and T is any tensor; $\hat{\boldsymbol{\phi}}$ is the unit vector in ϕ. Given our earlier assumption that the mean and turbulent components are uncorrelated, we also have that

$$\langle a\rangle = \langle\bar{a}\rangle,\quad \langle ab\rangle = \langle\bar{a}\bar{b}+a'b'\rangle,\quad \langle\rho a\rangle = \langle\overline{\rho a}\rangle = \langle\bar{\rho}a_0\rangle. \tag{E.10}$$

After applying $\langle\ldots\rangle$ to equations (E.6)–(E.8) (where we only take the ϕ component of the momentum equation), we obtain

$$\partial_t\langle\bar{\rho}\rangle + \partial_R\langle\bar{\rho}v_{R,0}\rangle = 0, \tag{E.11}$$

$$\begin{aligned}
&\partial_t\langle R\bar{\rho}v_{\phi,0}\rangle + \partial_R\langle R\bar{\rho}v_{R,0}v_{\phi,0} + R\overline{\rho v_R}v_{\phi,1}\\
&\quad + R(\rho v_R)'v'_\phi + (4\pi\mathrm{G})^{-1}Rg'_Rg'_\phi\rangle = 0,
\end{aligned} \tag{E.12}$$

$$\begin{aligned}
&\partial_t\langle\bar{\rho}(u+\tfrac{1}{2}v^2)_0\rangle + \langle\Phi_0\partial_t\bar{\rho}+\Phi_1\partial_t\bar{\rho}+\Phi'\partial_t\rho'\rangle\\
&\quad + \partial_R\langle\bar{\rho}v_{R,0}e_{\mathrm{tot},0} + \overline{\rho v_R}e_{\mathrm{tot},1} + (\rho v_R)'e'_{\mathrm{tot}}\rangle = -2\pi R\Lambda_{\mathrm{cool}}.
\end{aligned} \tag{E.13}$$

Here $\Lambda_{\mathrm{cool}} \equiv \langle -Q_{\mathrm{rad}}\rangle/(2\pi R)$ and can be interpreted (approximately) as the radiative cooling rate per

unit area, since the transport by radiative diffusion in the radial direction is much weaker when the disc is geometrically thin. We can use the mass equation (E.11) to eliminate a few terms from the angular-momentum and energy equations, (E.12) and (E.13), and rewrite them as

$$\begin{aligned}&\langle R\bar{\rho}\partial_t v_{\phi,0}\rangle + \langle \bar{\rho} v_{R,0}\partial_R(R v_{\phi,0})\rangle \\ &\quad + \partial_R\langle R\overline{\rho v_R} v_{\phi,1} + R(\rho v_R)' v'_\phi + (4\pi \mathrm{G})^{-1} R g'_R g'_\phi\rangle = 0,\end{aligned} \tag{E.14}$$

$$\begin{aligned}&\langle \bar{\rho} v_{R,0}\partial_R e_{\mathrm{tot},0}\rangle + \langle \bar{\rho}\partial_t(u + \tfrac{1}{2}v^2)_0\rangle + \langle \Phi_1\partial_t\bar{\rho} + \Phi'\partial_t\rho'\rangle \\ &\quad + \partial_R\langle \bar{\rho} v_{R,0}(p/\rho)_0 + \overline{\rho v_R} e_{\mathrm{tot},1} + (\rho v_R)' e'_{\mathrm{tot}}\rangle = -2\pi R\Lambda_{\mathrm{cool}}.\end{aligned} \tag{E.15}$$

Because $\dot{M} = -\langle \rho v_R\rangle = -\langle \bar{\rho} v_{R,0}\rangle$, the first term in the energy equation (E.15) is $\mathcal{O}(\epsilon_1)$ and corresponds to the accretion heating we discussed in Section 3.4.1. In the discussion below, we will show that the remaining terms on the left-hand side of equation (E.15) are all $\lesssim\mathcal{O}(\epsilon_1)$ (or, in dimensional form, $\lesssim -g_R\dot{M}$).

E.3 Perturbation amplitude estimates

In order to know the amplitude of each term in the energy equation, we need to first estimate the amplitudes of the vertical variation in the mean disc profile a_1 and the turbulent fluctuations a'.

We start by estimating the vertical variation in the mean disc profile, $\Phi_1, u_1, (p/\rho)_1$ and $v_{\phi,1}$. Φ_1 can be divided into contributions from the central protostar and from disc self-gravity; the former is of order $(H/R)^2\Phi_0 \sim \epsilon_2$, while the latter is of order $H\mathrm{G}\Sigma \sim \epsilon_2^{1/2}\epsilon_3$ (here $\mathrm{G} \sim \epsilon_3$ under our normalization because $\mathrm{G}M_{\mathrm{tot}} \sim 1$ and $M_{\mathrm{tot}} \sim \epsilon_3^{-1}$), which is $\lesssim\epsilon_2$ since $\epsilon_3 \lesssim \epsilon_2^{1/2}$. The relative vertical variation of u and p/ρ can both be order unity, so $u_1, (p/\rho)_1 \sim \bar{u}, \overline{(p/\rho)} \sim \epsilon_2$. Finally, for $v_{\phi,1}$, because it leads to a vertical variation of centrifugal acceleration of $\approx 2v_{\phi,0}v_{\phi,1}/R$ that must be canceled by the vertical variation of gravity and pressure gradient (both are $\sim\epsilon_2$) , we expect $v_{\phi,1} \sim \epsilon_2$. In summary,

$$\Phi_1, u_1, (p/\rho)_1, v_{\phi,1} \sim \epsilon_2. \tag{E.16}$$

Now we estimate $\overline{\rho v_R}$ and $\boldsymbol{v}'$ using the angular momentum equation (E.14). Consider its first term. Since $v_{\phi,0}$ evolves at a timescale of order $M_{\mathrm{tot}}/\dot{M}$, we expect

$$\langle R\bar{\rho}\rangle\partial_t v_{\phi,0} \sim \dot{M}/M_{\mathrm{tot}} \sim \epsilon_1\epsilon_3. \tag{E.17}$$

Then, for the second term, since $\dot{M} = -\langle \bar{\rho} v_{R,0} \rangle$ suggests that $v_{R,0} \sim \epsilon_1$, we have

$$\langle \bar{\rho} v_{R,0} \partial_R (R v_{\phi,0}) \rangle \sim \epsilon_1. \tag{E.18}$$

Given the amplitude of the first two terms, each of the three terms in $\partial_R \langle \dots \rangle$ has to be $\lesssim \epsilon_1$ (assuming that they do not exactly cancel out). Therefore,

$$(\overline{\rho v_R}/\bar{\rho}) v_{\phi,1} \lesssim \epsilon_1, \quad [(\rho v_R)'/\bar{\rho}] v'_\phi \lesssim \epsilon_1. \tag{E.19}$$

Because $v_{\phi,1} \sim \epsilon_2$ (as shown earlier), the first part of equation (E.19) gives

$$\overline{\rho v_R}/\bar{\rho} \lesssim \min(1, \epsilon_1/\epsilon_2). \tag{E.20}$$

Here we cap it to $\mathcal{O}(1)$ because the radial velocity should remain smaller than the rotational velocity. Now consider $(\rho v_R)'$ (which is $\gtrsim \bar{\rho} v'_R$, assuming that the fluctuations in ρ and v_R do not exactly cancel) and v'_ϕ, which corresponds to turbulent fluctuations. Assuming that these turbulent fluctuations are relatively well-correlated (so that $\langle (\rho v_R)' v'_\phi \rangle$ and $\langle |(\rho v_R)'| |v'_\phi| \rangle$ are of the same order of magnitude) and $v'_R \sim v'_\phi \gtrsim v'_z$, the constraint in equation (E.19) gives

$$|\boldsymbol{v}'| \lesssim \epsilon_1^{1/2}. \tag{E.21}$$

This also allows us to estimate $e_{\mathrm{tot},1}$:

$$e_{\mathrm{tot},1} \approx u_1 + (p/\rho)_1 + \Phi_1 + v_{\phi,0} v_{\phi,1} + \tfrac{1}{2} (\boldsymbol{v}' \cdot \boldsymbol{v}')_1 \lesssim \max(\epsilon_1, \epsilon_2). \tag{E.22}$$

Next we consider the amplitudes of ρ' and Φ'. We can assume that the perturbations have characteristic length scale $\sim H$ and characteristic frequency $\sim \Omega$. (This is true for both spiral waves and local turbulence; in the former case the characteristic scale is $\sim H$ in R, z but can be $\sim R$ in ϕ.) Considering the fluctuating component of the continuity equation gives

$$\rho'/\bar{\rho} \sim \Omega^{-1} H^{-1} v' \lesssim \epsilon_1^{1/2} \epsilon_2^{-1/2}. \tag{E.23}$$

Under the normalization we use ($\Sigma, r, \Omega \sim 1$), $\bar{\rho}$ is of order $\Sigma/H \sim \epsilon_2^{-1/2}$ at midplane, and the gravitational constant G is $\mathcal{O}(\epsilon_3)$ because M_{tot} is $\mathcal{O}(\epsilon_3^{-1})$ and $\mathrm{G}M_{\mathrm{tot}}$ is $\mathcal{O}(1)$. The Poisson equation then gives

$$H^{-2} \Phi' \sim \mathrm{G} \rho', \quad \Phi' \lesssim \epsilon_1^{1/2} \epsilon_3. \tag{E.24}$$

Then we consider the amplitude of $(p/\rho)'$ and u'; these two perturbations should have comparable amplitudes. They can come from two sources: compression, which gives $u'/\bar{u}, (p/\rho)'/\overline{(p/\rho)} \sim \rho'/\bar{\rho}$, and non-uniform heating, which at most gives $u', (p/\rho)' \sim \delta\Lambda_{\rm heat}\Omega^{-1}\Sigma^{-1}$ (since the heating rate varies at a characteristic rate of at least Ω). Here $\delta\Lambda_{\rm heat}$ is the amplitude of variation in heating rate, and should satisfy $\delta\Lambda_{\rm heat} \lesssim \Lambda_{\rm heat} \approx \Lambda_{\rm cool} \sim \epsilon_1$.[1] Therefore,

$$u', (p/\rho)' \lesssim \max(\epsilon_1, \epsilon_1^{1/2}\epsilon_2^{1/2}). \tag{E.25}$$

Finally, for $e'_{\rm tot}$ we have

$$e'_{\rm tot} \approx u' + (p/\rho)' + \Phi' + \bar{\boldsymbol{v}} \cdot \boldsymbol{v}' \lesssim \epsilon_1^{1/2}. \tag{E.26}$$

In summary, the amplitudes of the terms in the energy equation (E.15) are:

$$\begin{aligned} &\underbrace{\langle \bar{\rho} v_{R,0} \partial_R e_{\rm tot,0} \rangle}_{\sim\epsilon_1} + \underbrace{\langle \bar{\rho} \partial_t (u + \tfrac{1}{2}v^2)_0 \rangle}_{\sim\epsilon_1\epsilon_2+\epsilon_1\epsilon_3} + \underbrace{\langle \Phi_1 \partial_t \bar{\rho} \rangle}_{\sim\epsilon_1\epsilon_2} + \underbrace{\langle \Phi' \partial_t \rho' \rangle}_{\lesssim\epsilon_1\epsilon_2^{-1/2}\epsilon_3\lesssim\epsilon_1} \\ &+ \partial_R \langle \underbrace{\bar{\rho} v_{R,0} (p/\rho)_0}_{\sim\epsilon_1\epsilon_2} + \underbrace{\overline{\rho v_R} e_{\rm tot,1}}_{\lesssim\epsilon_1} + \underbrace{(\rho v_R)' e'_{\rm tot}}_{\lesssim\epsilon_1} \rangle = -2\pi R \Lambda_{\rm cool}. \end{aligned} \tag{E.27}$$

For the terms inside the operator $\partial_R\langle \dots \rangle$, the amplitudes labelled below them include the contribution from $\partial_R\langle \dots \rangle$. We see that all terms on the left-hand side are indeed $\mathcal{O}(\epsilon_1)$ or less, making ϵ_1 (or $-g_R\dot{M}$) a good order-of-magnitude estimate for the cooling rate required at thermal equilibrium.

We also note that when disc is infinitely thin ($\epsilon_2, \epsilon_3 \to 0$) and all transport is local (i.e., self gravity is negligible), the energy equation consists only of the mean accretion term, $\langle \bar{\rho} v_{R,0} \partial e_{\rm tot,0} \rangle$, and the divergence of a turbulent energy flux, which is simply Ω times the turbulent angular-momentum flux. In that case one can obtain a more precise estimate of the heating rate. (This result still holds when other local angular-momentum transport mechanisms, such as turbulent magnetic transport through Maxwell stress, are present; cf. Balbus et al. 1994.) In our case, the problem is complicated greatly by the presence of terms that transport only energy but no angular momentum (through advection of thermal energy and pressure work) or transport energy and angular momentum non-locally (through self-gravity). In this case, a more precise estimate of the heating rate is hardly possible without additional assumptions on the details of energy and angular-momentum transport mechanisms.

[1] Strictly speaking this is a circular argument, since we cannot assume that heating is $\mathcal{O}(\epsilon_1)$ yet. More rigorously, one could argue that even when $u', (p/\rho)' \sim \Lambda_{\rm heat}\Omega^{-1}\Sigma^{-1}$, terms with $u', (p/\rho)'$ should remain small compared to the cooling term.

Appendix F

Testing the analytic estimate of energy transport by spiral convection

In this appendix we test the accuracy of the analytic estimate of the rate of convective energy transport by spiral convection, equation (3.11). Fig. F.1 compares the vertical energy transport by convection with estimates under different level of approximations. In this figure, $(\delta v_z/H)_{\rm est}$ uses equation (3.9) and $(\delta u/u)_{\rm est}$ uses equation (3.10), where $\Lambda_{\rm heat}$ is estimated with $\Lambda_{\rm cool}$ measured from the simulation.[1] The comparison shows that the analytic estimates in Section 3.4.3 are all fairly accurate inside the disc.

[1]These estimates depend on $\Omega_{\rm s} \equiv |\Omega - \Omega_{\rm p}|$. For the curves in Fig. F.1 the spiral-wave pattern speed $\Omega_{\rm p}$ is measured from the largest spirals in the outer part of the disc. Once the GI region becomes extended (covering a large range in $\log R$), spiral waves in the inner disc are excited more locally ($\Omega_{\rm p} \sim \Omega$), and our choice of $\Omega_{\rm p}$ is no longer correct there. However, this should not significantly affect the value of $\Omega_{\rm s}$, as $\Omega_{\rm s} \sim \Omega$ for the inner disc either way.

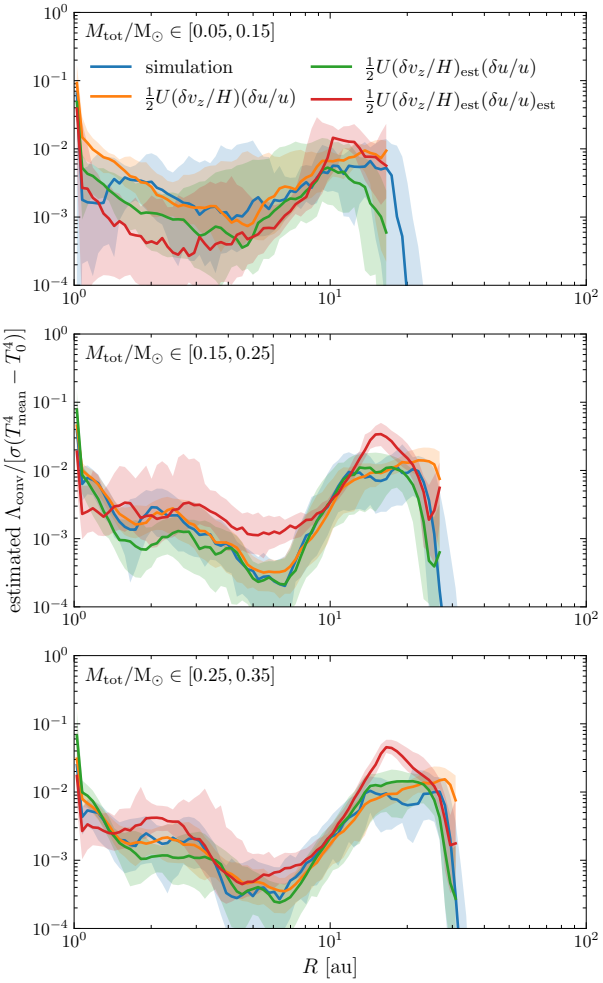

Figure F.1: Vertical energy transport by convection, as measured from the simulation (blue) and estimated under different levels of approximation. The analytic estimates are only plotted for $R \leq R_{\rm d}$. All estimates are fairly good inside the disc.

Appendix G

Are dust grains well-coupled to gas?

In our model we have assumed a constant dust-to-gas ratio throughout the disk; here we check whether this assumption is reasonable. We start by estimating the timescale of radial drift by aerodynamic drag, which is known to cause significant dust decoupling in Class II (e.g., Ansdell et al., 2018). Following Birnstiel et al. (2010), we estimate the radial profile of midplane Stokes number and radial drift timescale (Fig. G.1) for our estimated disk profiles at two different grain sizes, 1 mm and 1 cm. We find that grains as large as $\sim$1 mm are still well-coupled to dust (St $\ll 1$) with a radial drift timescale longer than the typical lifetime of Class 0/I (10^5 yr) for all of our disks. Even for cm-size grains, less than half of our disks have sufficiently short radial timescale to allow substantial drift during Class 0/I.

There are, however, other mechanisms of decoupling that might be more important in Class 0/I. For example, the dynamical timescale in a gravitationally unstable disk could be lower than Ω^{-1} (e.g., due to shocks) and that causes the effective Stocks number to increase and makes grains less well-coupled to the gas. Recent 3D simulations by Baehr et al. (2022) shows a significant degree of clumping at St= 1 (and much weaker clumping at St= 0.1). More significant decoupling could occur if the disk is fragmenting (which produces shorter dynamical timescale) and dust growth in the dense fragment is modeled (which further increases St), as demonstrated in Vorobyov et al. (2019). Magnetic field could also decouple ionized grains from the neutral gas. While this effect could be very significant in the envelope and the outflow, the resulting dust-to-gas variation in the disk, and the systematic difference between disk and pre-stellar dust-to-gas ratio, are generally a factor of a few except for very large (few cm) grains (Lebreuilly et al., 2020). In summary, while we generally should no expect a systematic difference in dust and gas disk size as we do in Class II, several mechanisms could produce enhanced and/or non-uniform dust-to-gas ratio in the disk.

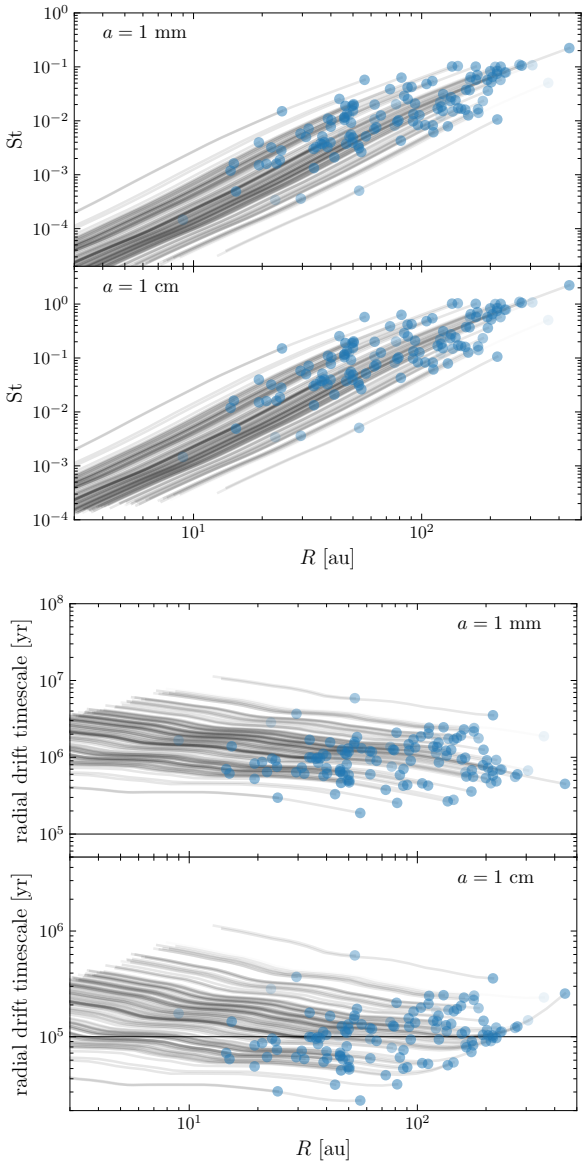

Figure G.1: Left panel: Stokes number profiles for 1 mm and 1 cm grains. Right panel: Timescales of radial drift by aerodynamic drag for 1 mm and 1 cm grains. For 1 mm grains in all disks and 1 cm grains in more than half of the disks, the grains are well-coupled and radial drift by aerodynamic drag should be negligible over the lifetime of Class 0/I.

Appendix H

Systematic uncertainties in fitting observed protostellar disks

In this appendix we evaluate the sensitivity of estimated disk properties on assumed parameters of our model, and estimate the resulting systematic uncertainties.

For a disk property y and a set of assumed model parameter p_i whose log have systematic uncertainties $\sigma(\log p_i)$, the uncertainty in $\log y$ is simply

$$\sigma^2(\log y) = \left\langle \sum_i \left(\frac{\partial \log y_j}{\partial \log p_i} \right)^2 \sigma^2(\log p_i) \right\rangle_j . \tag{H.1}$$

Here y_j is the estimated value of y for the jth system and $\langle \cdot \rangle_j$ denotes an average in j. In order to estimate $\sigma(\log y)$ we first directly evaluate $\partial \log y_j / \partial \log p_i$ for each system by comparing y_j fitted from different choices of p_i, and then assign physically reasonable $\sigma(\log p_i)$ to obtain $\sigma(\log y)$. For our model, the assumed parameters are the max dust grain size $a_{\rm max}$, the ratio between accretion rate and protostar mass $\dot{M}/M_\star$, and the Toomre Q parameter. The first two are highly uncertain, so we assume that their 1σ uncertainties are both a factor of 10, i.e. $\sigma(\log a_{\rm max}) = \sigma(\log \dot{M}/M_\star) = \log(10)$. For Q, since a gravitationally self-regulated

	$R_{\rm d}$	$M_{\rm d}$	$M_\star$	$M_{\rm d}/M_\star$	$T_{\rm mean}$
$a_{\rm max}$	$+0.03 \pm 0.10$	$+0.01 \pm 0.05$	$+0.04 \pm 0.13$	-0.03 ± 0.11	-0.04 ± 0.05
$\dot{M}/M_\star$	-0.03 ± 0.13	-0.16 ± 0.06	-0.45 ± 0.20	$+0.30 \pm 0.22$	-0.01 ± 0.08
Q	-0.24 ± 0.39	-0.96 ± 0.18	$+1.04 \pm 0.52$	-2.01 ± 0.51	$+0.18 \pm 0.20$

Table H.1: Sensitivity of disk property estimates to assumed model parameters. The numbers before and after $\pm$ are mean and standard deviation of $\partial \log(\text{disk property}) / \partial \log(\text{assumed model parameter})$, respectively. These statistics are evaluated on the sub-sample with $\chi^2_{\rm mean} \leq 2$ for our fiducial model.

	$R_{\rm d}$	$M_{\rm d}$	$M_\star$	$M_{\rm d}/M_\star$	$T_{\rm mean}$
$a_{\rm max}$	0.25	0.11	0.31	0.27	0.15
$\dot{M}/M_\star$	0.31	0.39	1.15	0.85	0.19
Q	0.16	0.34	0.40	0.72	0.09
Total	**0.43**	**0.53**	**1.25**	**1.15**	**0.26**

Table H.2: Estimated 1σ uncertainties of log disk property due to each parameter. Here we assume $\sigma(\log a_{\rm max}) = \log(10)$, $\sigma(\log \dot{M}/M_\star) = \log(10)$, and $\sigma(\log Q) = \log(2)/2$. We also assume that the errors in assumed model parameters are uncorrelated. These statistics are evaluated on the sub-sample with $\chi^2_{\rm mean} \leq 2$ for our fiducial model. The uncertainties in disk property estimates are between a factor of $\exp(0.26) \approx 1.3$ and $\exp(1.99) \approx 3$.

disk generally have $Q = 1 - 2$, we let this range correspond to $\pm 1\sigma$ and choose $\sigma(\log Q) = \log(2)/2$.

In Table H.1 and H.2 we summarize the evaluated $\partial \log y / \partial \log p_i$ and the estimated uncertainties of log disk properties $\sigma(\log y)$. The resulting 1σ uncertainties in disk properties range from a factor of 2-7. Therefore, disk properties estimated with our current model should generally be considered as a rough estimate. In the future, these uncertainties can be significantly reduced if data (or theory) can constrain the grain size and the accretion rate (or protostellar mass) better.

Since different disk properties are affected by the same set of parameters, their systematic errors are generally correlated. To capture this correlation, we evaluate the covariance between the systematic error of two different disk properties (y, y') using

$$\mathrm{cov}(\log y, \log y') = \left\langle \sum_i \frac{\partial \log y_j}{\partial \log p_i} \frac{\partial \log y'_j}{\partial \log p_i} \sigma^2(\log p_i) \right\rangle_j . \tag{H.2}$$

This covariance is used for plotting the 2D uncertainties in Figs. 5.10 and 5.11.

Appendix I

Scalings of radial disk profile and disk mass

In this appendix we discuss the origin of the scalings observed in Figs. 5.11 and 5.12. We begin by considering the scalings of Σ and $T_{\rm mid}$. Approximately, our physical constraints are

$$\Sigma \propto T^{1/2}\Omega \tag{I.1}$$

$$\Omega^2\dot{M} \propto \begin{cases} T^4\Sigma^{-1}\kappa_{\rm R}^{-1} & \text{(optically thick)} \\ T^4\Sigma\kappa_{\rm P} & \text{(optically thin)} \end{cases} \tag{I.2}$$

Here the first scaling corresponds to gravitational self-regulation ($Q \sim$constant) and the second corresponds to the thermal budget. $\kappa_{\rm R,P}$ are the average opacities at the given radius and T is the average temperature (which should be comparable to the midplane temperature). This gives

$$\Sigma \propto \begin{cases} R^{-15/7}M_\star^{5/7}\dot{M}^{1/7}\kappa_{\rm R}^{1/7} & \text{(optically thick)} \\ R^{-5/3}M_\star^{5/9}\dot{M}^{1/9}\kappa_{\rm P}^{-1/9} & \text{(optically thin)} \end{cases} \tag{I.3}$$

$$T \propto \begin{cases} R^{-9/7}M_\star^{3/7}\dot{M}^{2/7}\kappa_{\rm R}^{2/7} & \text{(optically thick)} \\ R^{-1/3}M_\star^{1/9}\dot{M}^{2/9}\kappa_{\rm P}^{-2/9} & \text{(optically thin)} \end{cases} \tag{I.4}$$

Here we have assumed that $M(<R) \sim M_\star$, which is a reasonable order-of-magnitude approximation since we never have $M_{\rm d}/M_\star \gg 1$ in our sample. The scalings of Σ, T with respect to R are consistent with those observed in Fig. 5.12. (Here $\kappa_{R/P}$ should also vary in radius, but since Σ and T depend weakly on $\kappa_{R/P}$, we can ignore this dependence.)

Now we consider the scaling of disk mass. The disk mass is simply given by

$$M_{\rm d} = \int 2\pi\Sigma R^2 {\rm d}\log R. \tag{I.5}$$

From Eq. I.3 we see that the ΣR^2 scales with $M_\star$ at a slope between $5/7 = 0.71$ and $5/9 = 0.56$, which is consistent with the empirical scaling of $M_{\rm d} \propto M_\star^{0.6}$. ΣR^2 is much less sensitive to other parameters (including R), and that explains why $M_{\rm d}$ shows such tight correlation with $M_\star$. Besides, the weak dependence of ΣR^2 on R means that the relatively steep correlation between $M_{\rm d}$ and $R_{\rm d}$ in Fig. 5.10 for larger disks is not just a generic feature of gravitationally self-regulated disks.

Bibliography

Alves J. F., Lada C. J., Lada E. A., 2001, Nature, 409, 159

Andersen B. C., et al., 2019, ApJ, 873, 54

Anderson D. E., Cleeves L. I., Blake G. A., Bergin E. A., Zhang K., Carpenter J. M., Schwarz K. R., 2022, arXiv e-prints, p. arXiv:2202.00709

Andre P., Ward-Thompson D., Barsony M., 1993, ApJ, 406, 122

Andre P., Ward-Thompson D., Barsony M., 2000, in Mannings V., Boss A. P., Russell S. S., eds, Protostars and Planets IV. p. 59 (`arXiv:astro-ph/9903284`)

André P., et al., 2010, A&A, 518, L102

Andrews S. M., Rosenfeld K. A., Kraus A. L., Wilner D. J., 2013, ApJ, 771, 129

Ansdell M., et al., 2016, ApJ, 828, 46

Ansdell M., et al., 2018, ApJ, 859, 21

Armitage P. J., 2011, ARA&A, 49, 195

Auddy S., Myers P. C., Basu S., Harju J., Pineda J. E., Friesen R. K., 2019, ApJ, 872, 207

Babcock H. W., Cowling T. G., 1953, MNRAS, 113, 357

Baehr H., Zhu Z., Yang C.-C., 2022, arXiv e-prints, p. arXiv:2204.13310

Balbus S. A., Hawley J. F., 1998, Rev. Mod. Phys., 70, 1

Balbus S. A., Gammie C. F., Hawley J. F., 1994, MNRAS, 271, 197

Barranco J. A., Goodman A. A., 1998, ApJ, 504, 207

Basu S., 1997, ApJ, 485, 240

Basu S., Mouschovias T. Ch., 1994, ApJ, 432, 720

Basu S., Mouschovias T. Ch., 1995, ApJ, 453, 271

Bate M. R., 2018, MNRAS, 475, 5618

Béthune W., Latter H., Kley W., 2021, A&A, 650, A49

Birnstiel T., Dullemond C. P., Brauer F., 2010, A&A, 513, A79

Birnstiel T., Ormel C. W., Dullemond C. P., 2011, A&A, 525, A11

Birnstiel T., et al., 2018, ApJ, 869, L45

Bjerkeli P., van der Wiel M. H. D., Harsono D., Ramsey J. P., Jørgensen J. K., 2016, Nature, 540, 406

Bodenheimer P., Yorke H. W., Rozyczka M., Tohline J. E., 1990, ApJ, 355, 651

Booth A. S., Walsh C., Ilee J. D., Notsu S., Qi C., Nomura H., Akiyama E., 2019, ApJ, 882, L31

Braiding C. R., Wardle M., 2012, MNRAS, 422, 261

Caselli P., Benson P. J., Myers P. C., Tafalla M., 2002, ApJ, 572, 238

Chen C.-Y., Ostriker E. C., 2014, ApJ, 785, 69

Chen C.-Y., Ostriker E. C., 2018, ApJ, 865, 34

Ciolek G. E., Königl A., 1998, ApJ, 504, 257

Clarke C. J., 2009, MNRAS, 396, 1066

Commerçon B., Hennebelle P., Audit E., Chabrier G., Teyssier R., 2008, A&A, 482, 371

Contopoulos I., Ciolek G. E., Königl A., 1998, ApJ, 504, 247

Crutcher R. M., 1999, ApJ, 520, 706

Crutcher R. M., Nutter D. J., Ward-Thompson D., Kirk J. M., 2004, ApJ, 600, 279

D'Alessio P., Calvet N., Hartmann L., 1997, ApJ, 474, 397

Dapp W. B., Basu S., Kunz M. W., 2012, A&A, 541, A35

Desch S. J., Mouschovias T. C., 2001, ApJ, 550, 314

Desch S. J., Turner N. J., 2015, ApJ, 811, 156

Draine B. T., 2006, ApJ, 636, 1114

Eisner J. A., et al., 2018, ApJ, 860, 77

Evans Neal J. I., et al., 2009, ApJS, 181, 321

Falgarone E., Troland T. H., Crutcher R. M., Paubert G., 2008, A&A, 487, 247

Fiedler R. A., Mouschovias T. Ch., 1993, ApJ, 415, 680

Fischer W. J., et al., 2017, ApJ, 840, 69

Fischer W. J., Hillenbrand L. A., Herczeg G. J., Johnstone D., Kóspál Á., Dunham M. M., 2022, arXiv e-prints, p. arXiv:2203.11257

Fromang S., Balbus S. A., Terquem C., De Villiers J.-P., 2004, ApJ, 616, 364

Galli D., Shu F. H., 1993, ApJ, 417, 220

Galli D., Lizano S., Shu F. H., Allen A., 2006, ApJ, 647, 374

Galván-Madrid R., Liu H. B., Izquierdo A. F., Miotello A., Zhao B., Carrasco-González C., Lizano S., Rodríguez L. F., 2018, ApJ, 868, 39

Gammie C. F., 1996, ApJ, 457, 355

Gammie C. F., 2001, ApJ, 553, 174

Girart J. M., Rao R., Marrone D. P., 2006, Science, 313, 812

Goodman A. A., Barranco J. A., Wilner D. J., Heyer M. H., 1998, ApJ, 504, 223

Greene T. P., Wilking B. A., Andre P., Young E. T., Lada C. J., 1994, ApJ, 434, 614

Guillet V., Hennebelle P., Pineau des Forêts G., Marcowith A., Commerçon B., Marchand P., 2020, A&A, 643, A17

Gullbring E., Hartmann L., Briceño C., Calvet N., 1998, ApJ, 492, 323

Hennebelle P., Commerçon B., Lee Y.-N., Charnoz S., 2020, A&A, 635, A67

Hernández J., Hartmann L., Calvet N., Jeffries R. D., Gutermuth R., Muzerolle J., Stauffer J., 2008, ApJ, 686, 1195

Hirano S., Tsukamoto Y., Basu S., Machida M. N., 2020, ApJ, 898, 118

Hirashita H., 2012, MNRAS, 422, 1263

Hubeny I., 1990, ApJ, 351, 632

Hull C. L. H., et al., 2013, ApJ, 768, 159

Hull C. L. H., et al., 2014, ApJS, 213, 13

Hunter C., 1977, ApJ, 218, 834

Indriolo N., et al., 2015, ApJ, 800, 40

Ishimaru A., 1978, Wave propagation and scattering in random media. Volume 1 - Single scattering and transport theory. Vol. 1, doi:10.1016/B978-0-12-374701-3.X5001-7,

Jijina J., Myers P. C., Adams F. C., 1999, ApJS, 125, 161

Joos M., Hennebelle P., Ciardi A., 2012, A&A, 543, A128

Joos M., Hennebelle P., Ciardi A., Fromang S., 2013, A&A, 554, A17

Kawasaki Y., Koga S., Machida M. N., 2021, MNRAS, 504, 5588

Kawasaki Y., Koga S., Machida M. N., 2022, MNRAS, 515, 2072

Kimura K., Hosokawa T., Sugimura K., 2021, ApJ, 911, 52

Kirk J. M., Ward-Thompson D., André P., 2005, MNRAS, 360, 1506

Königl A., 1987, ApJ, 320, 726

Könyves V., et al., 2015, A&A, 584, A91

Krasnopolsky R., Königl A., 2002, ApJ, 580, 987

Krasnopolsky R., Li Z.-Y., Shang H., 2011, ApJ, 733, 54

Krasnopolsky R., Li Z.-Y., Shang H., Zhao B., 2012, ApJ, 757, 77

Kratter K., Lodato G., 2016, ARA&A, 54, 271

Kristensen L. E., et al., 2012, A&A, 542, A8

Krumholz M., Burkert A., 2010, ApJ, 724, 895

Kuffmeier M., Haugbølle T., Nordlund Å., 2017, ApJ, 846, 7

Kunz M. W., Mouschovias T. Ch., 2009, ApJ, 693, 1895

Kunz M. W., Mouschovias T. Ch., 2010, MNRAS, 408, 322

Kuznetsova A., Hartmann L., Heitsch F., 2020, ApJ, 893, 73

Lada C. J., 1987, in Peimbert M., Jugaku J., eds, Vol. 115, Star Forming Regions. p. 1

Larson R. B., 1969, MNRAS, 145, 271

Lebreuilly U., Commerçon B., Laibe G., 2020, A&A, 641, A112

Levermore C. D., Pomraning G. C., 1981, ApJ, 248, 321

Li Z.-Y., McKee C. F., 1996, ApJ, 464, 373

Li Z.-Y., Krasnopolsky R., Shang H., 2011, ApJ, 738, 180

Li Z.-Y., Krasnopolsky R., Shang H., 2013, ApJ, 774, 82

Li J. I.-H., Liu H. B., Hasegawa Y., Hirano N., 2017, ApJ, 840, 72

Lin D. N. C., Pringle J. E., 1987, MNRAS, 225, 607

Lin D. N. C., Pringle J. E., 1990, ApJ, 358, 515

Lubow S. H., Spruit H. C., 1995, ApJ, 445, 337

Machaieie D. A., Vilas-Boas J. W., Wuensche C. A., Racca G. A., Myers P. C., Hickel G. R., 2017, ApJ, 836, 19

Machida M. N., Basu S., 2019, ApJ, 876, 149

Machida M. N., Basu S., 2020, MNRAS, 494, 827

Machida M. N., Hosokawa T., 2013, MNRAS, 431, 1719

Machida M. N., Inutsuka S.-i., Matsumoto T., 2006, ApJ, 647, L151

Machida M. N., Inutsuka S.-i., Matsumoto T., 2007, ApJ, 670, 1198

Machida M. N., Inutsuka S.-i., Matsumoto T., 2014, MNRAS, 438, 2278

Machida M. N., Higuchi K., Okuzumi S., 2018, MNRAS, 473, 3080

Marchand P., Masson J., Chabrier G., Hennebelle P., Commerçon B., Vaytet N., 2016, A&A, 592, A18

Masson J., Chabrier G., Hennebelle P., Vaytet N., Commerçon B., 2016, A&A, 587, A32

Mathis J. S., Rumpl W., Nordsieck K. H., 1977, ApJ, 217, 425

Maury A. J., et al., 2018, MNRAS, 477, 2760

Maury A. J., et al., 2019, A&A, 621, A76

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

Meyer C. D., Balsara D. S., Aslam T. D., 2014, J. Comput. Phys., 257, 594

Miyake K., Nakagawa Y., 1993, Icarus, 106, 20

Mouschovias T. C., 1991, ApJ, 373, 169

Mouschovias T. C., Ciolek G. E., 1999, in Lada C. J., Kylafis N. D., eds, NATO Advanced Study Institute (ASI) Series C Vol. 540, The Origin of Stars and Planetary Systems. p. 305

Mouschovias T. C., Paleologou E. V., 1979, ApJ, 230, 204

Mouschovias T. Ch., Paleologou E. V., 1980, ApJ, 237, 877

Mouschovias T. Ch., Paleologou E. V., 1986, ApJ, 308, 781

Mouschovias T. Ch., Spitzer L. J., 1976, ApJ, 210, 326

Myers P. C., Benson P. J., 1983, ApJ, 266, 309

Natta A., 1993, ApJ, 412, 761

Neufeld D. A., Wolfire M. G., 2017, ApJ, 845, 163

Offner S. S. R., McKee C. F., 2011, ApJ, 736, 53

Ormel C. W., Paszun D., Dominik C., Tielens A. G. G. M., 2009, A&A, 502, 845

Ossenkopf V., 1993, A&A, 280, 617

Pattle K., Fissel L., Tahani M., Liu T., Ntormousi E., 2022, arXiv e-prints, p. arXiv:2203.11179

Penston M. V., 1969, MNRAS, 144, 425

Pineda J. E., et al., 2012, A&A, 544, L7

Pollack J. B., Hollenbach D., Beckwith S., Simonelli D. P., Roush T., Fong W., 1994, ApJ, 421, 615

Powell D., Murray-Clay R., Pérez L. M., Schlichting H. E., Rosenthal M., 2019, ApJ, 878, 116

Rafikov R. R., 2007, ApJ, 662, 642

Rafikov R. R., 2009, ApJ, 704, 281

Rice W. K. M., Mayo J. H., Armitage P. J., 2010, MNRAS, 402, 1740

Riols A., Latter H., 2018, MNRAS, 476, 5115

Riols A., Xu W., Lesur G., Kunz M. W., Latter H., 2021, MNRAS, 506, 1407

Rossi S. C. F., Benevides-Soares P., Barbuy B., 1991, A&A, 251, 587

Rowther S., Meru F., Kennedy G. M., Nealon R., Pinte C., 2020, ApJ, 904, L18

Sadavoy S. I., et al., 2019, ApJS, 245, 2

Segura-Cox D. M., et al., 2018, ApJ, 866, 161

Seifried D., Banerjee R., Pudritz R. E., Klessen R. S., 2012, MNRAS, 423, L40

Semenov D., Henning T., Helling C., Ilgner M., Sedlmayr E., 2003, A&A, 410, 611

Shakura N. I., Sunyaev R. A., 1973, Astronomy and Astrophysics, 24, 337

Sheehan P. D., Eisner J. A., 2017, ApJ, 851, 45

Sheehan P. D., Tobin J. J., Federman S., Megeath S. T., Looney L. W., 2020, ApJ, 902, 141

Sheehan P. D., Tobin J. J., Looney L. L., Megeath S. T., 2022, arXiv e-prints, p. arXiv:2203.00029

Shu F. H., 1977, ApJ, 214, 488

Shu F. H., Adams F. C., Lizano S., 1987, ARA&A, 25, 23

Simon J. B., Hawley J. F., Beckwith K., 2009, ApJ, 690, 974

Snow T. P. J., 1975, ApJ, 202, L87

Stammler S. M., Birnstiel T., 2022, arXiv e-prints, p. arXiv:2207.00322

Stephens I. W., et al., 2013, ApJ, 769, L15

Stone J. M., Tomida K., White C. J., Felker K. G., 2020, ApJS, 249, 4

Tassis K., 2007, MNRAS, 379, L50

Tassis K., Mouschovias T. C., 2005, ApJ, 618, 783

Tassis K., Mouschovias T. Ch., 2007, ApJ, 660, 388

Teixeira P. S., Lada C. J., Alves J. F., 2005, ApJ, 629, 276

Tobin J., 2019a, ALMA 870 micron Continuum Measurement Sets, doi:10.7910/DVN/5NXMLI, https://doi.org/10.7910/DVN/5NXMLI

Tobin J., 2019b, VLA Ka-band (9 mm) Continuum - A-configuration (0.08 arcsec), doi:10.7910/DVN/ICKYX0, https://doi.org/10.7910/DVN/ICKYX0

Tobin J. J., et al., 2016, ApJ, 818, 73

Tobin J. J., et al., 2020, ApJ, 890, 130

Tobin J. J., et al., 2022, ApJ, 925, 39

Tohline J. E., Hachisu I., 1990, ApJ, 361, 394

Tomida K., Okuzumi S., Machida M. N., 2015, ApJ, 801, 117

Tomida K., Machida M. N., Hosokawa T., Sakurai Y., Lin C. H., 2017, ApJ, 835, L11

Tripathi A., Andrews S. M., Birnstiel T., Wilner D. J., 2017, ApJ, 845, 44

Tsukamoto Y., Iwasaki K., Okuzumi S., Machida M. N., Inutsuka S., 2015, ApJ, 810, L26

Tsukamoto Y., Okuzumi S., Iwasaki K., Machida M. N., Inutsuka S., 2018, ApJ, 868, 22

Turner N. J., Fromang S., Gammie C., Klahr H., Lesur G., Wardle M., Bai X. N., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. p. 411 (arXiv:1401.7306), doi:10.2458/azu_uapress_9780816531240-ch018

Tychoniec Ł., et al., 2018, ApJS, 238, 19

Ulrich R. K., 1976, ApJ, 210, 377

Umebayashi T., 1983, Progress of Theoretical Physics, 69, 480

Umebayashi T., Nakano T., 1980, PASJ, 32, 405

Vaytet N., Commerçon B., Masson J., González M., Chabrier G., 2018, A&A, 615, A5

Verliat A., Hennebelle P., Maury A. J., Gaudel M., 2020, A&A, 635, A130

Vorobyov E. I., Basu S., 2006, ApJ, 650, 956

Vorobyov E. I., Basu S., 2007, MNRAS, 381, 1009

Vorobyov E. I., DeSouza A. L., Basu S., 2013, ApJ, 768, 131

Vorobyov E. I., Skliarevskii A. M., Elbakyan V. G., Pavlyuchenkov Y., Akimkin V., Guedel M., 2019, A&A, 627, A154

Ward-Thompson D., Scott P. F., Hills R. E., Andre P., 1994, MNRAS, 268, 276

Ward-Thompson D., Kirk J. M., Crutcher R. M., Greaves J. S., Holland W. S., André P., 2000, ApJ, 537, L135

Ward-Thompson D., André P., Crutcher R., Johnstone D., Onishi T., Wilson C., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 33 (arXiv:astro-ph/0603474)

Wardle M., 2004, Ap&SS, 292, 317

Wardle M., Ng C., 1999, MNRAS, 303, 239

Whitehouse S. C., Bate M. R., 2006, MNRAS, 367, 32

Williams J. P., Cieza L. A., 2011, ARA&A, 49, 67

Williams J. P., Cieza L., Hales A., Ansdell M., Ruiz-Rodriguez D., Casassus S., Perez S., Zurlo A., 2019, ApJ, 875, L9

Woitke P., et al., 2016, A&A, 586, A103

Woolfson M. M., 1993, QJRAS, 34, 1

Wurster J., Bate M. R., 2019a, arXiv e-prints, p. arXiv:1906.12276

Wurster J., Bate M. R., 2019b, MNRAS, 486, 2587

Wurster J., Bate M. R., Price D. J., 2019, MNRAS, 489, 1719

Xu W., Kunz M. W., 2021a, MNRAS, 502, 4911

Xu W., Kunz M. W., 2021b, MNRAS, 508, 2142

Zakri W., et al., 2022, ApJ, 924, L23

Zamponi J., Maureira M. J., Zhao B., Liu H. B., Ilee J. D., Forgan D., Caselli P., 2021, MNRAS, 508, 2583

Zhao B., Caselli P., Li Z.-Y., Krasnopolsky R., 2018a, MNRAS, 473, 4868

Zhao B., Caselli P., Li Z.-Y., 2018b, MNRAS, 478, 2723

Zhao B., et al., 2020a, Space Sci. Rev., 216, 43

Zhao B., Caselli P., Li Z.-Y., Krasnopolsky R., Shang H., Lam K. H., 2020b, MNRAS, 492, 3375

Zhu Z., Hartmann L., Gammie C. F., Book L. G., Simon J. B., Engelhard E., 2010, ApJ, 713, 1134

Zhu Z., et al., 2019, The Astrophysical Journal Letters, 877, L18

di Francesco J., Evans N. J. I., Caselli P., Myers P. C., Shirley Y., Aikawa Y., Tafalla M., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 17 (`arXiv:astro-ph/0602379`)