Planets in turbulent disks

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Alexandros Ziampras

aus Thessaloniki, Griechenland

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	08.04.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Klaus Werner
2. Berichterstatter/-in:	Prof. Dr. Richard Nelson
3. Berichterstatter/-in:	Prof. Dr. Hubert Klahr

Zusammenfassung

Akkretionsscheiben sind die Geburtsstätte von Planeten, deren Eigenschaften und Entwicklung empfindlich von der der Thermo- und Hydrodynamik von der Gasscheibe beeinflusst werden, in der sie wachsen. Das Verständnis der Prozesse, die die Gasstrukturen der Scheibe definieren, und der Mechanismen, die die Akkretion verursachen, ist essenziell um Interaktionen zwischen Planet und Scheiben mit numerischen Simulationen zu modellieren.

Über den Verlauf dieser Arbeit betrachten wir vielfältige Aspekte der protoplanetaren Scheibendynamik. Wir untersuchen den Einfluss von unterschiedlichen Strahlungseffekten auf die Interaktion von Planet und Scheibe, ergründen das Verhalten von Wirbel in der Gasscheibe durch den Planet und überprüfen unterschiedliche, physikalische Eigenschaften der vertikalen Scherinstabilität (VSI) als einen Akkretions treibenden Mechanismus in protoplanetaren Scheiben. Hierfür führen wir numerischen Hydrodynamische Simulationen von Akkretionsscheiben mit und ohne eingebettete Planeten durch mit einer großen Bandbreite von physikalischen und numerischen Parametern und vielfältigen Nachbearbeitungstechniken um zu validieren, zu vergleichen und die Resultate unserer Modele zu verstehen.

Unsere Erkenntnis ist, dass die Zustandsgleichung eine zentrale Wichtigkeit hat für die Modellierung der Entstehung von Ringen und Leerräumen in der Gasscheibe durch Planeten. In diesem Rahmen führen wir den Einfluss von verschiedenen Strahlungsmechanismen sowohl auf das Profil der Dichte und Temperatur der Scheibe, als auch auf die Lebensdauer von Wirbel vor, die von solchen Planeten erzeugt werden. Bezüglich der VSI demonstrieren wir, dass sie einen konkurrenzfähigen Kandidat zur Erklärung der beobachteten Akkretion in protoplanetaren Scheiben darstellt. Und wir identifizieren Bedingungen, unter denen der Planet die turbulenten Spannung durch die VSI unterdrückt oder mit ihr koexistiert. Schließlich zeigt sich, dass die Spannung erzeugt durch Wirbel und Spiralarme -wenn gleich hilfreich für die Akkretion- die VSI schwächen und abflachen kann und dadurch das vertikale Mischvermögen limitiert.

Unsere Resultate unterstreichen, dass der adäquate Einsatz von Strahlungseffekten in numerischen Modellen der Interaktion zwischen Planet und Scheibe entscheidend ist. Diese Arbeit legt außerdem nahe, dass es unwahrscheinlich ist Signaturen der VSI in der Gegenwart von schweren Planeten zu beobachten, aber potentiell in der Zukunft während der frühen Planetenentstehungsphasen.

Abstract

Accretion disks are the birthplace of planets, the properties and evolution of which are highly sensitive to the thermo- and hydrodynamics of the gaseous disk that they are embedded in during their growth. Understanding the processes that define the structure of gas in the disk and the mechanisms that drive accretion is essential to modeling planet–disk interaction with numerical simulations.

Over the course of this project, we look into various aspects of protoplanetary disk dynamics. We examine the impact of different radiative effects on planet–disk interaction, explore the behavior of planet-generated vortices, and investigate the physical properties of the vertical shear instability (VSI) as an accretion-driving mechanism in protoplanetary disks. To that end, we perform numerical hydrodynamics simulations of accretion disks with and without embedded planets, using a wide variety of physical and numerical parameters and various post-processing techniques in order to validate, compare, and understand the results of our models.

We find that the equation of state is of key importance to modeling the formation of rings and gaps by planets, and highlight the impact of different radiative mechanisms on the density and temperature profile of the disk as well as on the lifetime of vortices generated by such planets. Regarding the VSI, we show that it is a competitive candidate in interpreting accretion in observations of protoplanetary disks and identify conditions for which a planet can coexist with or suppress the turbulent stress that this mechanism can generate. Finally, we find that the stress generated by vortices and spiral arms, while conducive to accretion, can weaken or quench the VSI and therefore limit its vertical mixing capacity.

Our results outline that appropriate treatment of radiative effects is crucial in numerical models of planet–disk interaction. Our work also suggests that observing VSI signatures in the presence of massive planets is unlikely, but perhaps possible in the future and during earlier stages of planet formation.

Contents

1 Introduction

From time immemorial, humanity has gazed at the countless lights in the night sky and has put great effort in its attempts to rationalize its observations. Celestial bodies visible by the naked eye have been observed by ancient civilizations that date back to the Babylonians in the second millennium BCE (Sachs, 1974). The ancient Greeks interpreted the celestial bodies within our solar system as "wandering stars", as they would "drift" in the sky relative to the stationary, distant stars in the background. Since then, the ancient Greek word for "wanderer" (*"πλανήτης", planētēs*) has been used to refer to these celestial objects, known as *planets*.

The formation of stars and planets is thought to begin with the gravitational collapse of a massive molecular cloud of primarily hydrogen gas. A young stellar core is formed at the center of the cloud, while conservation of angular momentum flattens the remaining gas into a disk that continues to accrete onto the forming protostar over a few Myr (Haisch et al., 2001). The remnants of the disk are finally blown away via radiation pressure by the fully-formed star, and the latter continues to evolve by burning its gas fuel via nuclear fusion until the end of its life. A sketch of this process is shown in Fig. 1.

During the early stages of the universe, this process would form massive stars with very short lifespans that would convert hydrogen and helium into heavier elements such as carbon and silicon, before ejecting them into space at the end of their lives and enriching future gas clouds with a relatively small amount of tiny dust grains formed by complex chains containing these heavier elements. Later generations of molecular clouds and resulting accretion disks then contain approximately 1% of their mass in µm-sized dust particles (Mathis et al., 1977), which can aggregate into mm-sized grains and eventually into up to meter-sized *pebbles*, km-sized *planetesimals* or even 100-km-sized *planetary embryos* during the disk phase of stellar evolution. These embryos can then grow into young *protoplanets* by accreting pebbles, planetesimals and/or gas during the disk phase, or through gravitational interaction with other embryos once the disk has dispersed. Several Myr later, after the system has stabilized, what remains is a *planetary system* that consists of at least one central, hydrogen-burning star and one or more planets. This theory, which was pioneered by Kant (1755) and later adjusted by Laplace (1796), forms the *Kant–Laplace nebular hypothesis* and was confirmed much later by observations of forming star systems, thus constituting the basis of the most widely-accepted theory of planet formation with our current knowledge.

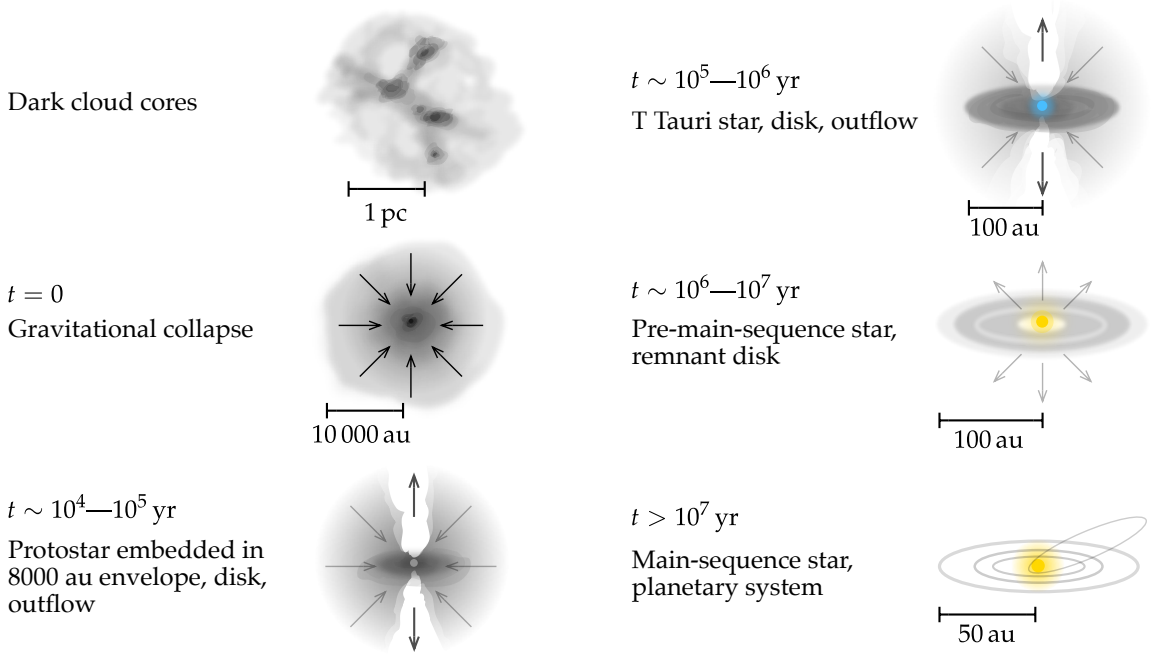

Figure 1. Sketch of the various stages of star and planet formation. Illustration by André Oliva, adapted from Casasola (2008).

Observations of protoplanetary disks (PPDs) date back to the measurement of the spectral distribution of energy (SED, left panel of Fig. 2) of young stellar objects (YSOs). While stars mostly emit in the UV and optical bands of the electromagnetic spectrum, the presence of dust grains within the accretion disk emitting in the μm–mm range results in an observable excess in the emission spectrum of YSOs in the infrared (IR). As a result, the first observations of PPDs were only indirect, through spatially-unresolved SEDs alone.

As the angular resolution of telescopes improved, it was eventually possible to observe the structure of YSOs in the optical spectrum as a system of a star and its accretion disk rather than as just a point source, as the optically thick dust distribution would appear as a dark region around the star (middle panel of Fig. 2). From the 1990s, observations of so-called *proplyds* ("**pro**to**pl**anetar**y** **d**isk**s**") would become more common as detection techniques and detector technology evolved rapidly. This development led up to the modern age of mm interferometry, with instruments such as the Very Large Array (VLA) and the Atacama Large Millimeter Array (ALMA) providing high-resolution, high-fidelity observations of PPDs and revealing features such as rings, gaps, and non-axisymmetries in their structure (right panel of Fig. 2).

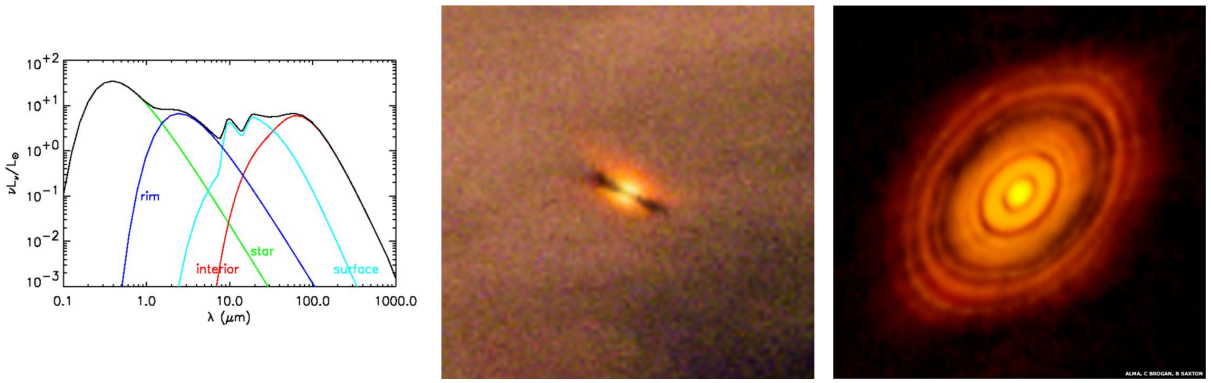

Figure 2. Evolution of observations of protoplanetary disks over the years. Left: a sample spectral energy distribution (SED) showing the contribution of the star (green) and the characteristic IR excess of the different regions of the disk (blue, cyan, red) Image credit: slides by Cornelis P. Dullemond. Middle: edge-on direct observation of a protoplanetary disk around the YSO 473-245 in the Orion Nebula with the Hubble Space Telescope. Image Credit: NASA/ESA and L. Ricci (ESO). Right: the rich radial structure of the system HL Tau, observed in the IR with the ALMA instrument (ALMA Partnership et al., 2015).

Planets, on the other hand, are incredibly faint compared to the luminosity of their central star. As a result they are typically observed indirectly via techniques such as the radial velocity method (RV, left panel of Fig. 3), where the planet's gravitational pull on the star causes a Doppler shift in the star's emission spectrum, or the transit method (middle panel of Fig. 3), where the passage of a planet between an observer and the star results in a minuscule but periodic drop in the star's observed brightness. It was in fact the RV method that led to the discovery of the first exoplanet by Mayor et al. (1995), earning them a share of the Nobel Prize in Physics in 2019. At the moment of submission of this dissertation, 5009 exoplanets have been confirmed through various methods[1].

Modern instruments such as the VLA and ALMA have revolutionized the field of planet formation not only by constraining the properties of protoplanetary disks (e.g., the DSHARP survey, see Andrews et al., 2018), but also by achieving the first observation of a planet embedded in an accretion disk (Keppler et al., 2018; Haffert et al., 2019, right panel of Fig. 3). The datasets by these surveys have provided constraints to theoretical models of the evolution of accretion disks and planet formation, but at the same time highlighted the need for a better understanding of the underlying mechanisms of accretion and radiation transport that these models hinge on.

The rapid advancement of technology has also enabled theorists to explore complex models of accretion disks, planet–disk and planet–planet interaction using numerical simulations of nonlinear, coupled systems of equations that cannot be approached analytically beyond a linear approximation. Turbulence

[1]NASA exoplanet archive, accessed on June 3, 2022. https://exoplanetarchive.ipac.caltech.edu/

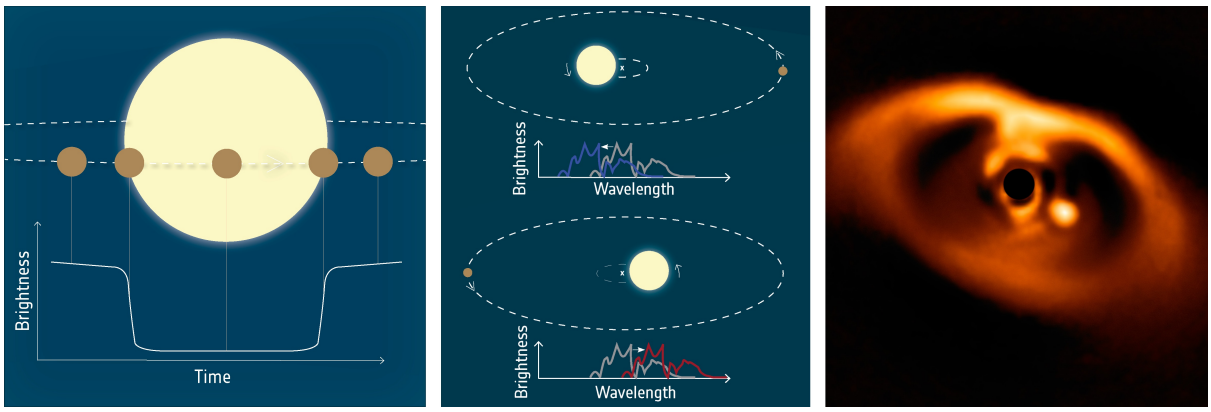

Figure 3. Different techniques used in observations of exoplanets. Left: transit photometry, which tracks the periodic dimming of a source due to an occulting companion. Middle: the radial velocity method, which relies on a blue- or red-shift of the star's spectrum as the star–planet system orbit their common center of mass. Right: direct imaging of an embedded planet in the system PDS 70 (Keppler et al., 2018). Image credit for left and center panels: ESO (2019).

as an accretion-driving mechanism, the random emission of thermalized dust grains as a means of radiation transport, the nonlinear angular momentum exchange between a massive planet and the disk, and the chaotic interaction among planets in a planetary system are very computationally expensive problems that can nowadays be approximately solved for specific, well-defined initial conditions and physical parameters. Motivating these models, understanding their limitations and interpreting their results appropriately is crucial to bridging the gap between our current theory and observations of planet formation.

With this project, we aim to investigate the thermo- and hydrodynamics of planet–disk interaction using analytical and numerical models of protoplanetary disks in the context of planet formation. We approach this by modeling various radiative effects using different degrees of approximating assumptions, and studying their impact on the formation of rings, gaps, and vortices by embedded planets. Furthermore, we analyze the behavior of the vertical shear instability, a mechanism that could possibly explain accretion in protoplanetary disks, and examine its interplay with an embedded planet. The key questions we aim to answer are:

How important are radiative effects in the formation of planet-generated features?
and
Can the vertical shear instability coexist with a planet in an accretion disk?

We introduce our theoretical framework in Sect. 2 and our numerical setup in Sect. 3. We then present an overview of our published results and a short discussion of their implications in Sects. 4–6. We present our analysis of the VSI in Sects. 7 and 8, while also discussing our results. Finally, we summarize the key points of our VSI analysis in Sect. 9, and discuss their implications.

2 Theory

In this section we outline the hydrodynamics of a protoplanetary disk and describe possible modifications to the energy equation and their impact on disk thermodynamics. In addition, we provide a brief summary of the mechanism behind the vertical shear instability, and an introduction to planet–disk interaction. For further details see Armitage (2009), Kley and Nelson (2012), and Nelson et al. (2013).

2.1 Protoplanetary disk hydrodynamics

We assume an ideal gas with mass density ρ, pressure P, specific internal energy ϵ and a velocity field $\boldsymbol{v}$. The gas is orbiting around a star with mass $M_\star$ and luminosity $L_\star$. The Navier–Stokes equations for such a gas then read as follows:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) &= 0, \\ \frac{\partial (\rho \boldsymbol{v})}{\partial t} + \nabla \cdot (\rho \boldsymbol{v} \otimes \boldsymbol{v}) &= -\nabla P - \rho \nabla \Phi + \nabla \cdot \overline{\sigma}, \\ \frac{\partial (\rho \epsilon)}{\partial t} + \boldsymbol{v} \cdot \nabla (\rho \epsilon) &= -\gamma \rho \epsilon \nabla \cdot \boldsymbol{v} + (\overline{\sigma} \cdot \nabla) \cdot \boldsymbol{v} + S, \end{aligned} \tag{2.1}$$

where $\overline{\sigma}$ is the viscous stress tensor (e.g., Tassoul, 1978), $\Phi = \Phi_\star = -\mathrm{G}M_\star/r$ is the gravitational potential due to the star at distance r, and S encompasses any additional radiative source terms, which we will discuss in Sect. 2.2. The gravitational constant is denoted with G. We have closed this set of equations by defining the gas pressure as $P = (\gamma - 1)\rho\epsilon = \mathcal{R}\rho T/\mu$ through the ideal gas assumption. Here, T is the gas temperature, γ is the adiabatic index, μ the mean molecular weight, and $\mathcal{R}$ the gas constant, respectively. Through the above we can also define the isothermal sound speed $c_\mathrm{s} = \sqrt{P/\rho}$, which relates to the adiabatic sound speed c_s^ad such that $c_\mathrm{s}^\mathrm{ad} = \sqrt{\gamma} c_\mathrm{s}$.

In a cylindrical coordinate system $\{R, \varphi, z\}$ where $r = \sqrt{R^2 + z^2}$ is the spherical radius, we can derive a hydrodynamic equilibrium state for an axisymmetric ($\partial/\partial\varphi = 0$), non-accreting disk ($\boldsymbol{v} = u_\varphi \hat{\boldsymbol{\varphi}}$). For simplicity, we can assume that the disk is vertically isothermal ($\partial T/\partial z = 0$) and that the gas density at the midplane ρ_mid as well as the temperature both follow a power law in the radial direction such that

$$\rho_\mathrm{mid}(R) = \rho_{z=0} = \rho_0 \left(\frac{R}{R_0}\right)^p, \qquad T(R) = T_0 \left(\frac{R}{R_0}\right)^q, \tag{2.2}$$

where R_0 is a reference radius. With that in mind, we can write (see for example Nelson et al., 2013)

$$\begin{aligned} \rho(R, z) &= \rho_\mathrm{mid}(R) \exp\left[\frac{\mathrm{G}M_\star}{c_\mathrm{s}^2}\left(\frac{1}{r} - \frac{1}{R}\right)\right], \\ u_\varphi(R, z) &= R\Omega_\mathrm{K}\left[1 + (p+q)\left(\frac{H}{R}\right)^2 + q\left(1 - \frac{R}{r}\right)\right]^{1/2}. \end{aligned} \tag{2.3}$$

In the above, $\Omega_\mathrm{K} = \sqrt{\mathrm{G}M_\star/R^3}$ is the Keplerian frequency at radius R and $H = c_\mathrm{s}/\Omega_\mathrm{K}$ is the pressure scale height. Since disks are generally geometrically thin ($H \ll R$), we also define the disk aspect ratio $h = H/R$, which typically varies between 3–10%.

From the above relations, we can finally define the *column* or *surface density* Σ:

$$\Sigma(R) = \int_{-\infty}^{\infty} \rho(R, z) \mathrm{d}z \approx \sqrt{2\pi} \rho_\mathrm{mid} H \tag{2.4}$$

In the absence of radiative effects that can induce heating or cooling over long timescales, we can simply assume that the temperature profile $T \propto R^q$ is constant in time. This implies that the temperature profile of the disk is decided on a timescale that is short enough that its evolution can be disregarded,

or equivalently that the disk is always in thermal equilibrium. This approach, which neglects the energy equation in Eq. (2.1) by prescribing $T(R)$, is commonly referred to as *locally isothermal* and is a quite popular *ansatz* when modeling protoplanetary disks at ALMA-observed radii, where the cooling timescale $t_{\rm cool} \approx \epsilon/\dot{\epsilon}$ is expected to be a small fraction of the local dynamical timescale $t_{\rm dyn} = \Omega_{\rm K}^{-1}$ (e.g., Zhang et al., 2018).

In general, however, constructing realistic temperature structures requires modeling heat sources (e.g., viscous heating or stellar irradiation) as well as prescribing an appropriate cooling solution, be it through thermally relaxing the gas to a predetermined "equilibrium" state or by solving the radiation transport problem throughout the disk. These radiative effects will be discussed in the next section.

2.2 Radiative effects

The temperature profile of a protoplanetary disk is determined by the energy equation (see Eq. (2.1)). For a steady state and in the absence of heating through compression ($\nabla \cdot \boldsymbol{v} = 0$), thermal equilibrium is given by a balance among radiative heating and cooling terms. In this section, we describe some examples of mechanisms that motivate such terms.

The surface layers of the disk are typically exposed to heating from starlight by the young stellar object at the center of the system (Chiang and Goldreich, 1997). Radiation energy is then absorbed by dust grains, which re-emit in all directions similar to a black body. This isotropic thermal emission provides a cooling solution as excess heat is radiated away from the disk, while also illuminating the disk midplane as dust particles near the surface layers intercept photons from the central source and partially re-emit radiation energy towards the midplane. At the same time, the disk can heat up internally via turbulent dissipation. These mechanisms are sketched in Fig. 4.

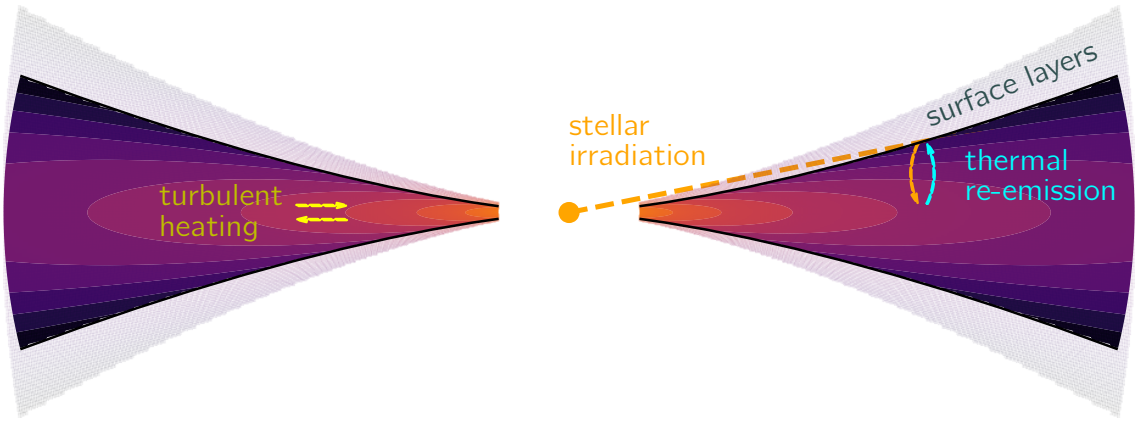

Figure 4. A sketch of the main radiative mechanisms that set the temperature profile of a typical protoplanetary disk in thermal equilibrium. In a *passive* disk, stellar irradiation deposits radiation energy at the surface layers of the disk (shown in pale colors), which is then redistributed via thermal re-emission of dust grains. For an *active* disk model, internal heating via turbulent dissipation can also take place.

In the next subsections, we will describe how radiation transport and stellar irradiation can be modeled, as well as a simplified approach that aims to capture radiative effects by modeling them with a single, "thermal relaxation" parameter.

2.2.1 Radiation transport

Depending on the dominant source of heating, protoplanetary disks are often characterized as *passive* (heated by stellar irradiation) or *active* (heated by internal viscous/turbulent dissipation). In both cases—as long as the gas and dust grains are thermally coupled such that the dust can carry thermal energy away

from the gas and emit it away—radiation energy is transported by the thermal emission of dust grains, the efficiency of which depends strongly on the emission optical depth of the gas:

$$\tau_{\rm emit} = \int_{z=0}^{\infty} \kappa_{\rm R}\rho {\rm d}z \sim \kappa_{\rm R}\rho H \tag{2.5}$$

where $\kappa_{\rm R}$ is the Rosseland mean opacity. The higher this $\tau_{\rm emit}$, the more difficult it is for radiation to escape a given region of the disk. For a sufficiently optically thick gas, we can treat the radiation problem as a diffusion of a radiative flux rather than the absorption and re-emission of individual photons. The transition between the optically thick ($\tau_{\rm emit} \gg 1$) and optically thin ($\tau_{\rm emit} \ll 1$) regimes is then handled by a "flux limiter" λ that is a function of the photon mean free path and the gradient of radiation energy. In this *flux-limited diffusion* approximation (FLD, Levermore and Pomraning, 1981), we can incorporate the radiative flux $\boldsymbol{F}_{\rm rad}$ into the energy equation and add an equation to solve for the time-dependent radiation energy $E_{\rm rad}$:

$$\begin{aligned} \frac{\partial(\rho\epsilon)}{\partial t} + \boldsymbol{v}\cdot\nabla(\rho\epsilon) &= -\gamma\rho\epsilon\nabla\cdot\boldsymbol{v} + (\overline{\boldsymbol{\sigma}}\cdot\nabla)\cdot\boldsymbol{v} - \kappa_{\rm P}\rho c(a_{\rm R}T^4 - E_{\rm rad}) \\ \frac{\partial E_{\rm rad}}{\partial t} + \nabla\cdot\boldsymbol{F}_{\rm rad} &= \kappa_{\rm P}\rho c(a_{\rm R}T^4 - E_{\rm rad}). \end{aligned} \tag{2.6}$$

Here, $\kappa_{\rm P}$ is the Planck mean opacity, $a_{\rm R}$ is the radiation constant, and c is the speed of light. The radiative flux is then given by

$$\boldsymbol{F}_{\rm rad} = -\lambda\frac{c}{\kappa_{\rm R}\rho}\nabla E_{\rm rad}. \tag{2.7}$$

To define the flux limiter we follow the approach of Kley (1989):

$$R = \frac{1}{\kappa_{\rm R}\rho}\frac{|\nabla E_{\rm rad}|}{E_{\rm rad}} \qquad \Rightarrow \qquad \lambda(R) = \begin{cases} \frac{2}{3+\sqrt{9+10R^2}}, & 0 \le R \le 2 \\ \frac{10}{10R+9+\sqrt{180R+81}}, & 2 \le R \le \infty. \end{cases} \tag{2.8}$$

As a result, $\lambda \to 1/3$ for optically thick regions (diffusion limit) and $\boldsymbol{F}_{\rm rad} \to cE_{\rm rad}$ for optically thin regions (free-streaming limit).

This approach does not account for radiation pressure (which would enter into the momentum equation in Eq. (2.1)), scattering effects, or the frequency-dependent nature of the opacities $\kappa_{\rm P}$ and $\kappa_{\rm R}$. Nevertheless, it provides a reasonable cooling mechanism that can release excess heat into the interstellar medium through the disk surfaces, while also treating radiative diffusion in the radial and vertical directions.

2.2.2 Stellar irradiation

As the disk is illuminated by starlight at its surface layers, it intercepts a portion of stellar photons based on its *absorption* optical depth $\tau_{\rm abs}$. Assuming a point-source star at $r = 0$ with properties that would be defined by its real radius $R_\star$, we can write

$$\tau_{\rm abs}(\boldsymbol{r}) = \int_{R_\star}^{r} \kappa_{\rm P}\,\rho\,{\rm d}r', \qquad r = |\boldsymbol{r}| \tag{2.9}$$

The amount of energy δE per unit time δt that strikes a fluid element is then

$$\frac{\delta E_{\rm in}(\boldsymbol{r})}{\delta t} = L_{\rm in}(\boldsymbol{r}) = A^r_\perp F_\star e^{-\tau_{\rm abs}(\boldsymbol{r})} = A^r_\perp \frac{L_\star}{4\pi r^2} e^{-\tau_{\rm abs}(\boldsymbol{r})}, \tag{2.10}$$

where $A^r_\perp$ is the surface normal to the direction of the ray at distance r. Considering a spherical geometry, a fluid element with a solid angle $\delta\Omega$ and radial extent $\delta r = |\boldsymbol{\delta r}|$ would then have a volume $V \approx r^2\delta r\delta\Omega$

and an area $A^r_\perp = r^2\delta\Omega$. This element would then intercept part of this incoming energy, and the remaining amount of energy would penetrate into the next element:

$$\frac{\delta E_{\rm out}(\boldsymbol{r})}{\delta t} = L_{\rm out}(\boldsymbol{r}) = L_{\rm in}(\boldsymbol{r}+\boldsymbol{\delta r}) = A^{r+\delta r}_\perp \frac{L_\star}{4\pi(r+\delta r)^2} e^{-\tau_{\rm abs}(\boldsymbol{r}+\boldsymbol{\delta r})}. \tag{2.11}$$

The absorbed energy density per unit time would then be:

$$S_{\rm irrad}(\boldsymbol{r}) = \frac{L_{\rm in}-L_{\rm out}}{V} \approx \frac{L_\star}{4\pi r^2\delta r} e^{-\tau_{\rm abs}(\boldsymbol{r})}\left(1-e^{-\kappa_{\rm P}\rho\delta r}\right). \tag{2.12}$$

This equation defines a source term that can be added to the energy equation in Eq. (2.1) or (2.6) to account for heating due to stellar irradiation.

It should be noted that most of the starlight is absorbed in the radial region between the star and the disk surface where $\tau_{\rm abs} \approx 1$. This surface typically corresponds to a height of $z \approx 4H$ (Chiang and Goldreich, 1997), and is followed by a sharp drop in gas temperature (see Fig. 5). Radiation energy then diffuses towards the midplane via thermal re-emission, even though the midplane is not directly illuminated by starlight. As a result, the vertical temperature profile is essentially isothermal between the midplane and the $\tau_{\rm abs} = 1$ surface, and then T rises dramatically near the top layers of the disk, which are directly exposed to the radiation field of the star (see right panel of Fig. 5).

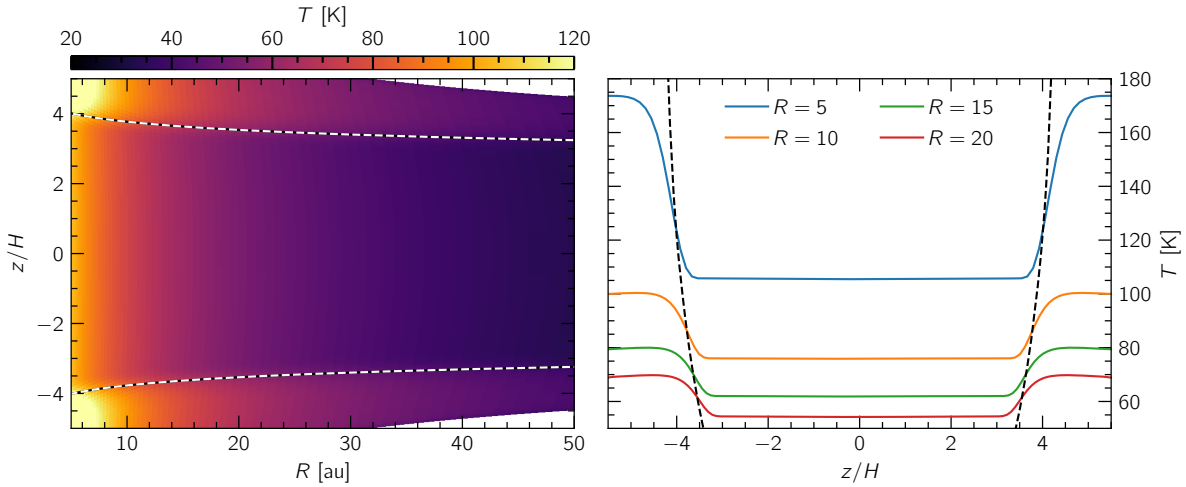

Figure 5. An example model of an irradiated disk, showcasing the radial and vertical temperature profile. Left panel: $T(R,z)$. The black-white dashed line marks the $\tau_{\rm abs} = 1$ surface. Right panel: $T(z)$ at several radii, quoted in au. The transition from a vertically isothermal interior to the hot disk surface can be seen at roughly 3–4 scale heights above the midplane, with the $\tau_{\rm abs} = 1$ surfaces traced in dashed black lines. The temperature profile at the disk midplane follows a power law of approximately $T(R) \propto R^{-1/2}$, in agreement with the passive disk profile of Chiang and Goldreich (1997).

2.2.3 A simplified approach: β-cooling

While the treatment of various radiative effects (namely stellar irradiation and radiation transport) is an attempt to model disks in more realistic conditions, it comes with several assumptions about the state of the disk that are often very difficult to constrain. Modeling gas and dust opacities requires estimating the abundances of hundreds of different species through observations and then constructing an opacity model based on those estimates (e.g., Lin and Papaloizou, 1985; Bell and Lin, 1994; Semenov et al., 2003; Birnstiel et al., 2018). In addition, estimating the midplane density and temperature of observed disks relies on assumptions about the mass fraction of observed species such as CO (Molyarova et al., 2017) and complex astrochemistry models (e.g., Rab et al., 2016).

We can instead parameterize radiative effects by defining a cooling timescale $t_{\rm cool}$ and relaxing the temperature profile of the disk to a predetermined state T_0 over that timescale. For example, assuming

that the gas cools through thermal emission, we can estimate the thermal relaxation timescale (Flock et al., 2017):

$$t_{\rm cool} = t_{\rm thin} + t_{\rm thick} \approx \frac{l_{\rm thin}^2}{3D_{\rm rad}} + \frac{H^2}{D_{\rm rad}}. \tag{2.13}$$

Here, $l_{\rm thin} = 1/\kappa_{\rm R}\rho$ is the photon mean free path and $D_{\rm rad}$ is the radiation diffusion coefficient

$$D_{\rm rad} = \frac{16\sigma_{\rm SB}T^3}{3\kappa_{\rm R}\rho^2 c_{\rm v}}, \qquad c_{\rm v} = \frac{\mathcal{R}}{\mu(\gamma-1)}, \tag{2.14}$$

with $\sigma_{\rm SB}$ and $c_{\rm v}$ being the Stefan–Boltzmann constant and the heat capacity at constant volume, respectively. This cooling timescale is then often expressed in units of the orbital time $t_{\rm dyn}$:

$$t_{\rm cool} = \beta\, t_{\rm dyn} \Rightarrow \beta(R,z) = t_{\rm cool}\Omega_{\rm K}. \tag{2.15}$$

See for example Gammie (2001). We can now define a relaxation source term for the energy equation

$$S_{\rm relax} = -\rho c_{\rm v}\frac{T-T_0}{t_{\rm cool}} \Rightarrow \frac{\partial T}{\partial t} = -\frac{T-T_0}{\beta}\Omega_{\rm K}. \tag{2.16}$$

This "β-cooling" approach greatly simplifies both theoretical and numerical modeling by combining all uncertainties about the thermodynamics of the gas into a single parameter β. In that sense, β is now a proxy for the efficiency of radiative effects: a very short cooling timescale ($\beta \to 0$) would correspond to a nearly locally isothermal disk, where the gas can readjust its temperature quickly enough to the point that there is no apparent change in time. An example of this scenario would be the cool, optically thin outskirts of protoplanetary disks (at $R \gtrsim 100\,$au). Conversely, a very long cooling timescale ($\beta \to \infty$) would refer to the hot, optically thick inner regions of such disks ($R \lesssim 1\,$au). An example of β as a function of distance and height can be seen in Fig. 6.

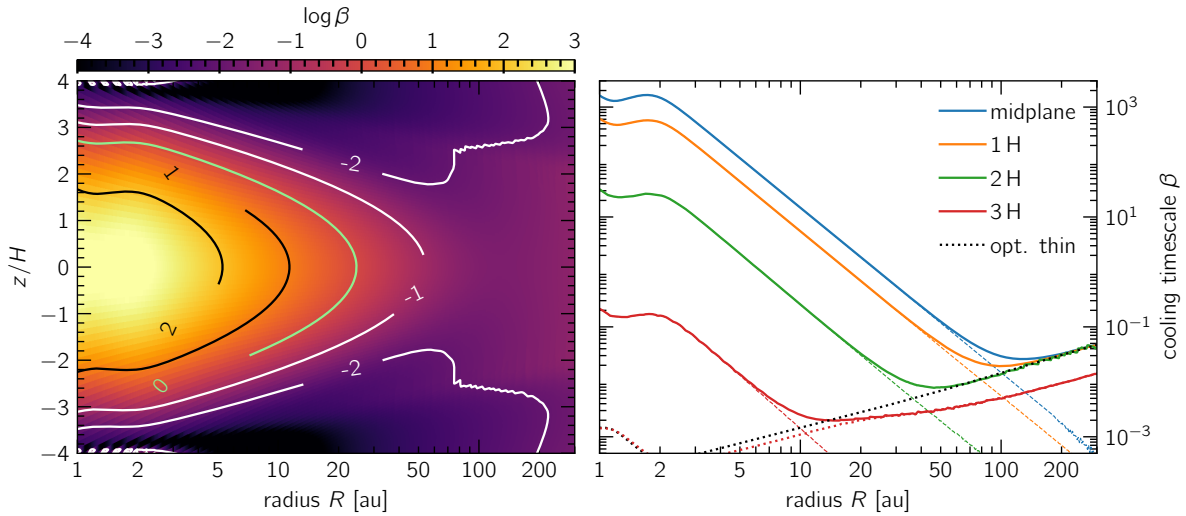

Figure 6. The dimensionless cooling timescale β as a function of height and radius. Left: a sample model of a passive protoplanetary disk with a typical density and temperature profile and an opacity model following Lin and Papaloizou (1985). Contours mark several values of β often used in the literature. Right: radial slices of β at different elevations above the midplane. Solid lines correspond to the cooling timescale given by Eq. (2.15), dashed lines only account for the optically thick contribution, and a black dotted line only accounts for the optically thin term (see Eq. (2.13)). Here, the cooling timescale is primarily dominated by the optically thick inner disk at $R \lesssim 50\,$au, while the outer disk and surface layers are optically thin and only contribute if we account for $t_{\rm thin}$ in Eq. (2.13). It should be noted that this picture is very sensitive to the choices of gas density, temperature, and opacity model.

2.3 A 2D approximation

In previous sections we saw how we can make predictions for the vertical structure of a protoplanetary disk with some reasonable assumptions. Specifically, the vertically isothermal assumption (see Fig. 5) combined with vertical hydrostatic equilibrium (see Eq. (2.3)) suggest that we can integrate Eq. (2.1) along the z axis to arrive at the *vertically integrated* Navier–Stokes equations:

$$\begin{aligned} \frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \boldsymbol{v}) &= 0, \\ \frac{\partial (\Sigma \boldsymbol{v})}{\partial t} + \nabla \cdot (\Sigma \boldsymbol{v} \otimes \boldsymbol{v}) &= -\nabla P_{\mathrm{2D}} - \Sigma \nabla \Phi + \nabla \cdot \overline{\boldsymbol{\sigma}}, \\ \frac{\partial (\Sigma \epsilon)}{\partial t} + \boldsymbol{v} \cdot \nabla (\Sigma \epsilon) &= -\gamma \Sigma \epsilon \nabla \cdot \boldsymbol{v} + (\overline{\boldsymbol{\sigma}} \cdot \nabla) \cdot \boldsymbol{v} + S, \end{aligned} \tag{2.17}$$

where Σ is defined through Eq. (2.4) and $P_{\mathrm{2D}} = (\gamma - 1)\Sigma\epsilon = \mathcal{R}\Sigma T/\mu$ is the vertically integrated pressure.

We can further calculate the contribution of individual radiative terms in this vertically-integrated approach. By assuming that $T_{\mathrm{rad}} = (E_{\mathrm{rad}}/a_{\mathrm{R}})^{1/4} = T_{\mathrm{gas}} = T$, which is a reasonable assumption below the $\tau_{\mathrm{abs}} = 1$ surface, Eq. (2.6) simplifies greatly to

$$\frac{\partial (\Sigma \epsilon)}{\partial t} + \boldsymbol{v} \cdot \nabla (\Sigma \epsilon) = -\gamma \Sigma \epsilon \nabla \cdot \boldsymbol{v} + Q_{\mathrm{visc}} + Q_{\mathrm{irrad}} + Q_{\mathrm{cool}} + Q_{\mathrm{FLD}}, \tag{2.18}$$

where Q_{visc} corresponds to the source term due to the vertically-integrated dissipation function

$$Q_{\mathrm{visc}} = (\overline{\boldsymbol{\sigma}} \cdot \nabla) \cdot \boldsymbol{v} \approx \frac{9}{4} \nu \Sigma \Omega_{\mathrm{K}}^2, \tag{2.19}$$

Q_{cool} and Q_{FLD} result from integrating the radiative flux $\boldsymbol{F}_{\mathrm{rad}}$ vertically,

$$\begin{aligned} Q_{\mathrm{cool}} &= -2\sigma_{\mathrm{SB}} \frac{T^4}{\tau_{\mathrm{eff}}}, \qquad \tau_{\mathrm{eff}} = \frac{3\tau}{8} + \frac{\sqrt{3}}{4} + \frac{1}{4\tau}, \qquad \tau = c_1 \frac{1}{\sqrt{2\pi}} \kappa_{\mathrm{R}} \Sigma \\ Q_{\mathrm{FLD}} &= -2H \nabla \cdot \boldsymbol{F}_{\mathrm{rad}}^{\mathrm{mid}}, \qquad \boldsymbol{F}_{\mathrm{rad}}^{\mathrm{mid}} = -\lambda \frac{c}{\kappa_{\mathrm{R}} \rho_{\mathrm{mid}}} \nabla a_{\mathrm{R}} T^4, \end{aligned} \tag{2.20}$$

and Q_{irrad} accounts for stellar irradiation following Menou and Goodman (2004):

$$Q_{\mathrm{irrad}} = 2 \frac{L_\star}{4\pi R^2} (1 - \varepsilon) \left(\frac{\mathrm{d}\log H}{\mathrm{d}\log R} - 1 \right) h \frac{1}{\tau_{\mathrm{eff}}}. \tag{2.21}$$

In the above, the effective optical depth at the midplane τ_{eff} is defined following Hubeny (1990), and for simplicity we assume that $\tau_{\mathrm{abs}} = \tau_{\mathrm{emit}} = \tau$. The factor $c_1 = 1/2$ is chosen as it provides a good agreement between 2D and 3D radiative simulations (Kley et al., 2009). Since only half of the intercepted irradiated flux is re-emitted towards the midplane, we define the disk albedo $\varepsilon = 1/2$.

Similarly to the 3D picture, a cooling timescale can be defined as

$$t_{\mathrm{cool}} \approx -\frac{\Sigma \epsilon_{\mathrm{mid}}}{Q_{\mathrm{cool}}} \Rightarrow \beta \approx \frac{\Sigma c_{\mathrm{v}} \tau_{\mathrm{eff}}}{2\sigma_{\mathrm{SB}} T^3} \Omega_{\mathrm{K}}, \tag{2.22}$$

and included using a β-cooling approach similar to Eq. (2.16). The optically thick and optically thin contributions to the cooling timescale shown in Eq. (2.13) are encoded in the effective optical depth τ_{eff}.

2.4 Accretion in protoplanetary disks

The Navier–Stokes equations incorporate the viscous stress tensor $\overline{\boldsymbol{\sigma}}$, which in general has the form

$$\sigma_{ab} = 2\eta\left[\frac{1}{2}\left(\frac{\partial u_a}{\partial x_b} + \frac{\partial u_b}{\partial x_a}\right) - \frac{1}{3}\left(\nabla\cdot\boldsymbol{v}\right)\delta_{ab}\right] + \zeta\left(\nabla\cdot\boldsymbol{v}\right)\delta_{ab}, \qquad a,b\in\{1,2,3\} \tag{2.23}$$

where η and ζ denote the *shear* and *bulk* viscosity coefficients respectively, and δ_{ab} is the Kronecker delta (see for example Tassoul, 1978). For a shear flow such as a protoplanetary disk, where η is expected to dominate, we can write that $\zeta = 0$ and $\eta = \nu\rho$, where ν is the kinematic viscosity coefficient.

The viscous stress tensor appears in both the momentum and energy equations, and its implications include angular momentum transport and dissipation of kinetic energy into heat. The former can result in an effective accretion (inflow) rate throughout the disk, while the latter is a source of heating.

Classical viscous disk theory predicts that, for a vertically integrated axisymmetric flow in steady state, the angular momentum $j = \Sigma R u_\varphi$ is transported radially outwards according to

$$\frac{\partial j}{\partial t} + \frac{1}{R}\frac{\partial}{\partial R}\left(Rju_R - R^2\sigma_{R\varphi}\right) = 0, \tag{2.24}$$

over a typical *viscous accretion timescale* which can be defined as $t_{\rm acc} = R^2/\nu$. This formalism suggests an explicitly defined viscosity ν, or by extension an explicit viscous stress tensor. In that regard, the result is a *laminar* accretion mechanism which provides a steady outward transport of angular momentum.

We can now estimate the lifetime of a disk by defining the mass accretion rate $\dot{M} := -2\pi R\Sigma u_R$. Observations suggest typical accretion rates of 10^{-9}–10^{-7} $\mathrm{M}_\odot$/yr with a median of $\sim 10^{-8}$ $\mathrm{M}_\odot$/yr for T Tauri stars (Hartmann et al., 1998), such that the typical lifetime of a disk with $M_{\rm disk} \sim 10^{-2}$–$10^{-1} M_\star$ corresponds to approximately 1–10 Myr (Haisch et al., 2001; Ribas et al., 2014).

In principle, however, while accretion is observed in protoplanetary disks, the accretion rates expected by kinetic gas theory alone fail to explain the observations. More specifically, if driven by the collision of particles at a molecular level, the estimated viscosity and therefore accretion timescales would be approximately

$$\nu_{\rm mol} \sim \lambda_{\rm f} u_{\rm th} \Rightarrow t_{\rm acc} \sim \frac{(1\text{–}10\,\mathrm{au})^2}{\nu_{\rm mol}} \sim 10^8\,\mathrm{Myr} \tag{2.25}$$

where $\lambda_{\rm f}$ and $u_{\rm th}$ correspond to the mean free path and thermal velocity of the gas, respectively. As a result, mechanisms that can generate large-scale turbulence and therefore *turbulent* accretion are necessary to explain the observed accretion rates and disk lifetimes.

Turbulent accretion can be driven by instabilities active at different regions of the disk, and the generated turbulent stress does not require the prescription of an explicit viscous stress tensor. Instead, it emerges from the dissipation of kinetic energy due to the redistribution of angular momentum of the gas in a region that is subject to an instability. Following Balbus and Papaloizou (1999), we can replace the "steady state" assumption in Eq. (2.24) with a "mean flow" approximation, where the gas is subject to an instability that perturbs its velocity components u_r, u_φ such that they fluctuate around a mean residual $\langle v_R\rangle = \langle u_R\rangle$ and $\langle v_\varphi\rangle = \langle u_\varphi - R\Omega_{\rm K}\rangle$. We can then rewrite Eq. (2.24) as

$$\frac{\partial\langle\Sigma R^2\Omega\rangle}{\partial t} + \frac{1}{R}\frac{\partial}{\partial R}\left(R^3\Omega\Sigma\langle v_R\rangle + R^2\Sigma\langle v_R v_\varphi\rangle\right) = 0, \tag{2.26}$$

where $\langle X\rangle$ denotes a density-weighted mean of a quantity X, smoothed radially over a length scale ΔR:

$$\langle X\rangle = \frac{1}{2\pi\Sigma\Delta R}\int_{-\infty}^{\infty}\int_{R-\Delta R/2}^{R+\Delta R/2}\int_{0}^{2\pi}\rho X\,\mathrm{d}\varphi\mathrm{d}R\mathrm{d}z. \tag{2.27}$$

Here, Σ is defined through Eq. (2.4) and then subsequently smoothed radially. A typical smoothing length would be $\Delta R \sim H(R)$. The quantity $\langle v_R v_\varphi\rangle$ then corresponds to the $R\varphi$ component of the Reynolds

stress tensor $\overline{\mathbf{T}}$, or $T_{R\varphi}$ for short. This term, which describes the density-weighted correlation between v_R and v_φ for this mean flow, is ultimately responsible for driving accretion in the absence of magnetic fields.

Assuming a turbulent flow of gas with a nonzero $T_{R\varphi}$, we can now calculate an *effective* viscosity $\nu_{\rm eff}$ by combining Eqs. (2.24) and (2.26):

$$-R^2\sigma_{R\varphi} = R^2\Sigma\langle v_R v_\varphi\rangle \Rightarrow \nu_{\rm eff} = -\langle v_R v_\varphi\rangle\left(R\frac{{\rm d}\Omega}{{\rm d}R}\right)^{-1}. \tag{2.28}$$

This procedure, while requiring extensive temporal averaging and spatial smoothing, allows us to extract the turbulent stress levels from a hydrodynamical instability and model its impact on accretion as a pseudo-laminar viscous stress tensor according to Eq. (2.23). We can further parameterize it by employing the Shakura and Sunyaev (1973) "α-viscosity" prescription, where

$$\nu = \alpha c_{\rm s}^{\rm ad} H. \tag{2.29}$$

One interpretation of the above relation would be the modeling of large-scale turbulent motions (like eddies) with a size H, and a turnover speed that is a small fraction of the local sound speed. The dimensionless viscosity parameter α then encapsulates the properties of the turbulence-driving instability.

Several different candidates of (magneto-)hydrodynamical instabilities that can drive accretion have been studied over the years. A major candidate has been the magneto-rotational instability (MRI, Balbus and Hawley, 1991) which can generate effective α of up to 10^{-3}–10^{-2} (Hawley et al., 1995) as long as the gas is sufficiently ionized. For realistic, weakly-ionized disks however, the MRI can be either inactive due to an insufficient ionization fraction (Gammie, 1996) or suppressed by non-ideal MHD effects (Cui and Bai, 2020). An alternative that has been extensively studied for the last decade is the vertical shear instability (VSI, Nelson et al., 2013), which is driven purely from the vertical gradient of $u_\varphi(R, z)$ that naturally exists for steady-state protoplanetary disks (see Eq. (2.3)) and produces strong vertical motion that can generate α values of 10^{-5}–10^{-3} (Nelson et al., 2013; Stoll and Kley, 2014; Flock et al., 2017). While neither MRI nor VSI signatures have been directly observed, VSI-generated stress levels are compatible with observational constraints on disk turbulence from ALMA observations (Dullemond et al., 2018), and it might be possible to detect vertical motion induced by the VSI within the next few years using ALMA CO kinematics (Barraza-Alfaro et al., 2021). For an overview of various instabilities as accretion-driving mechanisms, see Lyra and Umurhan (2019).

The VSI is central to this work, and is explained in more detail in the next section.

2.5 The vertical shear instability

The vertical shear instability (VSI) was first investigated by Goldreich and Schubert (1967) and Fricke (1968) in the context of stellar interiors, to test the stability of differential rotation in the radiative zones of a star. It was fittingly dubbed the Goldreich–Schubert–Fricke (or GSF) instability, and the condition for stability was given by the lack of a vertical shear ($\partial\Omega/\partial z = 0$) and an outwardly-growing angular momentum profile ($\partial j/\partial R \geq 0$).

The GSF instability has since then been reanalyzed in the context of protoplanetary disks (e.g., Urpin and Brandenburg, 1998; Nelson et al., 2013; Stoll and Kley, 2014; Flock et al., 2017), where a vertical shear arises naturally in hydrodynamic equilibrium (see Eq. (2.3)). More importantly, since the VSI "feeds" on rotational energy (the vertical shear), its turbulent dissipation generates angular momentum transport and, therefore, accretion. This makes it a popular candidate for modeling accretion in very weakly-ionized disks.

2.5.1 The mechanism behind the VSI

We can illustrate an example of a VSI-unstable configuration in its linear growth phase by looking at a typical disk profile in hydrodynamic equilibrium (see Fig. 7), where $\rho(R, z)$, $T(R)$ and $\Omega(R, z)$ are given

through Eqs. (2.2) and (2.3). We displace a fluid element upwards ($(R, z) \rightarrow (R', z')$) along a line of constant specific angular momentum $j = j' = \Omega R^2$. Comparing the kinetic energy of the fluid element $E_{\rm kin}^{\rm el} = \Omega^2 R^2$ to that of its surroundings, $E_{\rm kin}^{\rm env} = \Omega'^2 R'^2$, we find that $E_{\rm kin}^{\rm el} > E_{\rm kin}^{\rm env}$ and therefore the fluid element should experience an acceleration in the direction of the displacement, creating a positive feedback cycle that allows the instability to grow. The configuration is unstable if the vertical shear ($\propto \partial\Omega/\partial z$), which here encourages upward movement, is strong enough to overcome the buoyancy of the gas, which acts as a restoring force. This is relatively easily achievable in the context of protoplanetary disks, where cooling timescales are often short enough (see Fig. 6) such that the perturbed fluid element can quickly attain the entropy of the surrounding fluid, effectively ignoring the effects of buoyancy.

The result is the development of unstable modes near the disk surface layers at $z \sim 4H$ (where the vertical shear rate is strongest), which then grow and saturate as they travel towards the midplane from both sides of the disk before joining into vertical corrugation modes that span the full extent of the disk (see Fig. 8). A more detailed derivation of the stability criterion for the VSI can be found in Nelson et al. (2013).

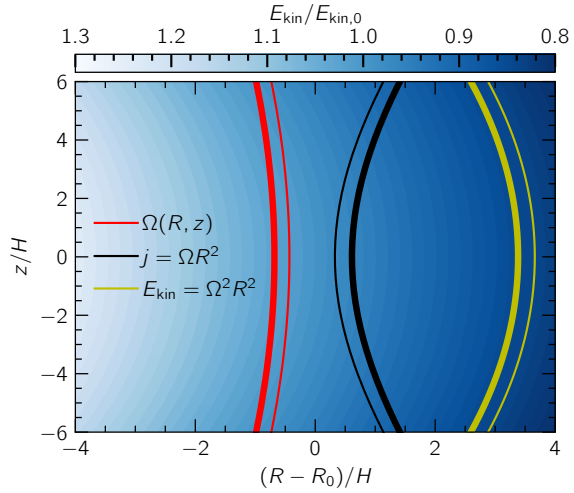

Figure 7. Illustration of specific angular momentum j and kinetic energy $E_{\rm kin}$ for a disk in hydrodynamic equilibrium. Each quantity is constant along a given contour, and increases in the direction from the thinner to the thicker curve.

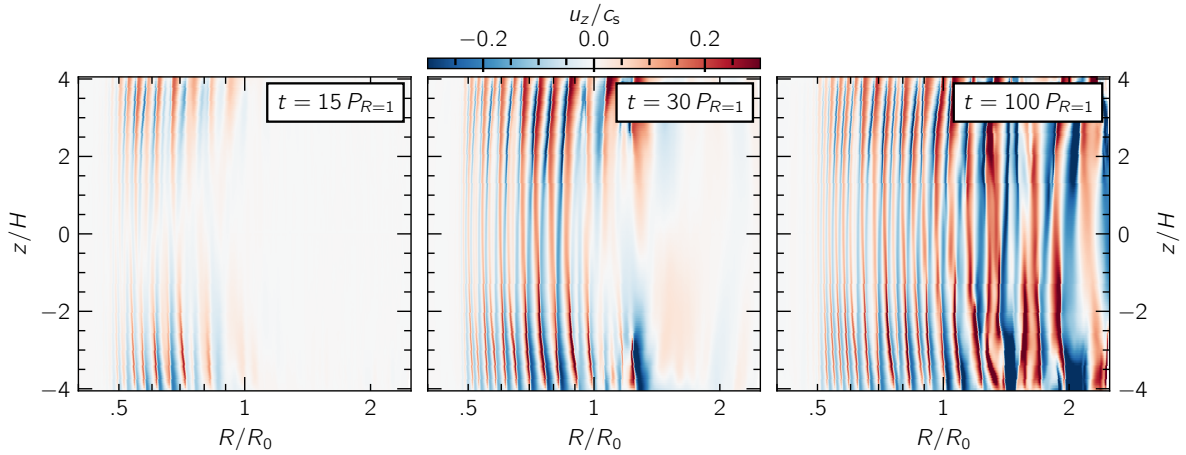

Figure 8. Development and saturation of VSI modes in a VSI-unstable disk. Left: VSI modes develop near the disk surface ($z \sim 4H$) and travel towards the midplane from both directions. Since the vertical shear rate scales with $\Omega_{\rm K}$, this happens first for small radii. Middle: the development of VSI modes continues traveling outwards, while modes within $R \lesssim R_0$ have joined together vertically and saturated. Right: the VSI is saturated in the entire radial extent of the disk. The region $R < 0.5\,R_0$ is damped according to Eq. (3.1) and therefore does not exhibit VSI activity.

2.5.2 Dependence on disk parameters

For a typical, axisymmetric accretion disk with a negative radial density and temperature gradient (see e.g., Eq. (2.2)), the vertical shear rate can be derived through the equilibrium state given in Eq. (2.3):

$$R\frac{\partial\Omega}{\partial z} = \frac{1}{2}q\frac{\Omega_{\rm K}^2}{\Omega}\frac{R^2 z}{r^3} \approx -\frac{1}{2}|q|h\Omega_{\rm K}\left(\frac{z}{H}\right). \tag{2.30}$$

The VSI therefore thrives in geometrically thick disks ($h = H/R$) with steep radial temperature gradients ($q = \frac{\mathrm{d}\log T}{\mathrm{d}\log R}$), and first develops near the disk surfaces ($|z| \approx 4H$). The observed turbulence similarly saturates at different levels depending on those parameters. For protoplanetary disks at radii observable with ALMA, we have $h \approx 0.1$, $q \approx -0.5$, and an effective turbulent $\alpha \sim 10^{-4}$ (Flock et al., 2017).

On the other hand, as mentioned in the previous section, if the perturbed fluid element cannot reach thermal equilibrium with its environment quickly enough, buoyancy can overcome the vertical shear and restore the fluid element to its starting position, stabilizing the flow. In a thin disk ($h \ll 1$), we can write the vertical buoyancy frequency N_z as

$$N_z \approx \sqrt{\frac{\gamma-1}{\gamma}}\Omega_{\rm K}\left(\frac{z}{H}\right). \tag{2.31}$$

Based on the above, Lin and Youdin (2015) derived a critical cooling timescale $\beta_{\rm crit}$ that allows the instability to develop if $\beta < \beta_{\rm crit}$, with

$$\beta_{\rm crit} \lesssim \frac{R|\partial\Omega/\partial z|}{N_z^2}\Omega_{\rm K}^{-1} \approx \frac{h|q|}{\gamma-1}, \tag{2.32}$$

which corresponds to $\beta_{\rm crit} \approx 0.1$ for typical disk parameters and agrees with estimates of the cooling timescale at radii $R \gtrsim 30$ au (e.g., Flock et al., 2017).

It should be noted that since N_z^2 grows faster with height than the vertical shear rate, the VSI can be quenched at a certain height above the midplane, leading to possible damping of the surface modes. In addition, Pfeil and Klahr (2021) showed that accounting for dust–gas decoupling at the surface layers of the disk leads to significantly longer effective cooling timescales, which can also "sandwich" the VSI modes within a few scale heights above the midplane.

Finally, viscous dissipation can counteract/prevent the growth of VSI-unstable modes. Nelson et al. (2013) found that a background kinematic viscosity $\nu_{\rm bg}$ of the order of the generated VSI-turbulent viscosity $\nu_{\rm VSI}$ can significantly reduce VSI-induced motion, halting it completely when $\nu_{\rm bg} \gtrsim \nu_{\rm VSI}$. This hints at a possible damping of the VSI in the presence of other instabilities, such as the streaming instability (Schäfer et al., 2020).

2.6 Planet–disk interaction

Protoplanetary disks are the birthplace of young planets, which interact with the surrounding gas and dust as they orbit around the system's central object. Through this interaction planets can grow by amassing dusty or gaseous material from the disk in a process called *accretion*. The disk can also exert a torque on the planet, causing it to *migrate* inwards or outwards within the disk.

Planet–disk interaction can be modeled by treating the planet as a point source with mass M_p at location $\boldsymbol{r}_\mathrm{p}$ and considering the gravitational potential it subjects a disk element at position $\boldsymbol{r}$ to

$$\Phi_\mathrm{p} = -\frac{\mathrm{G}M_\mathrm{p}}{d}, \qquad \boldsymbol{d} = \boldsymbol{r} - \boldsymbol{r}_\mathrm{p} \tag{2.33}$$

Of course, numerically, this expression can result in numerical pathogenities around the planet, where $d \to 0$. For that reason, the planet's potential is typically replaced by a cubic interpolation very close to the planet (Klahr and Kley, 2006)

$$\Phi_\mathrm{p}^* = -\frac{\mathrm{G}M_\mathrm{p}}{r_\mathrm{sm}}\left[\left(\frac{d}{r_\mathrm{sm}}\right)^3 - 2\left(\frac{d}{r_\mathrm{sm}}\right)^2 + 2\right] \text{ if } d < r_\mathrm{sm}, \quad r_\mathrm{sm} = 0.5\,R_\mathrm{Hill}, \quad R_\mathrm{Hill} = R_\mathrm{p}\sqrt[3]{\frac{M_\mathrm{p}}{3M_\star}} \tag{2.34}$$

In vertically integrated models, the vertical stratification of the disk must be taken into account as well when computing the gravitational acceleration on a *column* of gas. In that case, the planet's gravity can be described well with a Plummer potential where

$$\Phi_\mathrm{p} = -\frac{\mathrm{G}M_\mathrm{p}}{\sqrt{d^2 + \epsilon^2}}, \qquad \epsilon \sim H(\boldsymbol{r}) \tag{2.35}$$

Using $\epsilon = 0.6H$ (Müller et al., 2012) is typical in the literature. It is important to note that H is evaluated *locally* instead of only at the planet's location.

Planet–disk interaction creates features in the gas. Spiral arms are perhaps the most recognizable planet-generated feature, which is in turn responsible for various observables attributed to or explainable through planets (e.g., rings, gaps). The characteristic trajectory of a spiral arm launched by a planet can be approximated through linear theory (Ogilvie and Lubow, 2002; Rafikov, 2002), by considering hydrodynamical waves excited by the planet at Lindblad-resonant locations and their shearing as they propagate radially through the disk. While this method works best for a planet with a vanishingly small mass and far from said planet, it can provide useful information regarding the planet's mass or the disk's background temperature profile when used to match spiral patterns seen in observations of protoplanetary disks (e.g., Zhu et al., 2015; Huang et al., 2018b). An example of a planet-generated wake and the related linear theory fit can be seen in the left panel of Fig. 9.

For large enough planet masses, nonlinear effects come into play (e.g., Lin and Papaloizou, 1993; Rafikov, 2002). A single, massive enough planet can launch secondary or even tertiary spiral arms that steepen into shocks, depositing angular momentum and heat into the disk. This angular momentum exchange can result in the "siphoning" of material away from the point of launch of a spiral arm, which effectively translates to a local depression in gas density. This is typically observed as a gap at the planet's radius, but can also result in the formation of multiple gaps by a single planet in the right conditions (e.g., Zhang et al., 2018; Miranda and Rafikov, 2020b; Ziampras et al., 2020b). At the same time, the heat deposited near the shock front can, depending on disk thermodynamics, strongly affect the global temperature profile and consequently the hydrodynamics of the disk (e.g., Rafikov, 2016; Ziampras et al., 2020a). An example of gap opening and the temperature increase caused by spirals launched by a massive planet can be seen in the right panel of Fig. 9.

The gap opening process also creates the conditions for the development of other interesting features. The local clearing of material from within gap regions effectively results in the buildup of pressure bumps at the edges of said gaps, forming "traps" that can accumulate dust particles. The thermal emission of such collections of dust is then visible as bright rings in mm wavelengths with instruments such as ALMA (e.g., Huang et al., 2018a). Moreover, the steep density gradient at a gap edge satisfies the criteria

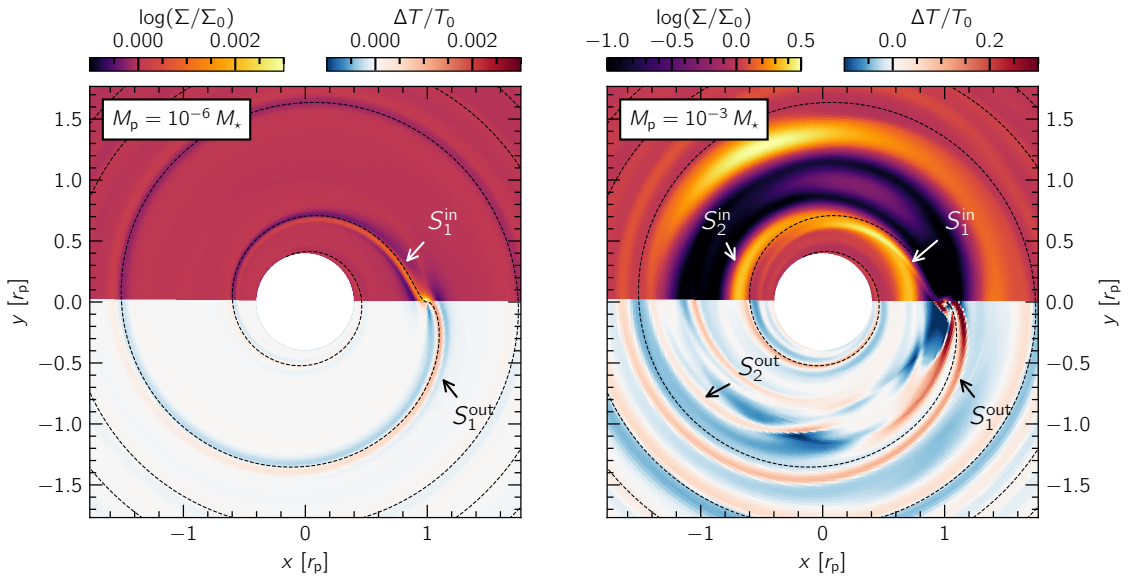

Figure 9. Planet–disk interaction in the low-mass, linear and high-mass, nonlinear regimes. Left: a low-mass planet excites an inner and outer spiral which permeate the disk in both radial directions without affecting the overall disk structure. Linear theory provides a very good prediction of the position of such spirals (dashed black curves). Right: a massive planet drives several nonlinear effects, namely the excitation of secondary spirals (marked by the index of the annotations in the figure), a deep gap around its orbit, and a massive vortex that orbits at the outer gap edge. Linear theory fails to predict the position of spirals, which also shock and can heat up the disk substantially. The models shown here simulate an inviscid, β-cooled disk with $\beta = 1$ and $h = 0.05$ for 100 planetary orbits.

for the Rossby-wave instability (RWI, Lovelace et al., 1999), resulting in the formation of vortices—non-axisymmetric anticyclonic features that orbit at Keplerian velocities and can decay over hundreds to tens of thousands of orbits (Hammer et al., 2019; Fung and Ono, 2021; Rometsch et al., 2021). Due to their anticyclonic (pressure maximum) nature, vortices can also act as dust traps and are therefore invoked to explain non-axisymmetric observables that can result from the planet formation scenario (Rodenkirch et al., 2021). Fig. 10 shows an ALMA observation of the system HD 163296, which contains both rings and a non-axisymmetric, vortex-like feature, as well as the result of a numerical simulation of this system that models these features as the result of planet–disk interaction.

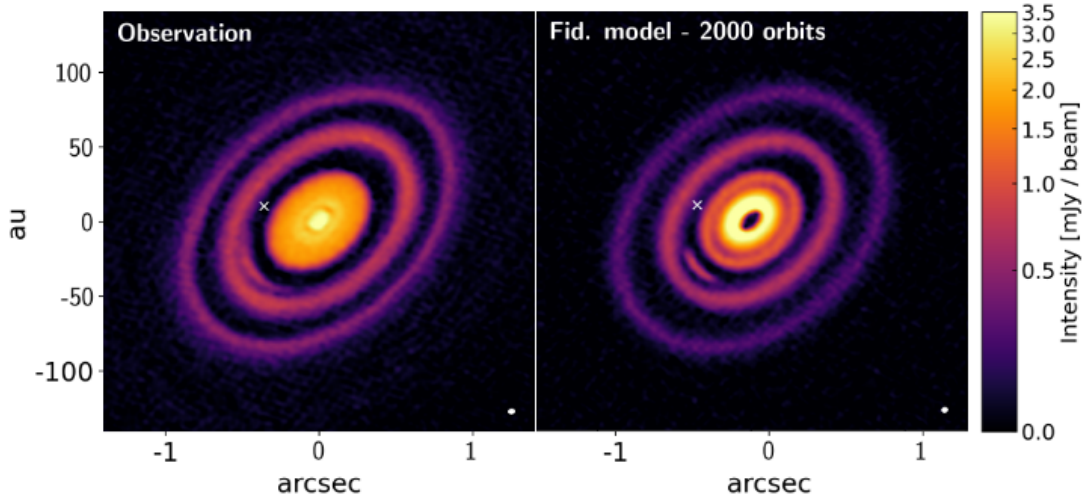

Figure 10. Left: an ALMA observation of HD 163296 (Andrews et al., 2018; Isella et al., 2018). Right: a numerical model of this system by Rodenkirch et al. (2021), aiming to reproduce the non-axisymmetric feature as the result of dust trapping inside a planet's gap region. Image credit: Rodenkirch et al. (2021) (see Fig. 13 therein).

3 Numerics

In this section we introduce the code and numerical methods employed in our study. This project involved both two- and three-dimensional simulations of protoplanetary disks, parts of which have already been published. To avoid repetition, we describe the common aspects of both families of simulations here, and direct the reader to related references for additional information. To disambiguate between different kinds of models, we will refer to simulations of the vertically-integrated equations using a cylindrical setup on the $R\varphi$ plane (see Eq. (2.17) and Sect. 2.3) as "$R\varphi$", models of axisymmetric, vertical slices on the $r\theta$ plane in spherical coordinates using Eq. (2.1) as "$r\theta$-2D", and models of the fully three-dimensional equations as "$r\theta\varphi$-3D". We will collectively refer to the $r\theta$-2D and $r\theta\varphi$-3D models as "$r\theta$" models.

Our $R\varphi$ models primarily focus on planet–disk interaction through the formation of rings and gaps for disks with different equations of state (mainly discussed in Ziampras et al., 2020b, and summarized in Sect. 5) as well as the formation and lifetime of planet-generated vortices (mainly discussed in Rometsch et al., 2021, and summarized in Sect. 6). On the other hand, our $r\theta$ models focus on the hydrodynamics of VSI-active disks. We present our results on these models in Sects. 7 and 8, and discuss them in Sect. 9.

3.1 PLUTO setup

We use the numerical hydrodynamics package PLUTO (Mignone et al., 2007), using version 4.2 for $R\varphi$ models and 4.3 for $r\theta$ ones. PLUTO utilizes a finite-volume, shock-capturing approach with access to a variety of Riemann solvers in order to integrate the equations of hydrodynamics (Eqs. (2.1) and (2.17)) in conservative form. The methods used are the same between versions.

Unless otherwise stated, our configuration is second-order accurate in both space and time, using an RK2 timestepping scheme and a 3-point-wide, piecewise-TVD[2] linear reconstruction stencil combined with the van Leer flux limiter (Van Leer, 1974). We chose the HLLC Riemann solver (Toro et al., 1994) for most of our models, since comparisons with the more accurate (but computationally more expensive) Roe solver yielded very similar results.

Whenever possible, the FARGO method is applied (Masset, 2000, implemented into PLUTO by Mignone et al. (2012)). This numerical scheme greatly relaxes the timestep limitation that the otherwise supersonic disk background would impose, by solving for the residual motion after this mean background speed has been subtracted. We typically use the IDEAL equation of state, which evolves all quantities in Eqs. (2.1) and (2.17), and the ISOTHERMAL EOS for locally isothermal models.

Viscous diffusion and dissipation are handled with PLUTO's VISCOSITY module using the Super-Time-Stepping scheme (Alexiades et al., 1996), which provides a reasonable balance between speed and stability for the otherwise strict timestep limitation due to the parabolic nature of viscous terms. Cooling, when applicable, is typically prescribed using PLUTO's TABULATED cooling module, unless stated otherwise.

The gravitational potential of the star and planet are included using the Nbody module of Thun and Kley (2018), which we ported to PLUTO 4.3. The entire star–planet–disk system orbits around the star's reference frame. Neither the planet nor the star feel feedback from the disk, unless stated otherwise. In $R\varphi$ simulations with embedded planets we use a Plummer potential with a smoothing length of $\epsilon = 0.6H$ as shown in Eq. (2.35), while in $r\theta\varphi$-3D simulations we opt for the cubic spline curve by Klahr and Kley (2006) (see Eq. (2.34)).

For more information on the modeling behind our $R\varphi$ simulations see Ziampras et al. (2020a), Ziampras et al. (2020b), or Rometsch et al. (2021). Additional physics such as radiation transport will be mentioned when used, and are described in detail in the Appendix.

[2] total variation diminishing (Harten, 1983)

3.2 Grid setup

For $R\varphi$ models, the computational grid typically extends between $0.2 \leq R/R_0 \leq 5$. We adopted a resolution of at least 8 cells per scale height (cps), a number which is commonly agreed upon as a sufficiently high numerical resolution in 2D planet–disk interaction for our purposes. Some models ran at a higher resolution to test for convergence. The grid is logarithmically spaced radially in order to maintain the cell aspect ratio, and the azimuthal cell count is chosen such that the grid consists of square cells.

For $r\theta$ runs, we use a grid with $0.4 \leq r/R_0 \leq 2.5$. The resolution is at least 16 cps in the radial direction, as we found that this is necessary to adequately resolve the VSI after a resolution study which we will discuss in Sect. 7.2.1. The polar direction covers 4 scale heights in both directions about the midplane (for a total of $8H$ from end to end) with a resolution of 16 cps. We found this extent to yield the same results as a model with a domain covering the full polar extent $\theta \in [0, \pi]$. Our $r\theta\varphi$-3D runs have an azimuthal resolution of 5 cps, which should be enough for our purposes, but we also tested against a run at 8 cps and found very similar results.

A sketch of our grid setup is shown in Fig. 11. In models where a planet is present, R_0 in that figure represents R_p. Otherwise, R_0 simply corresponds to a reference distance from the star.

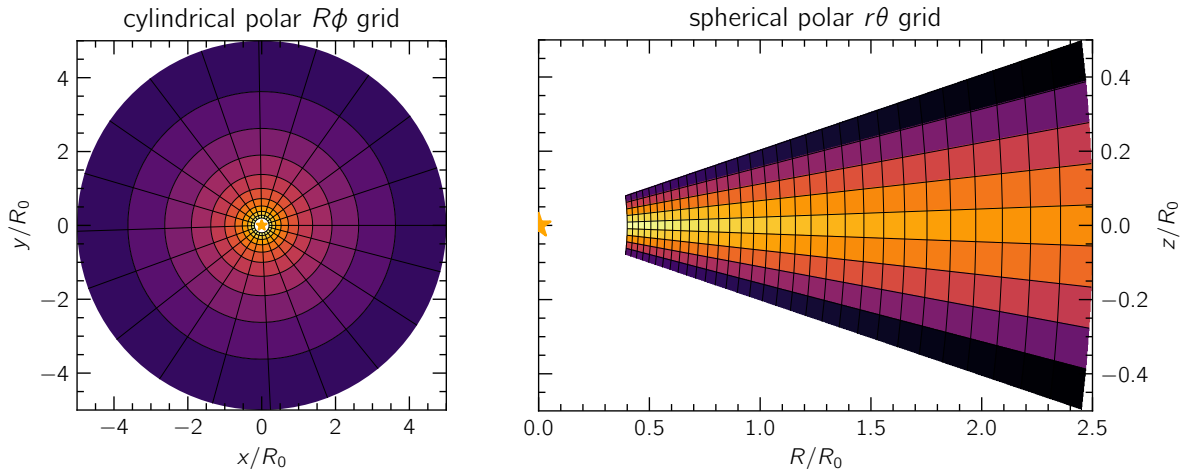

Figure 11. Illustration of our grid setups for $R\varphi$ models (left) and $r\theta$ models (right). Our $r\theta\varphi$-3D simulations extend vertically as shown in the right panel and azimuthally similar to the left, for a full-3D spherical grid. To preserve cell aspect ratio, grid cells are spaced logarithmically radially and uniformly in other directions. The color represents a typical, radially dropping $\Sigma(R)$ profile (left) and its corresponding $\rho(r, \theta)$ profile (right).

3.3 Initial and boundary conditions

All models are initialized in equilibrium, which translates to radial power law profiles for $\Sigma(R)$ and $T(R)$ for $R\varphi$ models (see Eqs. (2.2) and (2.4)), while the vertical density stratification $\rho(r, \theta)$ in $r\theta$ models follows Eq. (2.3). The gas velocity components are also initialized through these equilibrium profiles, with $u_R(R)$ adjusted to account for viscous accretion and $u_\varphi(R)$ corrected for radial (and vertical) pressure support through Eq. (2.3). In $r\theta$ models, a small amount of numerical noise with a maximum amplitude of 1% of the local sound speed is added to all velocity components. We refer the reader to Ziampras et al. (2020b) or Rometsch et al. (2021) for more information on the initialization of $R\varphi$ models, and to Nelson et al. (2013) for $r\theta$ models.

Boundaries are always periodic in the azimuthal direction. The boundary wall in our simulations is closed in the radial direction for $R\varphi$ models, and left open for $r\theta$ models such that gas is allowed to exit but not reenter the grid. The same "strict outflow" boundary is applied in the polar direction of $r\theta$ models, with the density and pressure extrapolated into the boundary zone to account for hydrostatic equilibrium. In all models, we utilize wave damping in a radial zone that extends from the inner/outer

boundary walls up to a small portion of the active domain, following the prescription of de Val-Borro et al. (2006):

$$\frac{\partial x}{\partial t} = -\frac{x - x_{t=0}}{t_{\rm damp}} R_{\rm damp}, \qquad x \in \{\boldsymbol{v}, \Sigma, \rho, P\} \tag{3.1}$$

with a damping timescale $t_{\rm damp} = 0.3$ and 10 for $R\varphi$ and $r\theta$ models respectively, in units of the orbital period in each radial boundary wall. The choice of this relatively long damping timescale for our $r\theta$ models relates to an unusual parasitic instability we observed near the inner boundary in those simulations for shorter $t_{\rm damp}$ (discussed in Sect. 7.2.2). Nevertheless, the open inner boundary helps minimize reflections back into the grid in those simulations.

Since our $r\theta\varphi$-3D models are meant to exhibit VSI activity, we initialized most of them by evolving an axisymmetric $r\theta$-2D disk until the VSI was fully saturated, and then expanded all fields in 3D with the addition of a small amount of noise (1% c_s) similar to Flock et al. (2020). We nevertheless checked whether this assumption is reasonable by allowing the VSI to develop from a quiet disk in 3D, and found the same results.

3.4 Computing turbulent stress in $r\theta$ models

As we discussed in Sect. 2.5, the VSI "feeds" on the angular momentum budget of the disk via the vertical shear and can therefore drive accretion. In fact, a VSI-active disk has been shown to exhibit a mean radial drift inwards near the disk midplane and outwards near its surfaces, resulting in accretion when accounting for the net, mass-weighted difference between the two (Stoll et al., 2017). We can now apply the averaging technique of Eq. (2.27) to our $r\theta$ models and consider the mean radial velocity field of a VSI-active, turbulent disk as the result of a mean Reynolds stress tensor following the strategy discussed in Sect. 2.4. From this analysis we can finally derive the effective turbulent viscosity parameter α and compare to values found in the literature (e.g., Stoll and Kley, 2014; Stoll et al., 2017; Flock et al., 2017).

We implement this in four stages: since the quasi-steady state of a saturated VSI-active disk still contains stochastic motion, we first extract the time-averaged state of the disk starting at a suitable timestamp where the VSI has fully developed and saturated (typically after 100 orbits at $R = R_0$) and spanning $N_{\rm files} = 200$–400 snapshots taken at intervals of one orbit at R_0. The result corresponds to a 3D, averaged disk state which we consider our "mean flow" and will denote with a bar (e.g., $\bar{\rho}$, $\bar{\boldsymbol{v}}$).

In the second stage, we compute the $R\varphi$ and $z\varphi$ components of the Reynolds stress tensor $\overline{\mathbf{T}}$ at every point in 3D space for every snapshot with index i

$$T^i_{x\varphi}(\boldsymbol{r}) = \rho^i \cdot (u^i_x - \bar{u}_x) \cdot (u^i_\varphi - \bar{u}_\varphi) = \rho^i v^i_x v^i_\varphi, \qquad x \in \{R, z\} \tag{3.2}$$

we then average over $N_{\rm files}$ snapshots to compute the mean stress tensor

$$\bar{T}_{x\varphi}(\boldsymbol{r}) = \frac{1}{N_{\rm files}} \sum_i^{N_{\rm files}} T^i_{x\varphi}(\boldsymbol{r}) \tag{3.3}$$

In the third stage, we compute the mean surface density Σ using $\bar{\rho}$ and Eq. (2.4), take its azimuthal average, and apply a radial smoothing over a window of a local scale height $H(R)$ according to Eq. (2.27). We apply the same radial smoothing (but not azimuthal averaging) to $\bar{T}_{x\varphi}$ as well:

$$\begin{aligned} W_{x\varphi}(R, \varphi, z) &= \frac{1}{\Delta R} \int_{R-\Delta R/2}^{R+\Delta R/2} \bar{T}_{x\varphi}(R', \varphi, z)\, {\rm d}R', \qquad \Delta R = H(R) \\ \Sigma(R) &= \frac{1}{2\pi\Delta R} \int_{R-\Delta R/2}^{R+\Delta R/2} \int_0^{2\pi} \int_{z_{\rm min}}^{z_{\rm max}} \bar{\rho}(R', \varphi, z)\, {\rm d}z\, {\rm d}\varphi\, {\rm d}R' \end{aligned} \tag{3.4}$$

The result is a 3D map in $\{R, \varphi, z\}$ for each of the two components $T_{R\varphi}$ and $T_{z\varphi}$, which are relevant for radial accretion and vertical mixing respectively.

In the fourth and final stage, we extract the vertically integrated profile of the effective turbulent accretion parameter $\alpha_{\rm acc}(R)$ by summing over all contributions of $W_{R\varphi}$ as a function of height and then averaging in the azimuthal direction if necessary:

$$\bar{W}_{R\varphi}(R) = \frac{1}{2\pi}\int_0^{2\pi}\int_{-z_{\rm max}}^{z_{\rm max}} W_{R\varphi}(R,\varphi,z)\,{\rm d}z\,{\rm d}\varphi \Rightarrow \alpha_{\rm acc}(R) = -\bar{W}_{R\varphi}\cdot\left(R\frac{{\rm d}\Omega}{{\rm d}R}\Sigma c_{\rm s}H\right)^{-1}, \tag{3.5}$$

where Ω, $c_{\rm s}$ and H are similarly calculated using the mean fields computed during the first stage. We note that, since the mean azimuthal velocity and disk temperature do not change significantly due to the VSI ($\Delta u_\varphi/u_{\varphi 0} \sim 10^{-2}$, $\Delta T/T_0 \sim 10^{-6}$), one can also instead use the initial, analytical profiles of Ω, $c_{\rm s}$ and H in these calculations. However, since the disk is constantly accreting (i.e., losing mass in our models), the mean density $\bar{\rho}$ is more appropriate than the initial profile ρ_0 that Eq. (2.3) suggests when calculating Σ.

The method described in Eq. (3.5) is not typically followed in the literature, where variants of $\bar{W}_{R\varphi}/\bar{P}$ are used (e.g., Nelson et al., 2013; Flock et al., 2017; Manger et al., 2020). In fact, approximating $\Omega \approx \Omega_{\rm K} \propto R^{-\frac{3}{2}}$ in Eq. (3.5) yields

$$\alpha_{\rm acc}(R) = -\bar{W}_{R\varphi}\cdot\left(R\frac{{\rm d}\Omega}{{\rm d}R}\Sigma c_{\rm s}H\right)^{-1} \approx -\bar{W}_{R\varphi}\cdot\left(R\frac{{\rm d}\Omega_{\rm K}}{{\rm d}R}\Sigma c_{\rm s}H\right)^{-1} = \frac{2}{3}\frac{\bar{W}_{R\varphi}}{\bar{P}} \tag{3.6}$$

As a result, the methods typically used in literature overestimate the turbulent $\alpha_{\rm acc}$ by a factor of 1.5 compared to the method of Balbus and Papaloizou (1999).

We also extract the vertical profile of an "effective turbulent vertical mixing" parameter $\alpha_{\rm mix}(z)$ at a reference radius R_0 through $W_{z\varphi}$. Following Stoll et al. (2017), we can approximate the $z\varphi$ component of the viscous stress tensor $\overline{\sigma}$ and write:

$$S_{z\varphi} \approx |q|\rho v_z\frac{\Omega_{\rm K}}{2}\frac{z}{r} \approx \frac{|q|}{2}P\alpha_z h\frac{z}{H} \Rightarrow \alpha_z = \frac{S_{z\varphi}}{P}\left(\frac{|q|h}{2}\frac{z}{H}\right)^{-1} \tag{3.7}$$

and, by replacing the viscous stress tensor with Reynolds stress

$$\alpha_{\rm mix}(z) = \frac{W_{z\varphi}(R_0,z)}{\bar{P}(R_0,z)}\cdot\left(\frac{|q|h}{2}\frac{z}{H}\right)^{-1} \tag{3.8}$$

We note that our method differs slightly from that of Balbus and Papaloizou (1999) and Stoll et al. (2017), in that we compare the velocity field against the *time-averaged mean field* instead of the *initial equilibrium solution*. While this does not affect the v_θ and v_φ components (since the VSI only excites vertical and azimuthal oscillations about the initial equilibrium states), accounting for the now-nonzero mean radial velocity correctly subtracts the disk background which is now accreting with the VSI in effect.

Fig. 12 showcases the difference between using $\bar{u}_R = 0$ and $\bar{u}_R = \langle u_R\rangle_t$ as the disk background velocity field when computing the turbulent stress tensor. While the result is practically the same near the disk midplane (where the bulk of the gas is located), using the time-averaged mean field provides higher-quality results at higher altitudes.

In the following sections, turbulent stress tensor components and the resulting turbulent α parameters are calculated using the method described here, unless otherwise stated.

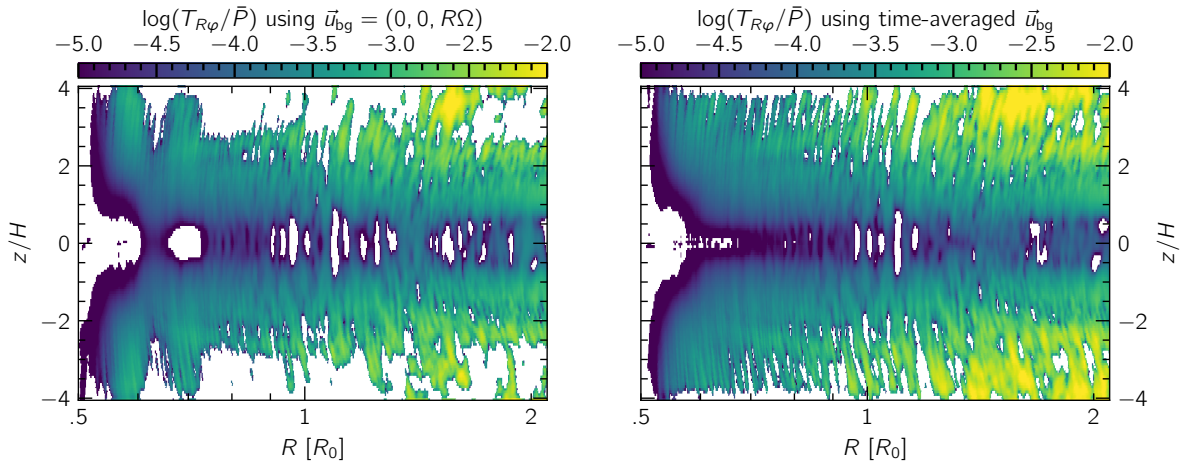

Figure 12. A comparison between two different methods used to extract turbulent motion from the disk background. Left: using the equilibrium state from Eq. (2.3) as the background. Right: using the time-averaged velocity field as the background. White regions denote negative values, which would be unphysical as a result of numerical noise. Using a time-averaged disk background yields better profiles, even though the result of a density-weighted vertical integration (which is needed to compute α) is practically the same.

4 Results I: Impact of planet shock heating on the location and shape of the water iceline (overview)

Here we present our results in Ziampras et al. (2020a), a study that was motivated by a previous project (an MSc thesis) but was considerably enriched during the course of this project, as our understanding of planet–disk dynamics developed. This work was inspired by the possibility of planetesimal growth being affected by the heating induced by planets. For the full publication, see Sect. C.1.

Specifically, we investigated scenarios where shock heating by planetary wakes can raise temperatures high enough to evaporate the icy coating of dust grains, effectively "moving" or deforming the water iceline in the disk and, in doing so, possibly impeding dust growth around it. To that end, we carried out a series of radiative hydrodynamics simulations with planets of different masses in disks with vastly different thermodynamics and investigated the effect of planet shock heating on the location and shape of the water iceline. The latter is defined as the location where ice sublimates into water vapor, or equivalently where gas temperature crosses $T \geq T_{\rm ice} = 170\,\mathrm{K}$ (Lin and Papaloizou, 1985; Bell and Lin, 1994).

Regarding the efficiency of shock heating, we showed that planetary wakes can efficiently heat up the dense and optically thick inner disk ($R \lesssim 5\,\mathrm{au}$), in agreement with Rafikov (2016). This typically results in less tightly wound spiral arms and a radial outward displacement of the water iceline, as the gas temperature increases practically in an axisymmetric manner in such an optically thick medium. More massive planets certainly accentuate this behavior, with Jupiter-sized planets being able to "push" the iceline radially outwards by a few au (left panel of Fig. 13).

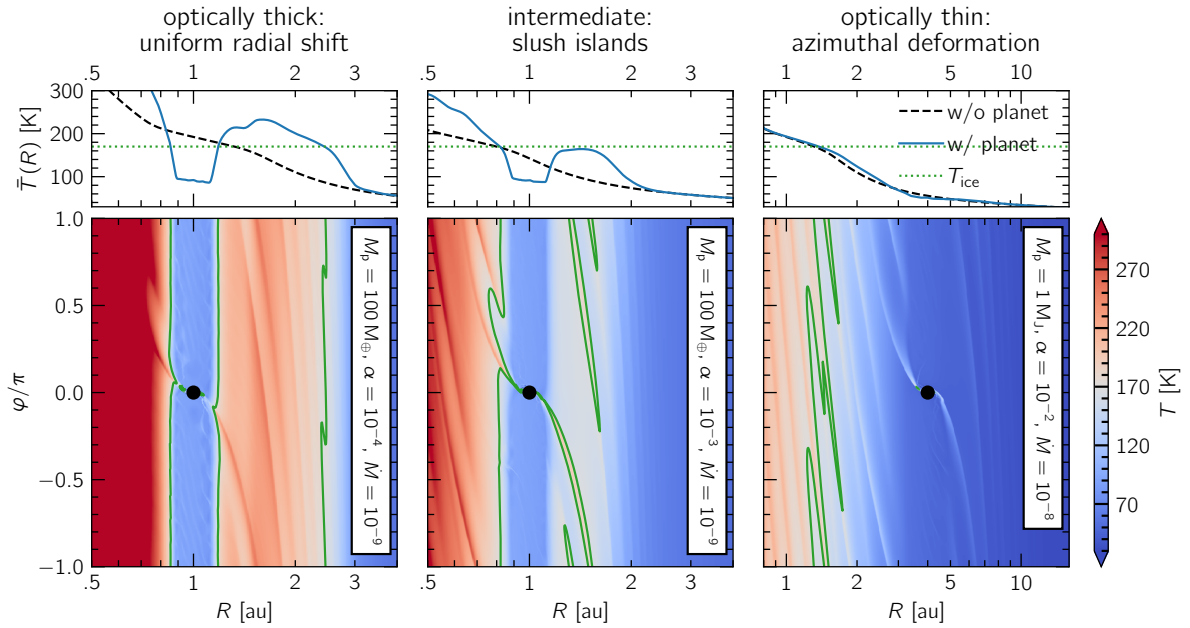

Figure 13. Three examples of how the planet's spiral arms can affect the location and shape of the water iceline (marked with green contours). Left: in an optically thick disk, the iceline is "pushed" radially outwards as the planet's spiral heating heats up the disk uniformly in azimuth. Right: spiral arms cannot heat up the planet's neighborhood efficiently, but the higher temperatures around their crests slightly deform the otherwise axisymmetric iceline as they propagate inwards. Middle: in an intermediate case, hot zones of dry grains and vapor form along the trajectory of the spirals.

The opposite is typically the case for optically thin disk regions, where shock heating is orders of magnitude weaker than the heating by stellar irradiation. In that case, the iceline might show very slight deformation along the azimuthal direction due to the non-axisymmetric nature of the inwardly propagating spiral arms by the planet, but no radial displacement (right panel of Fig. 13).

For certain combinations of intermediate planet masses and disk optical depth, where the iceline would originally lie near the planet's orbit, it is possible for spiral arms to deform the iceline in such

a way that the latter traces the trajectory of these spirals. This could result in the formation of regions containing a mixture of rocky and icy grains, which we dubbed "slush islands" (middle panel of Fig. 13).

Our results highlighted the ability of a planet to effectively heat up optically thick disks via spiral shock heating and quantified this effect in terms of the planet's capacity to move or reshape the water ice line. The latter can have implications on the water content of growing planets, migration torques, and the growth of planetesimals. In particular, the hot zones within "slush islands" could impede the growth of icy aggregates through repeated sublimation and recondensation of ice near the boundaries of such zones.

5 Results II: Impact of radiative effects on gap opening (overview)

Here we summarize the results of Ziampras et al. (2020b), a study which used $R\varphi$ simulations to model the gap opening process for different equations of state. The motivation for this project came from the multitude of ALMA observations of protoplanetary disks by the DSHARP project (Andrews et al., 2018) and the subsequent numerical modeling of those sources in an attempt to explain the rich annular structure observed in those systems with the planet-formation scenario, where a single planet was in some cases capable of explaining the location and contrast of multiple observed rings and gaps (Zhang et al., 2018). For the full publication, see Sect. C.2.

The focal point of our work in Ziampras et al. (2020b) was the equation of state. Specifically, we extended the locally isothermal models of Zhang et al. (2018), where the temperature profile is kept fixed as a function of radius, using a simplified modeling of radiative effects based on Ziampras et al. (2020a). Our aim was to investigate whether the however small but finite cooling timescale introduced in this way could have an effect on the gap opening capabilities of a planet, even though the global temperature profile of the disk would remain unaffected.

Our comparison between locally isothermal and radiative models showed a substantially different radial structure for these two equations of state. Locally isothermal models showed multiple gaps and in general more pronounced planet-generated features (higher-contrast gaps, sharper spiral arms) when compared to the smoother, less rich-in-structure radiative models. These main results are shown in Fig. 14.

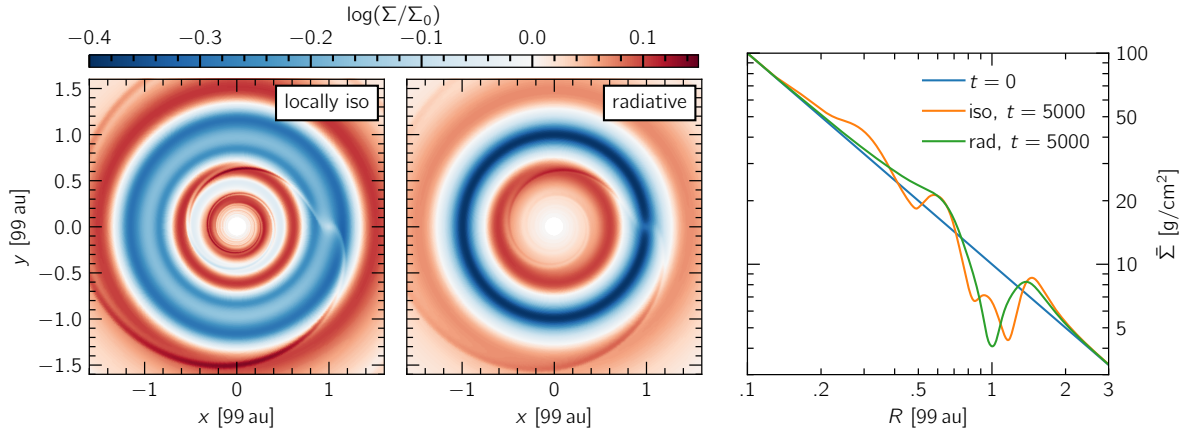

Figure 14. Numerical modeling of the system AS 209 using a locally isothermal or a radiative equation of state at $t = 5000$ orbits at $R = 99$ au. The two models show quite different structure, with the locally isothermal model having a sharper spiral arm contrast and multiple gaps, and the radiative model containing a single, deep gap around the planet.

Our results questioned the scenario where a singular planet is capable of opening multiple observable gaps in realistic protoplanetary disks, where a short but finite cooling timescale regulates disk thermodynamics even at large distances from the host star. This subject was further illuminated by Miranda and Rafikov (2019) and then Miranda and Rafikov (2020a) and Miranda and Rafikov (2020b), who analyzed the effects of cooling on gap opening and showed that, in addition to the above, the consideration of in-plane radiation transport (see $Q_{\rm FLD}$ in Eq. (2.20)) can significantly impact gap opening. Zhang and Zhu (2020) also showed that, even though these disks can be considered stable against the gravitational instability (Toomre, 1964), the inclusion of self-gravity can strongly affect the results of numerical models of planet–disk interaction.

Building on our results and motivated by the kinematic detection of planets at 83 and 137 au in the system HD 163296 (Teague et al., 2018), we constructed a model that could potentially explain the ring structure of the outer disk ($R \geq 60$ au) in that system using three planets, each of which could be present inside the gaps observed at 48, 86 and 145 au (Huang et al., 2018a), using the mass estimates of Zhang et al. (2018). We found that the two different equations of state agree very well at those distances, which

is to be expected given the short cooling timescale at $R \gtrsim 60$ au for this system. We also showed that the structure of the inner disk ($R \lesssim 50$ au) remains unchanged compared to a model with a single planet at 48 au, due to the innermost planet "shielding" the inner disk with its own gap and spiral arms. An overview plot is shown in Fig. 15.

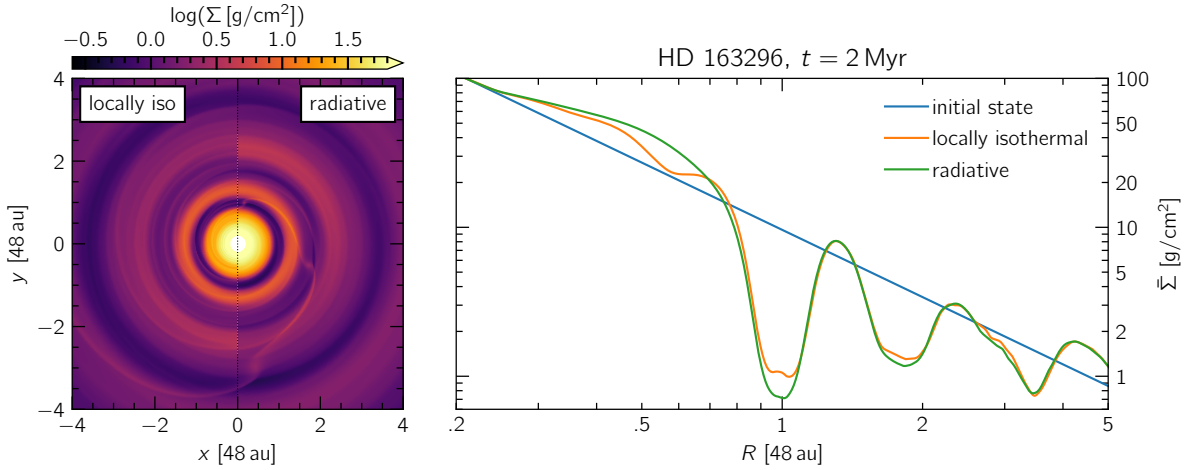

Figure 15. Numerical modeling of the system HD 163296 using three planets at 48, 86 and 145 au and two different equations of state. Left: a surface density heatmap of the disk at $t = 2$ Myr using a locally isothermal (left half) and a radiative (right half) equation of state. Right: a comparison of the azimuthally averaged surface density profiles at the same timestamp for the two different equations of state.

The effect of different physics on planet–disk dynamics has received significant attention over the last few years, and is an ongoing topic of research as the endeavor of theorists and observers to bridge the gap between the two ends of the spectrum continues.

6 Results III: Lifetime of planet-generated vortices (overview)

Here we present a summary of our work in Rometsch et al. (2021). This study was motivated by the observability of non-axisymmetric features in protoplanetary disks around young stellar objects (Andrews et al., 2018) and the possibility that they can be explained via the planet–disk interaction scenario as vortices generated at the planet's gap edge. For the full publication, see Sect. C.3.

The goal of this study was to explore the conditions for the formation of planet-generated vortices and the identification of trends in their lifetime as a function of different physical parameters—primarily the cooling timescale and to secondary importance the effects of viscosity and self-gravity. We contributed with a wide variety of $R\varphi$ numerical simulations using `PLUTO` alongside similar models executed by Thomas Rometsch with the `FARGO` code (Masset, 2000). The two codes were used in order to verify the integrity of our results using two fundamentally different numerical integration techniques (a finite-volume, shock-capturing Riemann solver with `PLUTO` versus a finite-difference upwind scheme with `FARGO`).

Our numerical models showed that planet-generated vortices follow a typical lifetrack similar to that found by previous studies (see Fig. 16), where multiple small-scale vortices form due to the RWI at the planet's outer gap edge and then merge together over a timescale of a few tens of planetary orbits into a massive vortex that decays over 200–2000 orbits, a number that strongly depends on the cooling timescale. We found that for cooling timescales similar to the orbital timescale (see Eq. (2.15), $\beta \sim 1$), cooling-induced vortex decay is strongest and vortices decay much faster compared to shorter or longer cooling timescales (see Fig. 17). We also reported several cases where vortices survived for abnormally long amounts of time ($> 10^4$ orbits), or experienced outward migration far into the outer disk, forming multiple density bumps outside of the planet's orbit.

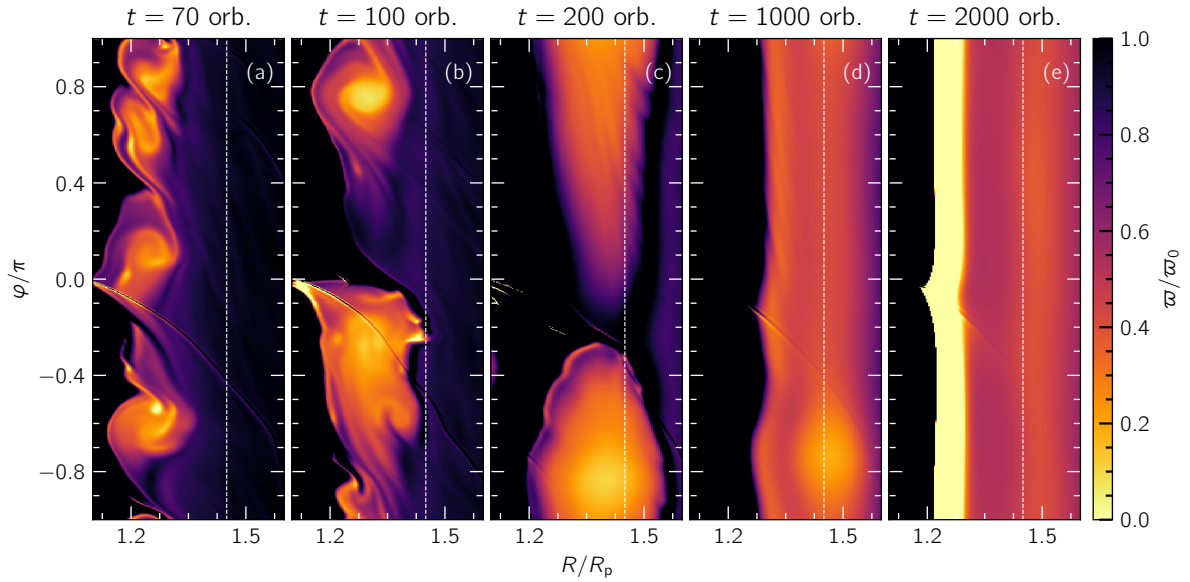

Figure 16. Maps of the gas vortensity as a function of time for a model with $\alpha = 10^{-5}$ and $\beta = 1$. Each bright spot constitutes a vortex at a given stage of its lifetime. Initially, the planet's gap edge grows Rossby-wave unstable and multiple small-scale vortices form on it (a). They quickly merge into bigger vortices (b) and finally into one massive vortex (c) that will decay over 1500–2000 orbits (d), before finally disappearing (e).

Our results suggest that the cooling timescale can influence the formation and lifetime of planet-generated vortices, which sometimes exhibit disproportionately long lifetimes for very short β and decay the fastest for intermediate $\beta \sim 1$. In conjunction with the short cooling timescales at large distances from the star and the relatively small expected turbulence levels at those radii, our results imply that planet-generated vortices should be observable and, inversely, that young planets are a viable candidate for the generation of observed vortices at ALMA-observable distances.

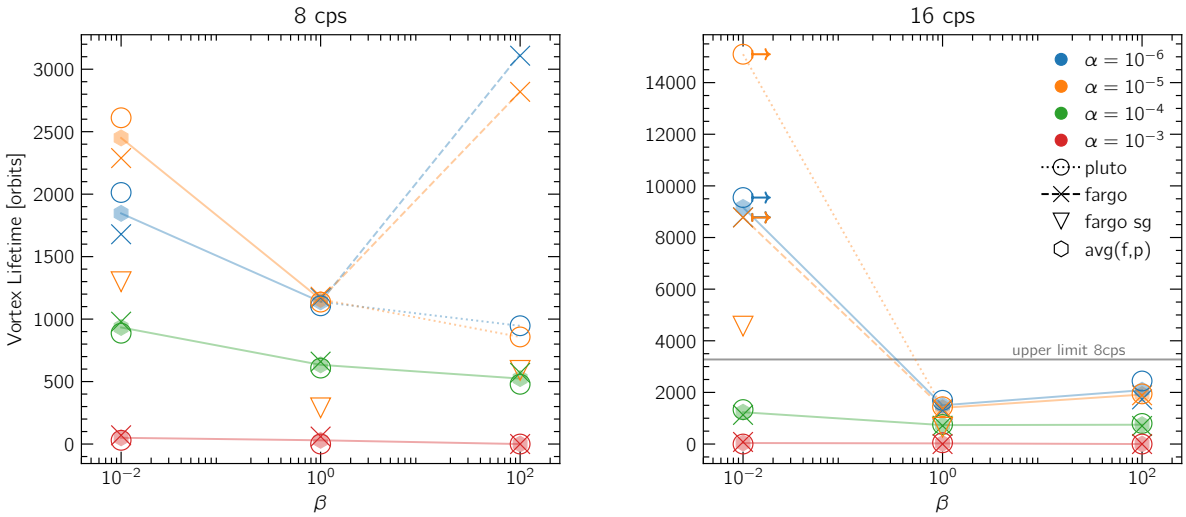

Figure 17. Vortex lifetime as a function of different parameters. Left: results at a resolution of 8 cells per scale height (cps). Vortices live longest for low viscosity, with FARGO exhibiting a "V"-shape centered around $\beta \sim 1$ for very low α. PLUTO does not share this behavior, with vortex lifetimes monotonically dropping with higher β instead. Right: results at 16 cps. Both codes show a minimum in vortex lifetime for $\beta = 1$, but we also observe significantly longer vortex lifetimes for $\beta = 10^{-2}$. These special cases are discussed in detail in the related paper. Image credit: Rometsch et al. (2021) (see Fig. 4 therein).

7 Results IV: Physical properties of the VSI

In this section we discuss our results on the characterization of the VSI. We begin with profiling the behavior of the instability by carrying out a fiducial test using configurations from the literature and measuring the growth rates and saturated turbulent stress levels that the VSI can develop in Sect. 7.1. We continue with an exploration of the effects of different numerical and physical parameters in Sects. 7.2 and 7.3, respectively.

7.1 Characterizing the VSI — Fiducial model

Our fiducial model uses an $r\theta$-2D setup that is similar to Stoll et al. (2017) and described in Sect. 3.2. For quick reference, our setup is summarized in Table I. The grid resolution corresponds to 16 cells per scale height (cps) in the radial and polar directions. The disk is initialized in an equilibrium state, where all velocity components are perturbed by 1% of the local sound speed to break any symmetries and excite VSI modes more easily.

Parameter	value	comment
Radial extent	$r \in [0.4\text{–}2.5]\ R_0$	$R_0 = 5.2$ au, but scale-free
Vertical extent	$z \pm 4H$	constant aspect ratio
Radial grid size	$N_r = 600$	log-spaced, 16 cps
Meridional grid size	$N_\theta = 128$	16 cps, square cells
Inner radial boundary	outflow, damping	$t_{\rm damp} = 10\ P_{\rm boundary}$
Outer radial boundary	reflective, damping	$t_{\rm damp} = 10\ P_{\rm boundary}$
Meridional boundaries	outflow	extrapolated for hydrostatic equilibrium
Aspect ratio h	0.05	constant throughout disk
Density exponent p	-1.5	$\Sigma \propto R^{-0.5}$
Temperature exponent q	-1	non-flared disk
Midplane density $\rho_{\rm mid}$ at R_0	2.2×10^{-11} g/cm^3	does not affect results (scale-free)
Surface density Σ at R_0	213 g/cm^2 $\times \left(\frac{h}{0.05}\right)$	does not affect results (scale-free)
Equation of state	ideal	using β-cooling to relax to initial $T(R)$
Relaxation parameter β	10^{-2}	very similar to locally isothermal EOS

Table I. Physical and numerical parameters used in our fiducial VSI model.

In terms of physics, we use typical disk parameters with $p = -1.5$, $q = -1$ and $h = 0.05$ (see Eq. (2.3)), and an ideal equation of state where we allow the temperature profile to relax to its initial equilibrium state over a constant cooling timescale $\beta = 10^{-2}$. We chose to use a β model over a more conventional locally isothermal one (i.e., where one would not evolve the thermal energy over time) for numerical reasons—namely to combat the parasitic instability discussed in Sect. 7.2.2—and for consistency with our planet–disk models in Sect. 8. The two equations of state (locally isothermal and $\beta = 10^{-2}$) otherwise produce functionally identical results in our area of interest within the disk, as will be discussed in Sect. 7.3.2. Our setup allows us to construct a scale-free problem and measure the properties of the VSI regardless of the reference distance from the star.

Our choice of surface density, while not relevant for our investigation of the VSI, allows us to ignore the effects of disk self-gravity. Namely, the Toomre parameter (Toomre, 1964) is estimated at $Q = 25\,(R/R_0)^{-3/2}$, falling below 9 at $R = 2\,R_0$. Understandably, this surface density would correspond to a relatively low mass reservoir at 100 au, which might be problematic from a planet formation perspective. We discuss this further in Sect. 8.4.

7.1.1 Growth phase and saturated state

We first measure the growth of the VSI by monitoring the total kinetic energy of the active domain in all three directions as a function of time:

$$E^x_{\mathrm{kin}} = \int_{V_{\mathrm{disk}}} \frac{1}{2}\rho u_x^2 \,\mathrm{d}V, \qquad x \in \{r, \theta, \varphi\} \tag{7.1}$$

As individual VSI modes grow during the linear phase of the instability's development, we expect an exponential rise of E_{kin} as a function of time. Following this, modes start interacting, merging, and saturating as they enter the nonlinear phase of development. Eventually the VSI is saturated throughout the disk, and its kinetic energy budget stabilizes. This behavior is shown in Fig. 18. Our measurement of the growth rate at ≈ 0.4/orbit is consistent with the estimate of 0.38/orbit by Stoll and Kley (2014) (see Fig. 1 therein), and reasonably close to that of Nelson et al. (2013), who reported a growth rate of 0.25/orbit during the linear phase (see Fig. 1 therein). We note, however, that the latter used a grid with a narrower radial extent ($r \in [1, 2]$) and therefore a slightly different result was to be expected. The kinetic energy budget stabilizes at similar levels in our model when compared to Nelson et al. (2013) and Stoll and Kley (2014) as well.

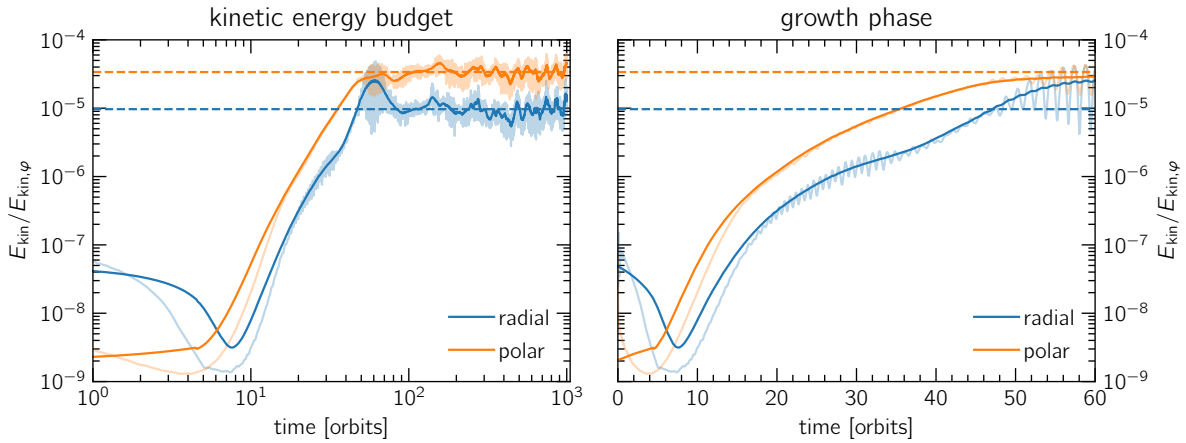

Figure 18. An example of the growth phase and subsequent saturation of the VSI, monitored via the measurement of the global kinetic energy budget of the disk in the radial and polar directions. The initial, linear growth phase can be approximated well with a growth rate of 0.4/orbit. Horizontal dashed lines mark the mean kinetic energy after the VSI has fully saturated. Pale curves denote our raw measurements, which are smoothed into the saturated curves using a rolling average in time.

Once fully saturated, VSI modes cover the entire disk while perturbing all velocity components of the gas. This results in both vertical and (to a lesser extent) radial mixing of the gas as it moves along elongated "sheets" that span the disk from end to end in the vertical direction (see Fig. 19). As a result, the observation of alternating vertical motion near the disk midplane is a typical signature of VSI activity. On the other hand, radial velocity perturbations are strongest at the surfaces of the disk, where the gas that travels towards the surface "turns over" along the radial direction and moves towards the midplane once again. Presumably, this radial rearrangement of gas near the disk surface is the reason why the azimuthal velocity is also affected more strongly at those regions, as u_φ is sensitive to the radial pressure (and therefore density) gradient.

7.1.2 Mean velocity field

In Sects. 2.4 and 2.5 we discussed how the VSI is driven by the vertical shear and, since it feeds on the angular momentum budget of the disk, can drive accretion. While typical α-disks (where α is constant and isotropic) show a radial velocity profile that is positive near the midplane and negative above a certain height, a VSI-active disk shows the exact opposite behavior (Stoll et al., 2017). An example

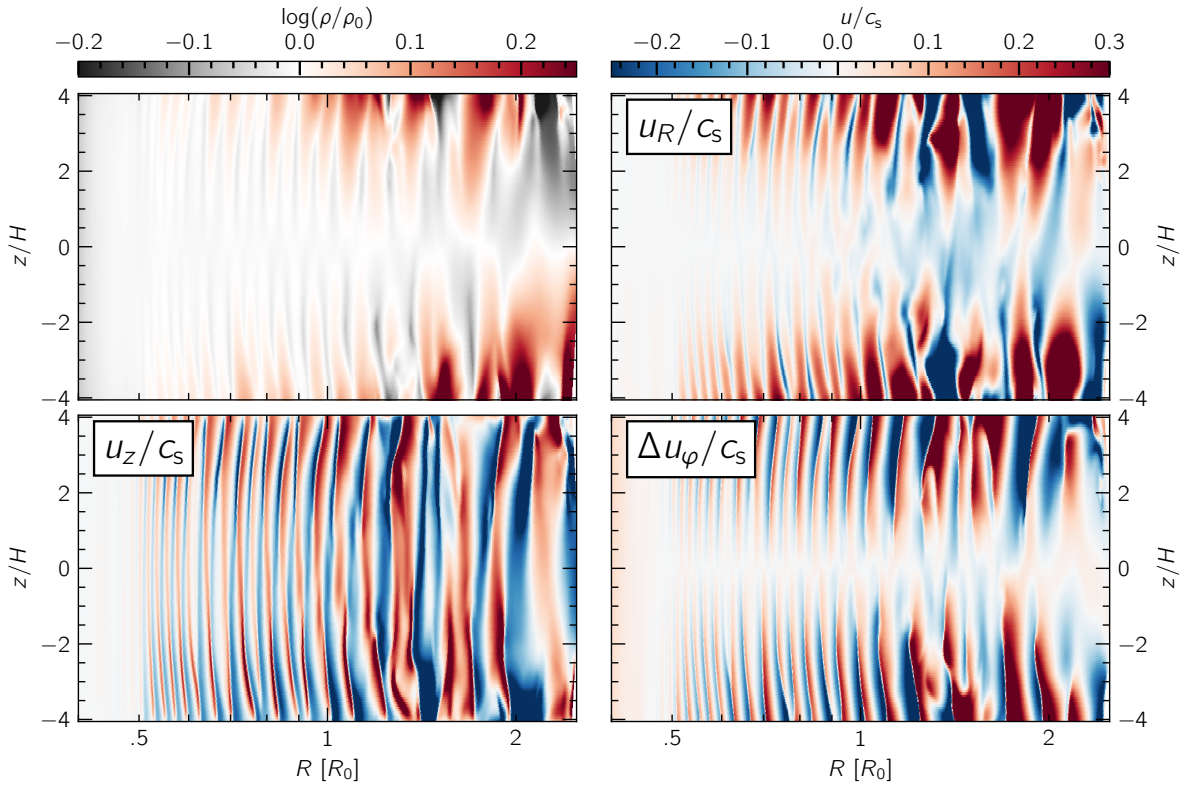

Figure 19. Depiction of a fully VSI-active disk after the instability has saturated. The top left panel shows density deviations, highlighting the movement of gas near the disk surfaces. The remaining three panels show deviations in all three velocity components. The "sheet"-like image in u_z is a characteristic signature of VSI activity, while deviations of u_R and u_φ grow with height as gas "turns over" radially near the disk surface.

of the radial velocity structure of such a disk can be seen in Fig. 20, where the mean profile of $\bar{u}_R$ is computed by time-averaging u_R over a few hundred snapshots once the VSI was fully saturated. A clear accretion channel is visible between $\pm H$ about the midplane, while gas moves radially outwards outside of this zone. We highlight that, since accretion is weighted by mass and the density profile decays exponentially with height, this configuration results in a net infall and therefore accretion. For our disk parameters, the resulting accretion rate is of the order $\dot{M} \sim 10^{-8}$–10^{-7} $M_\odot$/yr.

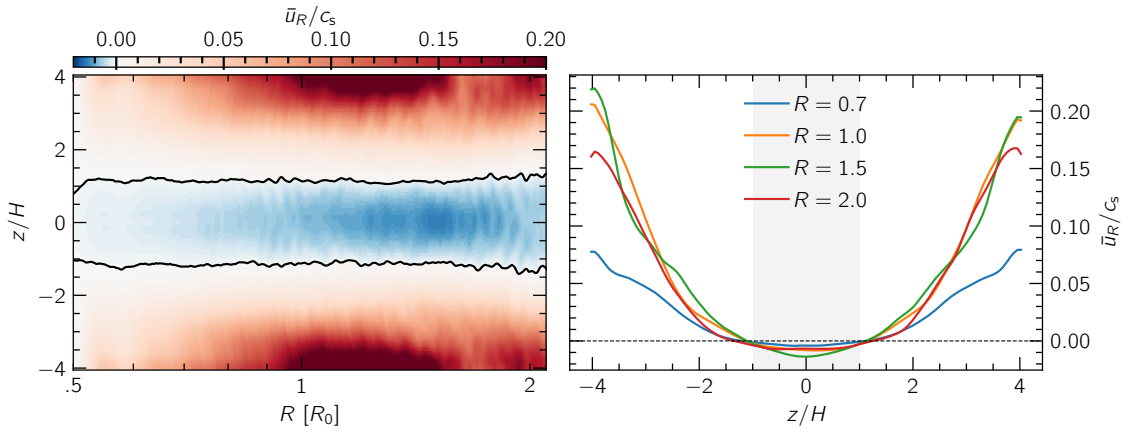

Figure 20. An example of the mean radial velocity field $\bar{u}_R$ in a VSI-active disk. The black contour on the left panel marks the $\bar{u}_R = 0$ surface. An accretion channel can be seen contained within a region $\pm H$ about the midplane. For the physical parameters used in this model (constant h), the ratio of $\bar{u}_R/c_s$ should be constant regardless of radius. The right panel highlights this ratio at several different radii quoted in R_0, showing that this holds true far from the inner boundary.

7.1.3 Turbulent stress

Next, we compute the turbulent viscosity parameter $\alpha_{\rm acc}(R)$ and the turbulent mixing parameter $\alpha_{\rm mix}(z)$, which drive radial accretion and vertical mixing respectively. This is achieved via the method described in Sect. 3.4, where the $R\varphi$ and $z\varphi$ components of the Reynolds stress tensor are averaged in time, integrated vertically, and smoothed radially accordingly. An example of typical stress levels for our fiducial model can be seen in Fig. 21, and the corresponding α are presented in Fig. 22. We find a turbulent $\alpha_{\rm acc}$ of the order of $1.5{\times}10^{-4}$, consistent with estimates of α for these physical and numerical parameters (Stoll and Kley, 2014; Manger et al., 2020). Similar to Stoll et al. (2017), we compute a vertical mixing parameter $\alpha_{\rm mix}$ that is roughly three orders of magnitude higher that $\alpha_{\rm acc}$, but with a ratio of $\alpha_{\rm mix}/\alpha_{\rm acc} \sim 1700$ as opposed to their reported value of 650 (which, after accounting for the factor of $3/2$ in Eq. (3.6), can be corrected to 975).

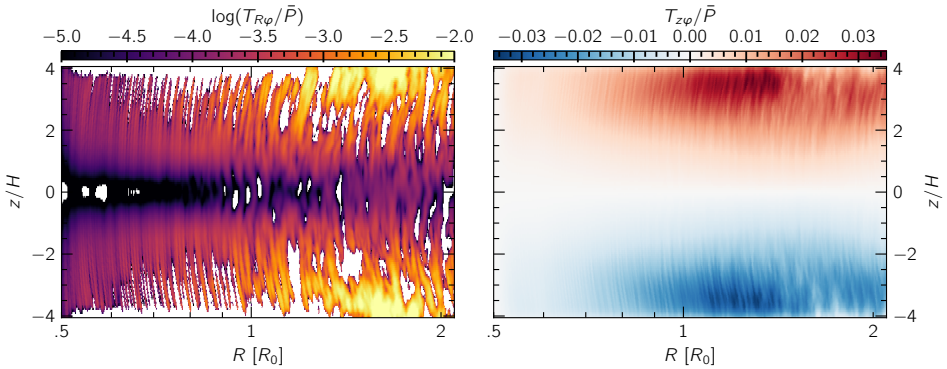

Figure 21. The $R\varphi$ and $z\varphi$ components of the Reynolds stress tensor $\overline{\mathbf{T}}$ as a function of height and radius, normalized to the mean background pressure $\bar{P}$. Left: the quantity shown can be used as a proxy for the accretion parameter $\alpha_{\rm acc}$, which relates to the net radial accretion when integrated vertically. A "V"-shape of $\alpha_{\rm acc}(z)$ about the midplane can be seen. Right: the quantity $T_{z\varphi}/\bar{P}$ roughly translates to a vertical mixing parameter $\alpha_{\rm mix}$. We note that $T_{z\varphi} \gg T_{R\varphi}$, typically by 2–3 orders of magnitude (Stoll et al., 2017).

We note that, due to the relatively narrow radial extent of our grid, the active domain is vulnerable to boundary effects that can damp or interfere with the VSI modes. This can be inferred by the relatively stable levels of α in the region 1–2 R_0, and the decaying levels of turbulence outside of that zone (as shown in Fig. 22). While a wider computational domain helps with alleviating this issue, our area of interest coincides with the region where the VSI has developed properly. For that reason, we chose not to extend the radial extent of our domain.

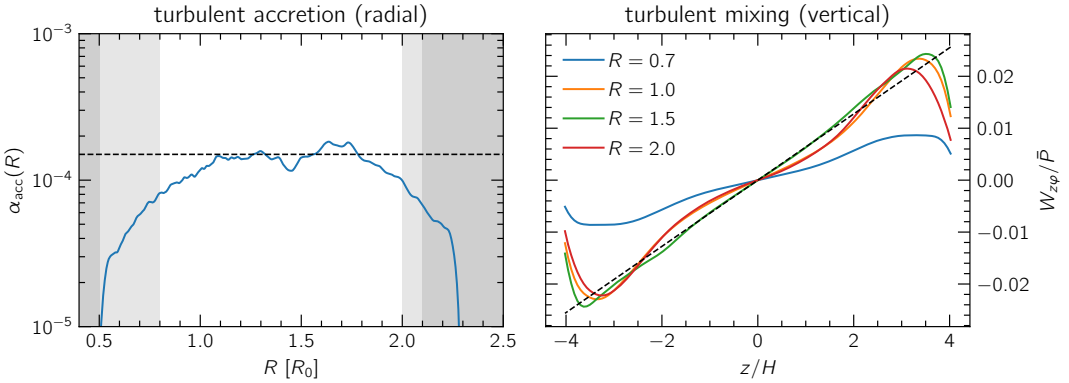

Figure 22. Turbulent stress in our example VSI-active disk. Left: the turbulent accretion parameter $\alpha_{\rm acc}$ as a function of distance. The damping zones are marked in dark gray, and zones likely affected by the damping region are marked in light gray. The horizontal dashed line corresponds to 1.5×10^{-4}. Right: normalized $z\varphi$ stress as a function of height at different radii. The slope of this curve corresponds to the turbulent mixing parameter, estimated at roughly $\alpha_{\rm mix} \approx 1700 \times \alpha_{\rm acc}$ (dashed black line).

7.1.4 Turbulent heating

As we discussed in Sect. 2.4, the turbulent stress generated by the VSI is responsible for accretion (compare Eqs. (2.24) and (2.26)). In addition to this, when translated to an effective "viscous" stress tensor $\overline{\sigma}_{\rm turb}$, Reynolds stress can contribute as a heating source with $Q_{\rm heat} = (\overline{\sigma}_{\rm turb} \cdot \nabla) \cdot \boldsymbol{v}$. Assuming a steady laminar disk heated by uniform, isotropic α-viscosity as in Eq. (2.29) and cooled with a constant β (Eq. (2.16)), we can write that $Q_{\rm heat} \approx \frac{9}{4}\nu\rho\Omega_{\rm K}^2$ (e.g., Tassoul, 1978), and the temperature balance is given by

$$\rho c_{\rm v} \frac{\partial T}{\partial t} = Q_{\rm heat} + S_{\rm relax} \quad \overset{\partial/\partial t=0}{\Longrightarrow} \quad \frac{9}{4}\nu\rho\Omega_{\rm K}^2 = \rho c_{\rm v} \frac{T_{\rm eq} - T_0}{\beta}\Omega_{\rm K}. \tag{7.2}$$

With the relations for $c_{\rm s}$ and H from Sect. 2.1 we can now solve for the equilibrium temperature $T_{\rm eq}$

$$T_{\rm eq} = \frac{T_0}{1 - k\alpha\beta}, \qquad k := \frac{9}{4}\sqrt{\gamma}(\gamma - 1) \approx 1.06. \tag{7.3}$$

In other words, the equilibrium temperature profile will be hotter than the relaxation profile T_0 by a constant factor that becomes larger for higher viscosity or slower cooling. We can now apply this approach to measure the turbulent heating rate in our VSI-active disks by comparing the time-averaged midplane temperature $\bar{T}_{\rm mid}(R)$ to the relaxation temperature T_0, assuming that heating is instead powered by the VSI and equivalent to that of a viscous stress tensor with $\nu_{\rm turb} = \alpha_{\rm acc} c_{\rm s}^{\rm ad} H$. Specifically for the VSI, which can only operate for $\beta \lesssim 0.1$ (Lin and Youdin, 2015) and produces relatively small effective turbulence $\alpha_{\rm acc} \sim 10^{-4}$, we expect very weak—but nevertheless noticeable—heating with $\Delta T/T_0 \lesssim 10^{-5}$.

Solving Eq. (7.3) for α, we obtain that

$$\alpha_{\rm heat} = \frac{1}{k\beta}\frac{T_{\rm eq} - T_0}{T_{\rm eq}} \tag{7.4}$$

we then compute this parameter for our fiducial model ($\beta = 10^{-2}$) and present our results on Fig. 23. The left panel shows temperature deviations, magnified to highlight the slightly hotter midplane. The right panel shows the computed $\alpha_{\rm heat} \approx 2.5 \times 10^{-4}$, which is very close to the accretion parameter $\alpha_{\rm acc} \approx 1.5 \times 10^{-4}$.

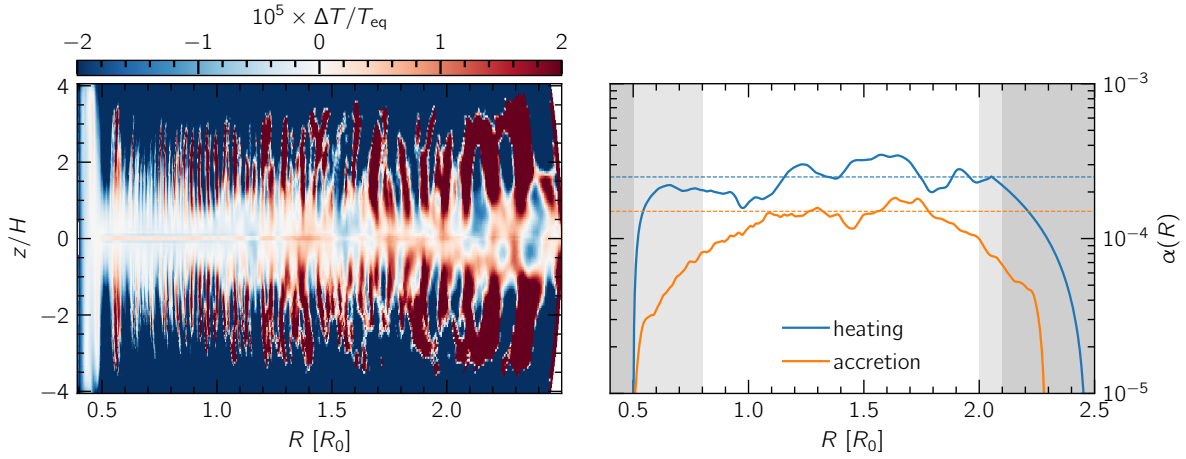

Figure 23. Left: deviations of the time-averaged mean temperature field as a function of radius and height. Right: comparison of the turbulent accretion parameter $\alpha_{\rm acc}$ to the estimated turbulent heating parameter $\alpha_{\rm heat}$ through the method described in Sect. 7.1.4 and Eq. (7.4). The horizontal blue and orange lines correspond to $\alpha = 2.5 \times 10^{-4}$ and 1.5×10^{-4}, respectively.

In the following sections we perform a resolution study with our fiducial model as the baseline, and then continue by exploring different physics as we progress towards a full-3D model with planet–disk interaction in a VSI-active disk.

7.2 Establishing convergence and numerical robustness

Having verified that our setup can produce a fully-saturated, VSI-active disk with turbulence levels consistent with the literature, we now carry out a series of tests at different resolution configurations and using different numerical parameters. In doing so, we can quantify the accuracy of our estimates of the turbulent α and test the robustness of our numerical setup.

7.2.1 Resolution study

The development and saturation of the VSI is the result of the growth of modes of different sizes, with the smallest-scale modes typically growing the fastest. After a local linear stability analysis, we found that the fastest-growing VSI mode has a wavelength of approximately $0.1H$, or would equivalently be resolved over at least 2 cells (the bare minimum to resolve it) at a resolution of 20 cps. Resolving the fastest-growing mode over 4 cells, which might be a reasonable requirement, demands a resolution of 40 cps instead. While we expect the growth rate to change slightly at higher resolution due to that, global turbulence levels do not necessarily depend on the fastest-growing mode but instead on the dominant modes after the VSI has saturated. Since we are interested in the latter, we carry out a resolution study by varying the radial and meridional resolution from 8 up to 48 cells per scale height (cps), with 16 cps being our reference point as discussed in Sect. 7.1.

Regarding the growth rate of the VSI, our results are shown in Fig. 24. We found that resolving smaller-scale radial modes does indeed affect the point in time when the instability begins to develop, and our results converge after doubling the radial resolution to 32 cps (left panel of Fig. 24). However, the measured growth rates are practically the same regardless of resolution at 16 cps or higher. We also found that a resolution of 8 cps in the radial direction (blue curves) is inadequate to correctly resolve the growth phase of the instability.

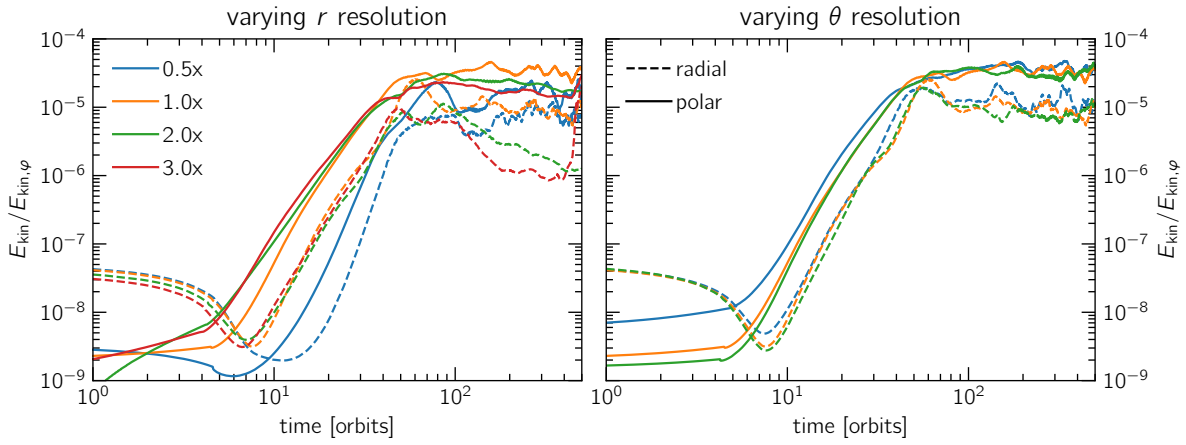

Figure 24. Growth rates and saturation levels of the VSI at different resolutions. The fiducial resolution is $N_r \times N_\theta = 600 \times 128$. Left: increasing the radial resolution results in slightly more accurate measurements of the growth rate, converging at double resolution. The drop in radial kinetic energy for double or triple radial resolution is the result of a parasitic instability developing near the inner boundary (see Sect. 7.2.2). Right: increasing or decreasing the meridional resolution by a factor of 2 does not have a noticeable effect.

On the other hand, changing the resolution in the polar direction yields the same results even when running at 8 cps (right panel of Fig. 24). This is to be expected when one considers the geometry of the instability: the vertically-elongated eddies of gas will correspond to radial wavenumbers k_R that are much larger than their vertical counterparts k_z (according to Nelson et al., 2013, who estimate $k_R/k_z \gtrsim 1/h = 20$ for our setup). As a result, resolution requirements are significantly more demanding in the radial direction. Since our results seem to converge at double resolution in R, it makes sense that we can safely halve or even quarter the resolution in the θ direction and still resolve the saturated VSI modes.

However, a resolution of only 4 cps is relatively low when it comes to resolving gas dynamics (e.g., Rometsch et al., 2021), and therefore we do not run such a test.

Concerning turbulent stress levels, a higher resolution in the radial direction results in a slightly lower α. With a measured $\alpha_{\rm acc} \approx 6\text{–}7 \times 10^{-5}$ at triple resolution, we find that the overall stress levels drop by a factor of 1.5–2 depending on resolution (left panel of Fig. 25). This result is once again consistent with Stoll and Kley (2014) and Manger et al. (2020). On the other hand, $\alpha_{\rm mix}$ behaves similarly ($W_{z\varphi}/\bar{P}$ increases linearly with height, see also Eq. (3.8)) regardless of resolution near the midplane, with the linear behavior extending to higher altitudes z/H at higher cps. Once again, results practically converge at double resolution in R.

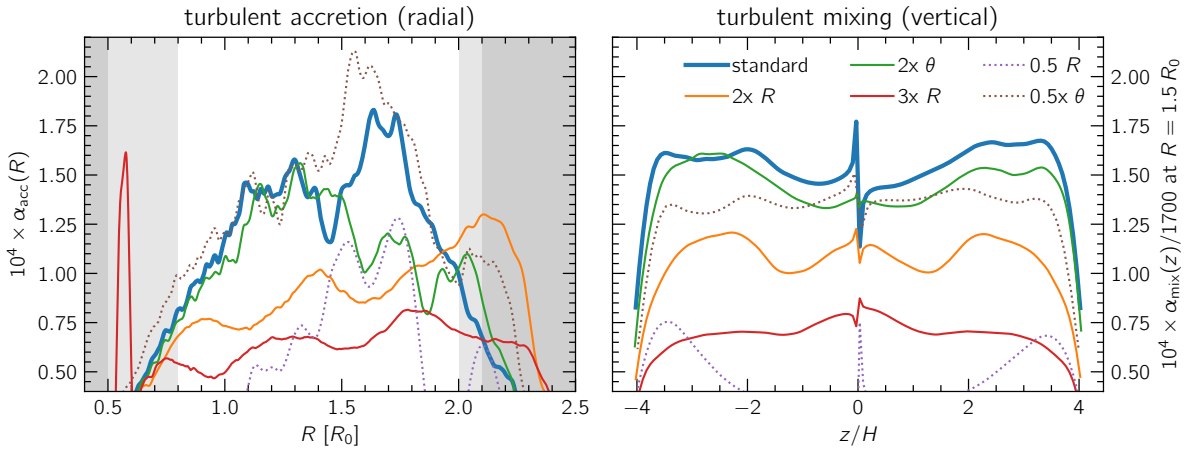

Figure 25. VSI stress parameters $\alpha_{\rm acc}$ and $\alpha_{\rm mix}$ at different resolutions. The fiducial model is shown in thicker blue curves. Dotted curves correspond to models with a resolution lower than the fiducial. Stress levels are lower by a factor of up to 2 for higher resolution. The turbulent mixing parameter $\alpha_{\rm mix}$ as a function of height on the right panel has been normalized by the factor 1700 we found in Fig. 22 and the limits of the y axis have been adjusted to highlight the consistency of our results for standard or higher resolution.

We note that as the resolution increases, so does the number of VSI eddies that are visible in the domain. This has been reported in the past (see for example Stoll and Kley, 2014) and is expected to not converge as the resolution continues to rise. Nevertheless, turbulent stress levels seem to not depend on the absolute number of VSI eddies and therefore resolving as many as possible is not a requirement.

With all of the above in mind, we run all following $r\theta$-2D models in Sects. 7.3.1 and 7.3.2 at a resolution of $N_r \times N_\theta = 1200 \times 128$ cells, which corresponds to 32×16 cps in the radial and polar directions respectively. We opt for such a high resolution such that we can confidently quantify the effect of different physics on our disk model, since $r\theta$-2D runs are reasonably computationally cheap even at such a high resolution. However, for our $r\theta\varphi$-3D models we stick to our fiducial resolution of $600 \times 128 \times N_\varphi$ cells, due to the very high computational cost of such models and given the acceptable results of our fiducial resolution.

7.2.2 Numerics, boundary conditions, parasitic instability

In addition to our resolution study, we tried several different combinations of numerical parameters while profiling the VSI. Below follows a list of parameters that we tested:

- We compared the results of our default solver, HLLC, against the more accurate but also more computationally expensive and slightly less stable Roe solver. We found the same results in terms of growth rates and stress levels regardless of solver, with the only difference being the starting point for the development of the VSI.

- The less-diffusive monotonized central difference limiter (`MC_LIM`) yields the same results as our default Van Leer limiter.

- A parabolic (3rd-order) reconstruction of the Riemann fluxes yielded the same results as our default, linear (2nd-order) scheme.
- For physical reasons we chose an outflow boundary condition in the polar direction, but a reflective (closed) boundary yields functionally the same results.

Our choice of radial boundary conditions (strict outflow with a damping zone that slowly restores the disk profile to its initial state over a damping timescale of 10 boundary orbits) was the outcome of multiple tests where the disk would eventually destabilize well after the VSI was fully saturated. This was the result of a physical or numerical instability that originates at the inner damping zone of the disk and develops in the form of body modes, or vortices in the vertical direction. These modes start from the midplane (one on each side) and grow in both the radial and polar directions, destabilizing the disk as they propagate outwards.

An example of this behavior can be seen in Fig. 26, and a typical signature of this happening is shown on the left panel of Fig. 28 for $h \geq 0.07$ and the right panel of that figure for $h = 0.1$ and $R \leq 0.6$. As it can be seen in these figures, a large aspect ratio strongly encourages the growth of these formations. They also are more likely to grow early for higher resolution, and slightly more so for a locally isothermal equation of state.

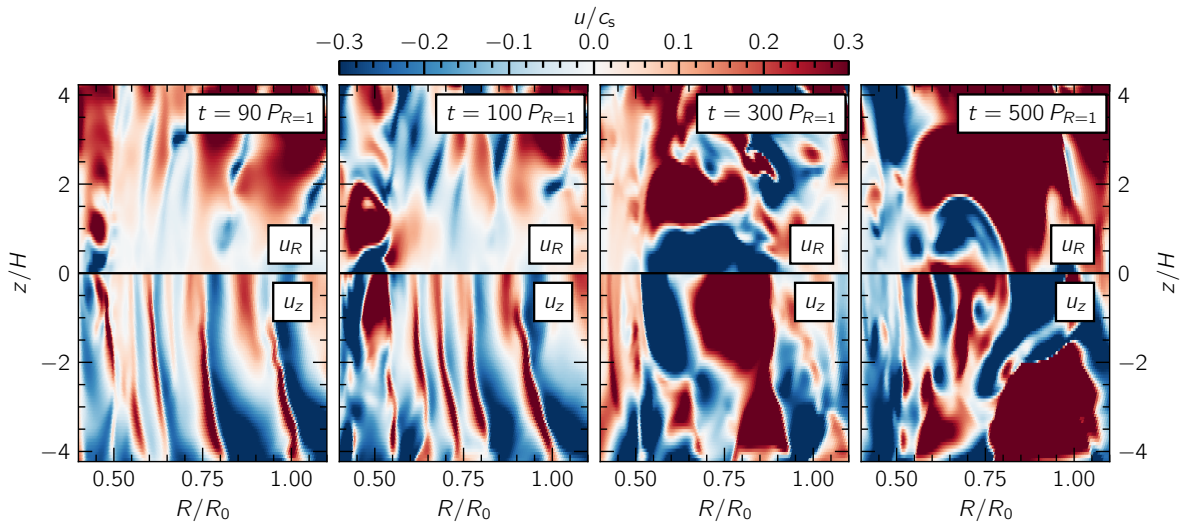

Figure 26. Example of a simulation where the parasitic instability at the inner boundary was left unchecked. The instability develops in the form of two vortices symmetric about the midplane and expands outwards, destabilizing the entire disk. The resulting dataset is unfortunately unusable, as all VSI modes are destroyed in the process.

We found that these unstable modes grow the fastest when using a reflective inner radial boundary and thus opted to open the inner boundary edge. Due to how a rapidly-damping boundary zone can similarly act as a "reflective wall", a damping timescale faster than ∼1–10 boundary orbits was prohibitive. Following Stoll and Kley (2014), we tried adding a small amount of artificial viscosity in the inner damping zone to no effect. Finally, we highlight the importance of damping *all* velocity components (with u_φ being the most important one) to the initial profile in order to limit the development of this instability.

We currently do not know whether this instability has numerical or physical origins. Its strong dependence on resolution and the aspect ratio suggests that there might be a physical mechanism driving it, but its dependence on boundary conditions implies that its growth might be exacerbated due to numerical reasons and should otherwise not be as fast or large. Unfortunately, studying this phenomenon is beyond the scope of this project, and we allow it to exist as long as it doesn't affect our results.

7.3 Exploring different physical parameters

In this section, we test the effect of different physical parameters on the VSI. Namely, since our end goal is the modeling of planet–disk interaction for at least two different reference radii (and therefore reference temperatures), we are interested in the effect of the temperature profile on the VSI. Secondly, we would like to measure the growth and saturation levels of the VSI for different equations of state, and test whether a simplified cooling approach that is compatible with the disk regions of our interest is viable. Finally, it is worth measuring the effect of accounting for the azimuthal extent of the disk and comparing it to our fiducial model, which assumes axisymmetry.

7.3.1 Temperature profiles

We can control the background temperature of the disk by adjusting its aspect ratio profile $h(R)$. Our reference model with $h = 0.05$ translates to $T_{\rm mid} = 121$ K at 5.2 au or 13 K at 50 au, but in the outer disk a higher aspect ratio is typically expected due to disk flaring (e.g., Chiang and Goldreich, 1997; Flock et al., 2020). In that case, $h = 0.1$ might be a suitable upper limit ($T_{\rm mid} = 50$ K at 50 au, 25 K at 100 au). We nevertheless consider the choice $h = 0.1$ to be an upper limit, since h hovers around 0.06–0.08 in the radial range 50–100 au for typical disk parameters (e.g., Ziampras et al., 2020b).

An example of VSI activity for two disks with different radial temperature profiles can be seen in Fig. 27. A cooler disk ($h = 0.05$) shows more, narrower VSI modes when compared to a hotter disk ($h = 0.1$). The hotter disk also shows much stronger radial and vertical motion, especially near the disk surface where the vertical shear rate is strongest.

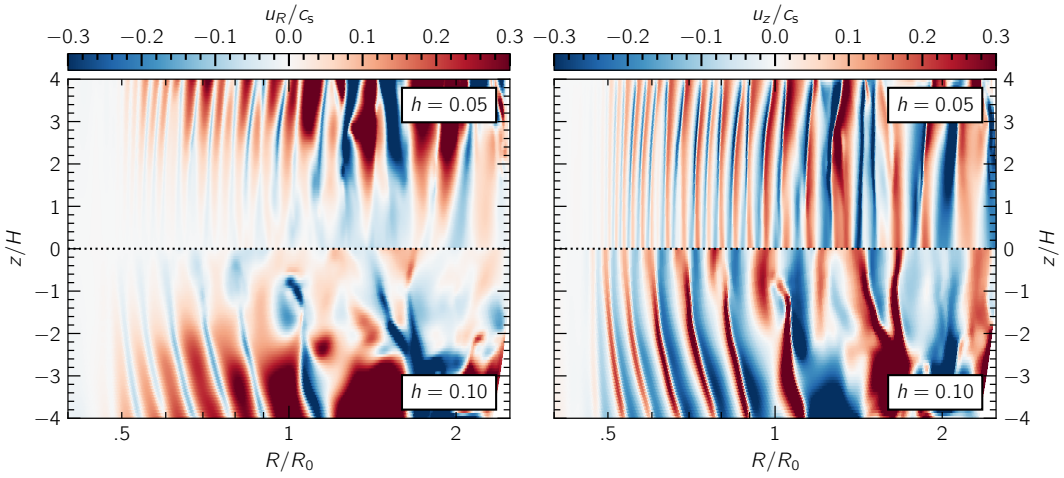

Figure 27. Two VSI-active disks with different reference temperatures, using the aspect ratio h as a proxy. The left and right panels depict the radial and vertical velocity components, respectively. The number of VSI modes depends on h, as can be seen by counting the number of "sheets" on either panel. A hotter disk (higher h) shows more and stronger VSI modes.

Fig. 28 shows a comparison between models of non-flared disks ($q = -1$) with a different aspect ratio. The VSI develops earlier and faster for hotter disks (higher h), consistent with both our theoretical expectations and previous results (Nelson et al., 2013). The corresponding α is also larger for higher h, in agreement with Manger et al. (2020).

Curiously, our computed scaling of α as a function of h yields a power law relation $\alpha \propto h^{1.1}$, contrary to the exponent of 2.6 ± 0.3 that Manger et al. (2020) found. There might be multiple reasons responsible for this contrast:

- Our grid setup uses a radial resolution of 32 cps compared to the 18 cps in that study. Based on our results from the resolution study in Sect. 7.2.1, we are inclined to believe that our models might resolve the VSI more appropriately.

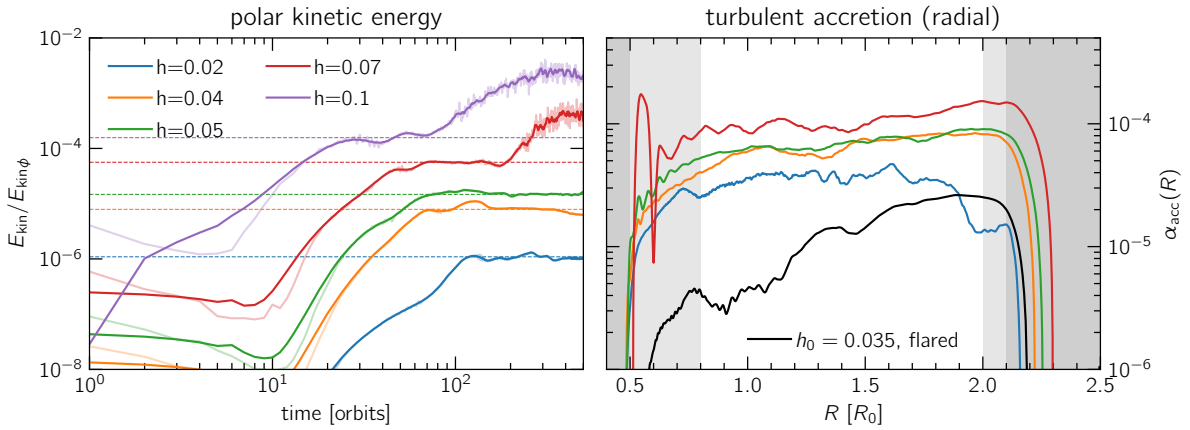

Figure 28. Growth rates and related turbulent stress for disks with different reference aspect ratios. The inner disk is often destabilized for large h, as can be seen by the upward-trending E_{kin} on the left panel shortly after the VSI has saturated. The right panel includes a model of a flared disk where $q = -0.5$ and $h_0 = 0.035$, highlighting the radial dependence of α_{acc} on h.

- We estimate $\alpha(R)$ through a series of time-averaging and radial smoothing, which should yield better results than their estimate of a single $\langle\alpha\rangle = \langle T_{R\varphi}\rangle/\langle P\rangle$, averaged over the entire active domain. Specifically, the latter is subject to pollution from the boundaries (where the VSI might not be active) or disk surfaces (where P drops very low).
- We have already established that a radial range of 0.4–2.5 R_0 is likely not enough to properly compute a robust $\alpha(R)$ as the latter is subject to boundary effects (see Fig. 22).

The combination of the above suggests that the true exponent in the relation between α and h likely lies somewhere between our result and that of Manger et al. (2020).

While our fiducial model does not consider disk flaring ($q = -1$), we discussed previously in Sect. 2.5.1 that the vertical shear rate scales linearly with $|q|$ and thus a flared disk should exhibit weaker VSI activity (Nelson et al., 2013). While we do not focus on flared disks in this project, we provide a comparison with a model with $q = -0.5$ in the right panel of Fig. 28. We note that, since the aspect ratio is a function of R for $q \neq -1$, the normalized shear rate $R\frac{\partial\Omega}{\partial z}\Omega_K^{-1}$ is no longer constant with radius (see Eq. (2.30)) and as a result α grows with distance from the star. Nevertheless, the accretion channel is still contained in the region $|z| < H$ even for the flared model, as can be inferred by the mean radial velocity profile in Fig. 29.

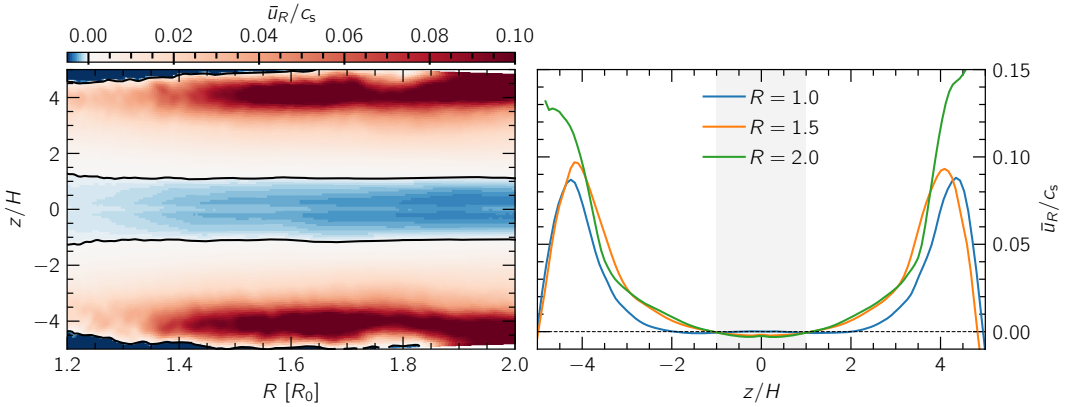

Figure 29. Time-averaged radial velocity field $\bar{u}_R$ similar to Fig. 20, but for a flared disk with $q = -0.5$ and $h_0 = 0.035$. We highlight the accretion channel contained about the midplane within $|z| < H$, even for this flared model. Since now h increases with R so does the vertical shear rate, and as a result $\bar{u}/c_s$ increases in absolute value with radius. This can be inferred by the deeper shades of blue and red with increasing distance on the left panel, or the higher velocities on the right panel.

7.3.2 Equation of state

So far, we have assumed an ideal equation of state with a constant cooling timescale $\beta = 10^{-2}$. For all practical reasons, this is indistinguishable from a locally isothermal equation of state, or equivalently that the cooling timescale of the disk is infinitely short, maintaining $T(R)$ constant with time. This creates optimal conditions for the development of the VSI, as the effects of buoyancy are practically ignored during the linear phase of the development of the instability (see Sect. 2.5.1). Realistically, however, protoplanetary disks have a longer cooling timescale ($\beta \sim 10^{-1}$ even at 50 au) that can also be a function of height and distance (see Sect. 2.2.3 and Fig. 6).

We first verify that our fiducial model with $\beta = 10^{-2}$ is consistent with a locally isothermal model, before setting up models with various β values. For simplicity, similar to the fiducial model, we choose a constant β throughout the entire domain and relax the temperature profile to the initial conditions, where $T(R)$ corresponds to a disk with a constant aspect ratio $h = 0.05$. Then, according to Eq. (2.32) and our choice of parameters for our fiducial model, the VSI should function for $\beta < \beta_{\rm crit} = 0.125$, or $\beta \lesssim \frac{R}{N_z}\left|\frac{\partial\Omega}{\partial z}\right| \approx 0.09$. To test this hypothesis, we run a grid of models for $\beta \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$ and then additional models based on our results until we converge on $\beta_{\rm crit}$.

We consolidate our results on Fig. 30 for selected values of β that highlight the transition of the VSI from fully saturated for $\beta < 0.05$ to inactive for $\beta > 0.1$. This transition seems to be centered around $\beta_{\rm crit} \approx 0.08$–$0.09$, with the $\beta = 0.08$ model generating stress levels weaker by an order of magnitude ($\alpha \approx 8 \times 10^{-6}$) compared to the typical $\alpha \approx 9 \times 10^{-5}$ for the fully saturated cases[3]. Curiously, the VSI seems to operate to some extent in the $\beta = 0.09$ model as well, as seen on the left panel of that figure (brown line), but the turbulence generated dies down over time. Finally we note that stress levels, while roughly the same for $R > R_0$, are noticeably weaker for $\beta = 0.05$ at $R \lesssim R_0$. This is consistent with the results of Manger et al. (2020), who suggested a sigmoidal dependence of α on β instead of a sharp VSI activity cutoff at $\beta_{\rm crit}$.

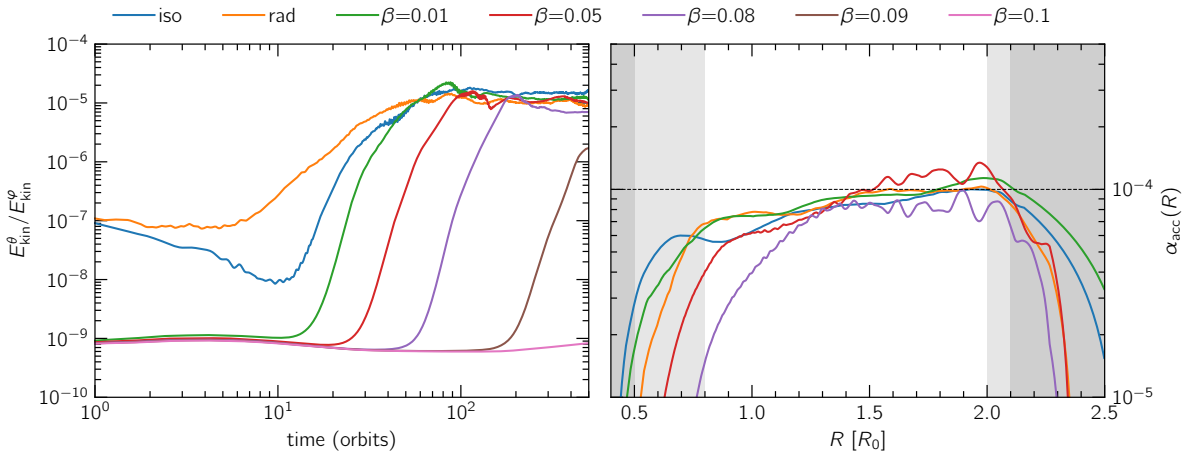

Figure 30. Growth rates and turbulent stress for disks with different equations of state. The computed $\alpha_{\rm acc}$ remains consistent between our locally isothermal setup ("iso") and our β-cooled models as long as $\beta < \beta_{\rm crit} \approx 0.1$. The cutoff around $\beta_{\rm crit}$ is not sharp, however, as can be seen by the slightly weaker stress at $R < R_0$ for $\beta = 0.05$ and the existent but much weaker VSI activity for $\beta = 0.09$. Our radiative model ("rad") agrees very well with our fiducial model. We used a radial smoothing window of $\pm 2H$ on the right panel to make it easier to distinguish different curves.

In addition to these models, we tried starting from the fully VSI-active state of our fiducial model and continuing its evolution using different values of β. We found that our results remain consistent, with models with $\beta < \beta_{\rm crit}$ sustaining the initial turbulence levels. On the other hand, the VSI is completely quenched over 100–200 orbits in models with higher β.

Finally, we would like to test whether a $\beta(R, z)$ profile could affect our results, provided that the conditions for the development of the VSI are met (i.e., β remains sufficiently short). To that end, we

[3]Models at double resolution, see Fig. 25.

ran a model where cooling was treated with the FLD approach discussed in Sect. 2.2.1 and implemented following Appendix A. Since there is no source of heating until the VSI can develop, we make the assumption that the temperature profile is vertically isothermal and that the radiation temperature $T_{\rm rad} = (E_{\rm rad}/a_{\rm R})^{1/4}$ at the vertical boundaries of our disk is equal to the local gas temperature $T_{\rm gas}(R) \propto R^{-1}$. This is a very good approximation as long as our boundary wall is below the $\tau_{\rm abs} = 1$ surface and therefore shielded from direct starlight.

We can verify this assumption by solving Eq. (2.6) for an equilibrium state ($\partial/\partial t = 0$, $u_R = u_z = 0$), from which we find that $T = T_{\rm rad}$ and $\partial T/\partial z = 0$. We confirmed this by modeling a disk using our irradiation module (discussed in Sect. 2.2.1, implemented following Appendix B) and comparing $T_{\rm gas}$ to $T_{\rm rad}$, finding the expected result below the $\tau_{\rm abs} = 1$ surface. We highlight that, while an irradiated disk model implies a flared radial temperature profile ($q \approx -1/2$), we are only interested in comparing against non-flared models while allowing β to be a function of height and radius. Since $T_{\rm rad}$ is equal to $T_{\rm gas}$ regardless of the gas temperature profile, we prescribe $T_{\rm rad} = T_0\, R^{-1}$ where T_0 corresponds to an aspect ratio of $h = 0.05$ at the vertical boundary walls.

For the FLD model, we used the dust opacity model of Lin and Papaloizou (1985) assuming a dust-to-gas ratio of 1%, and a reference radius of $R_0 = 5.2\,$au. The choice of reference surface density $\Sigma_0 = 20\,{\rm g/cm^2}$, while definitely unrealistic, is necessary to allow the VSI to develop as a test with a more realistic $\Sigma_0 = 200\,{\rm g/cm^2}$ did not allow the VSI to grow due to its prohibitively long cooling timescale. We note that for typical protoplanetary disks, the VSI can develop at radii $R \gtrsim 30\,$au (Flock et al., 2020). For our choice of parameters, our low-density model is similar to one for typical ALMA disks at $R_0 \sim 50\,$au.

Our results for the radiative model are incorporated into Fig. 30 with the tag "rad", and show that the VSI is fully active even though β varies with height, as expected. Based on these results, we chose to perform our $r\theta\varphi$-3D runs using the simplified β-cooling approach. This keeps our models as scale-free as possible (since proper radiation transport requires knowledge of the opacities, which are heavily dependent on the local temperature and density), and is far cheaper in terms of computational cost.

7.3.3 3D effects

The VSI is an inherently axisymmetric instability as it drives radial and vertical motion. Nevertheless, the azimuthal component of gas velocity u_φ is still perturbed as a result of this instability (see Fig. 19). It is therefore interesting to measure the impact of 3D effects onto the turbulent stress levels of a VSI-active disk, as well as the geometry of the instability in such a disk. For example, Flock et al. (2020) showed that if only part of the disk is VSI-unstable, it is possible for vortices to spawn on the interface between the stable and unstable zone due to the Rossby-wave instability. Additionally, Manger and Klahr (2018) showed that at a high enough resolution, large-scale vortices can still develop in the midplane even in fully VSI-active disks. Nevertheless, vertical modes remain dominant and the vertical velocity component u_z remains essentially axisymmetric (Barraza-Alfaro et al., 2021).

Fig. 31 shows a model where we extended the fiducial setup in the azimuthal direction using 600 cells (5 cps) in the full range $\varphi \in [0, 2\pi)$ and ran for an additional 200 orbits, similar to Flock et al. (2020). Due to the parasitic instability that develops at $R < 0.5\,R_0$ (discussed in Sect. 7.2.2), several non-axisymmetric modes form in the inner radial boundary and launch spirals that permeate the disk, disrupting VSI modes at $R \lesssim 0.7\,R_0$ but not affecting them outside of that zone. On the same figure a small vortex is visible, marked with a green contour. Fig 32 shows a close-up on the vortex and the planar velocity field around it.

A comparison between our 2D fiducial model and a 3D model is shown in Fig. 33. To calculate the mean field and stress tensor components, we used 150 snapshots between 50–200 P_0 from the start of the 3D run. We find that accounting for 3D effects slightly increases the turbulent stress and by extension α to approximately 2×10^{-4}. This is consistent with the literature (e.g., Stoll and Kley, 2014).

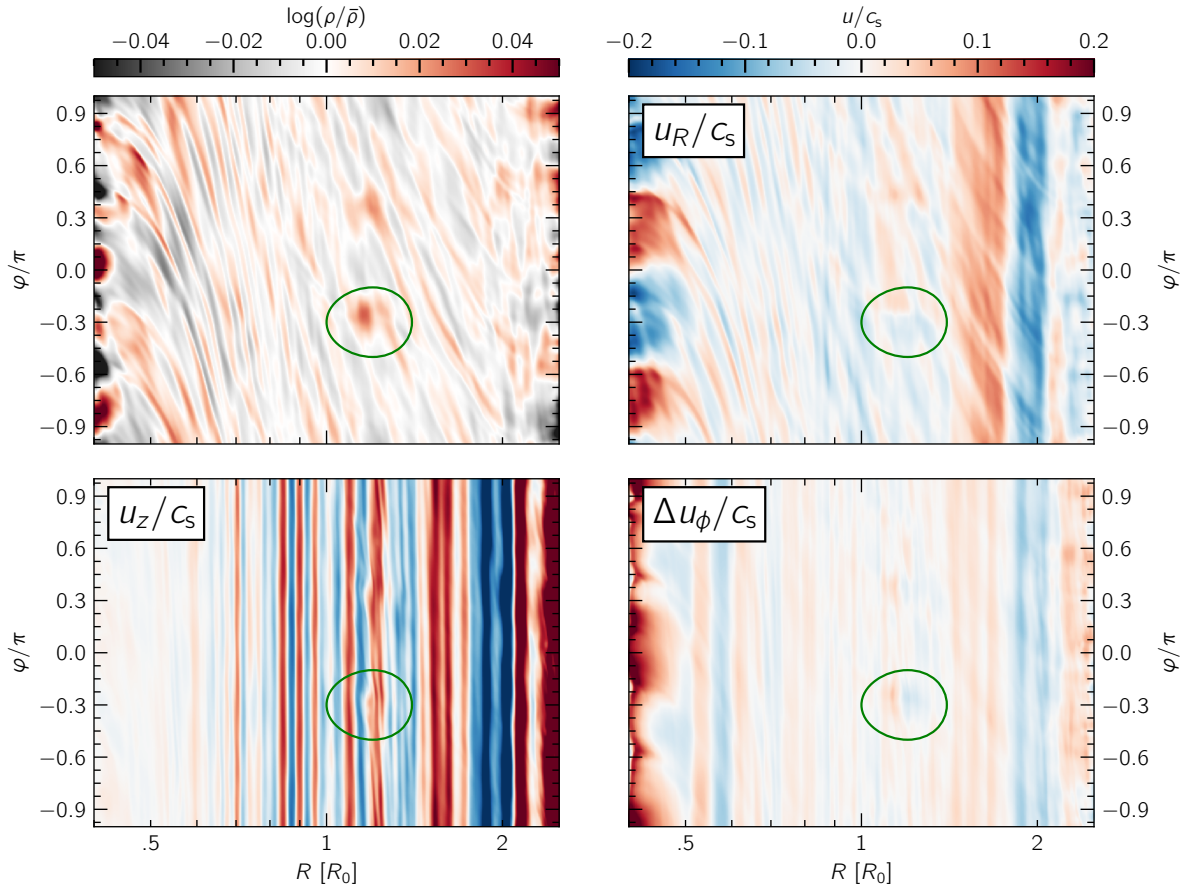

Figure 31. Midplane structure of a VSI-active disk. The top left panel shows density deviations from the azimuthal mean to highlight the non-axisymmetries in the form of spirals and the presence of a small vortex, marked with a green contour. The remaining panels show different velocity components at the midplane. Notable features include the inward drift of gas (u_R shows primarily shades of blue), the rather unperturbed, axisymmetric VSI signatures in u_z, and the swirling velocity field around the vortex. A close-up view on the vortex can be seen on Fig. 32. The inner boundary exhibits the parasitic instability discussed in Sect. 7.2.2, which is likely responsible for most of the spirals throughout the disk and the weaker VSI activity for $R \lesssim 0.7\,R_0$.

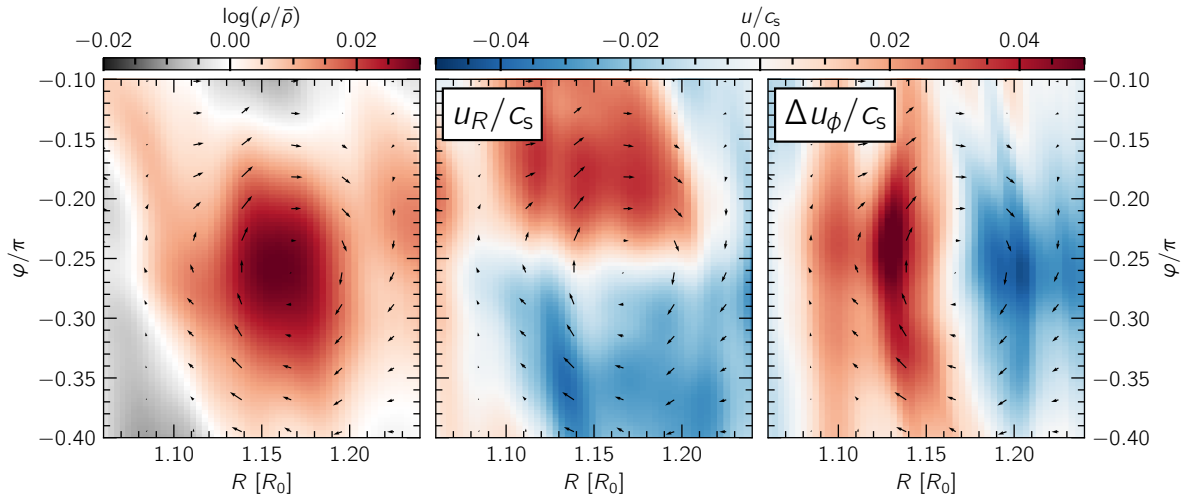

Figure 32. A close-up on the gas density and velocity field of the vortex highlighted in Fig. 31. The arrows denote the direction of streamlines of the planar velocity $(u_R, u_\varphi - R\Omega_K)$. The vertical velocity field, which is largely dominated by VSI motion, is not noticeably affected.

We note, however, that the spiral arms launched by the instability near the inner boundary as well as small-scale vortices contribute to the overall stress by affecting the radial velocity field. Larson (1990) showed that trailing spiral shocks such as those present in our simulations can produce an effective accretion parameter

$$\alpha_{\rm acc} \approx 0.013\ \sqrt{h^3 + 0.08h^2}, \tag{7.5}$$

or $\approx 2 \times 10^{-4}$ for $h = 0.05$. This estimate agrees very well with our findings and hints at the parasitic instability possibly overshadowing VSI activity in terms of accretion efficiency. To isolate the contribution of VSI eddies, we eliminate spiral arms throughout the domain by instead computing $\alpha_{\rm acc}$ using the azimuthally averaged velocity field. We find that, even though the overall stress is higher in 3D, the turbulence produced by VSI modes alone is weakened, with $\alpha_{\rm VSI} \approx 8 \times 10^{-5}$ (green curve on the right panel of Fig. 33).

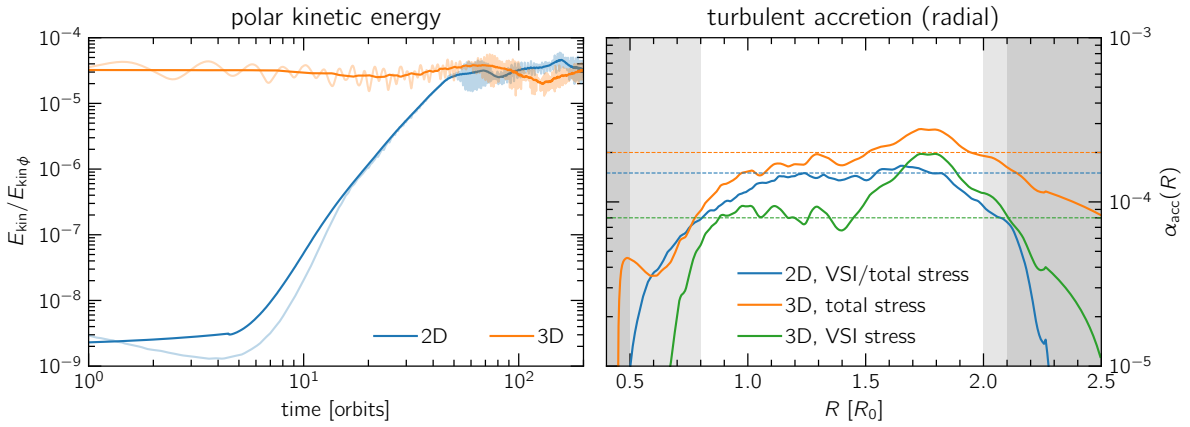

Figure 33. Comparison of the saturation and stress levels of the VSI between our 2D fiducial model and a 3D model. The turbulent accretion parameter $\alpha_{\rm acc}$ is slightly larger in the 3D model, due to the effect of spirals on the velocity field. Isolating the contribution of the VSI yields the green curve on the right panel instead.

Spiral arms launched by vortices act on the $R\varphi$ plane, in contrast with the VSI which generates vertical motion. While these spirals can induce very efficient accretion—often overshadowing the VSI—as discussed above, it is the VSI that is ultimately the primary driver of vertical mixing in the gas. This implies that as spirals and other non-axisymmetric features along the midplane perturb or even destabilize VSI modes, they also inhibit the ability of the VSI to lift material off the midplane even though the radial accretion profile is otherwise unaffected or even enhanced. In other words, while the effective viscosity generated through spiral shocks results in a higher $\alpha_{\rm acc}$ compared to that by the VSI alone, the turbulent mixing parameter $\alpha_{\rm mix}$ is lower in regions where spirals and vortices interfere with VSI activity. To highlight this behavior, we compare $\alpha_{\rm mix}(z)$ computed through Eq. (3.8) at two different radii for the models shown on the right panel of Fig. 33 in Fig. 34. We will revisit this argument in Sect. 8.3, where we will show how an embedded planet changes this picture.

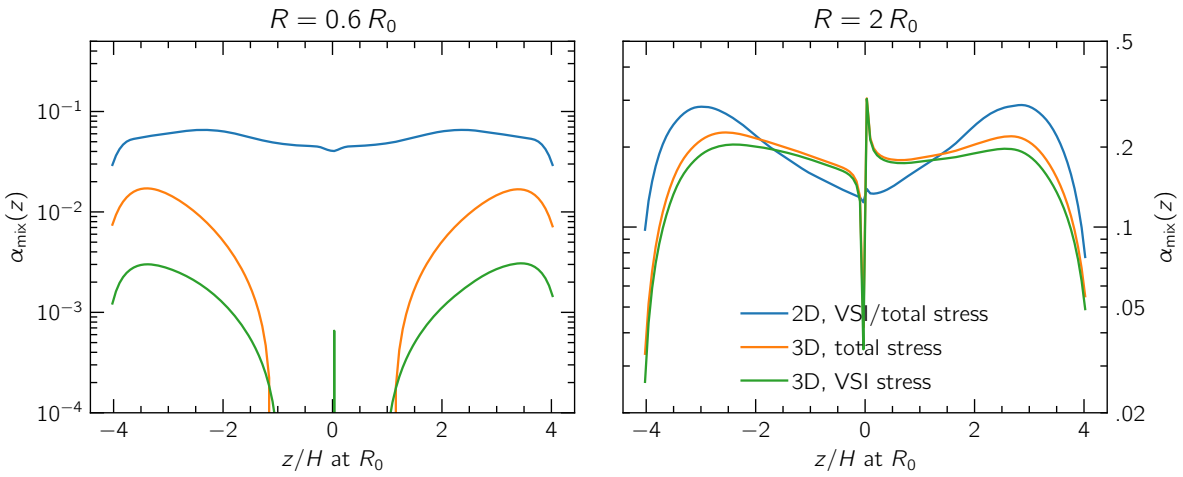

Figure 34. Vertical profiles of $\alpha_{\rm mix}$ at two different reference radii for the fiducial $r\theta$-2D model compared against an $r\theta\varphi$-3D model with the same parameters. The corresponding $\alpha_{\rm acc}$ is shown on the right panel of Fig. 33. Left: the contribution of the VSI to $\alpha_{\rm acc}$ drops significantly near the inner boundary for the 3D model, and as a result so does vertical turbulence. Since the VSI is the primary mechanism that drives vertical motion, $\alpha_{\rm mix}$ is very low when compared to the 2D model even if the stress by spiral arms is considered. Right: at $R = 2\,R_0$, VSI activity produces a total $\alpha_{\rm acc}$ that is higher than its 2D counterpart. This is reflected in the higher $\alpha_{\rm mix}$ for $|z| < 2H$. Once again, the orange and green curves agree very well because spiral arms only contribute to a very small fraction of $\alpha_{\rm mix}$, which is primarily supplied by VSI turbulence.

8 Results V: Planet–disk interaction in VSI-active disks

In this section we analyze planet–disk interaction in VSI-active disks. We investigate the impact of the planet's presence on VSI activity via the launching of spiral arms and the generation of vortices, and the opposite via planet migration in a VSI-active disk.

Our setup is similar to the 2D fiducial model in Sect. 7.1 in terms of physical parameters ($p = -1.5$, $q = -1$) and resolution ($N_R^{\rm cps} \times N_\theta^{\rm cps} \times N_\varphi^{\rm cps} = 16 \times 16 \times 5$). We use an ideal equation of state, and allow the disk to relax to its initial temperature profile with a constant $\beta = 10^{-2}$ to counteract the planet's shock heating while still allowing the VSI to fully develop.

To speed up calculations, we start from the quasi-equilibrium state of the VSI-active disk from a 2D model that has evolved for 1000 orbits. We expand the model into 3D by copying all fields N_φ times in the azimuthal direction, and then perturbing all velocity components by 1% of the local sound speed, similar to the method of Flock et al. (2020). We then run the model for an additional 200 orbits (total time: 1200 orbits) to reach a new quasi-steady state. Next, we insert the planet and allow it to grow to its final mass over a timescale of 100 orbits (total time: 1300 orbits) according to the formula of de Val-Borro et al. (2006) while keeping it fixed at $R_{\rm p} = R_0$, $z_{\rm p} = 0$. After 500 orbits (total time: 1700 orbits) we confirm that the disk has reached a new quasi-equilibrium state, and allow the planet to feel the disk feedback and possibly migrate. During this time period, the star feels the disk's gravitational pull as well. This continues for 300 orbits (total time: 2000 orbits). This procedure is summarized in Fig. 35. For simplicity, we refer to the time when the planet begins growing ($t = 1200\,P_0$) as t_0.

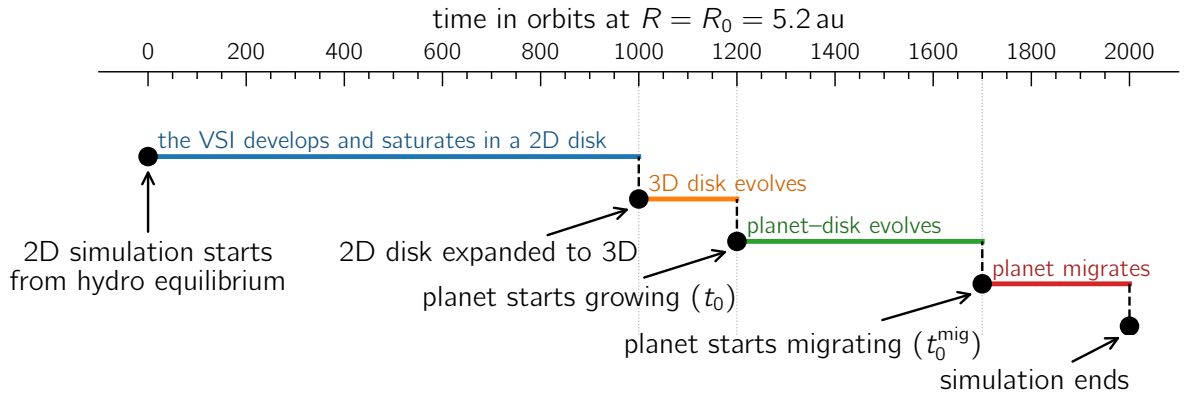

Figure 35. Timeline of 3D models involving planets. The planet is inserted at $t_0 = 1200\,P_0$ and grows over 100 orbits. It then starts migrating at $t_0^{\rm mig} = t_0 + 500\,P_0 = 1700\,P_0$.

To explore the behavior of both VSI activity and planet–disk interaction, we choose to vary the planet mass $M_{\rm p} \in \{0.3, 1\}\,{\rm M_J} = \{3 \times 10^{-4}, 10^{-3}\}\,{\rm M_\odot}$ and the reference aspect ratio $h \in \{0.05, 0.1\}$ for a total of 4 models. Our choices result in a grid of simulations that explore the full range from the linear, low-mass regime to the highly nonlinear, high-mass regime of planet–disk interaction.

We begin our analysis by identifying and monitoring the development of planet-generated features such as spiral arms, gaps and vortices in Sect. 8.1. We then continue by investigating the extent to which the VSI can function alongside a planet in protoplanetary disks in Sect. 8.2. Finally, we take a look at the planet migration scenario in a VSI-active disk in Sect. 8.4.

8.1 Planet–disk interaction: spiral arms, gaps, rings, and vortices

As we discussed in Sect. 2.6, planets interact with the disk around them by launching spiral arms that transport angular momentum radially, restructuring the gas distribution in the disk. In the case of high-mass planets, it is also possible for one or more gaps to open near the planet's orbit. To differentiate between the low- and high-mass regimes we can define the thermal mass $M_{\rm th}$ (Zhu et al., 2015)

$$M_{\rm th} = \left.\frac{{c_s}^3}{G\Omega_K}\right|_{R_p} \approx 1\,M_J \left(\frac{h_p}{0.1}\right)^3 \frac{M_\star}{M_\odot} \tag{8.1}$$

We expect to see nonlinear effects such as gap opening for $M_p \gtrsim M_{\rm th}$ (Lin and Papaloizou, 1993), and no such effects otherwise. Since the development of vortices through the RWI is largely tied to the formation of steep surface density gradients (e.g., Hammer et al., 2019; Rometsch et al., 2021), we expect to only see vortices in models with planets in the high-mass regime as well. A summary of our planet–disk models in the context of the thermal mass is provided in Table II.

h	M_p	M_p [$M_\odot$]	M_p [$M_{\rm th}$]	regime
0.10	$100\,M_\oplus$	3×10^{-4}	0.3	linear
0.10	$1\,M_J$	1×10^{-3}	1.0	moderately nonlinear
0.05	$100\,M_\oplus$	3×10^{-4}	2.4	nonlinear
0.05	$1\,M_J$	1×10^{-3}	8.0	strongly nonlinear

Table II. Physical and numerical parameters used in our 3D models of planet–disk interaction.

Fig. 36 showcases a typical image of planet–disk interaction in the high-mass regime ($M_p = 100\,M_\oplus = 2.4\,M_{\rm th}$ at $h = 0.05$). The two panels show a Cartesian (xy) and polar ($R\varphi$) view of the surface density structure of the disk after 500 planetary orbits. Various planet-generated features such as the deep gap around $R = R_p$, the shallower secondary gap at $R = 0.6\,R_p$ and multiple vortices are annotated.

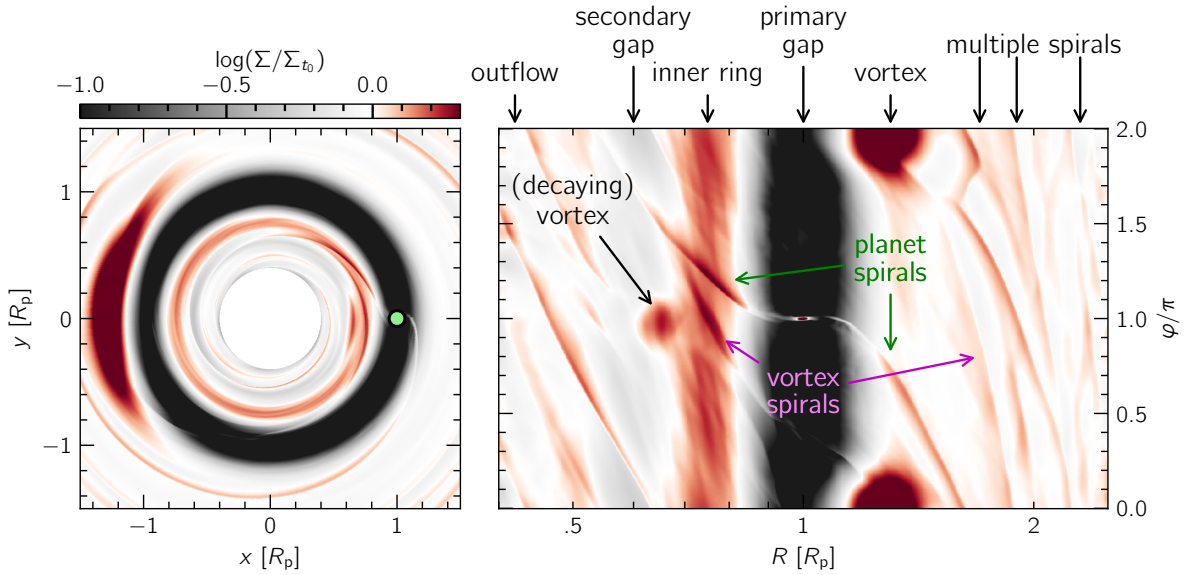

Figure 36. An example of planet–disk interaction in the high-mass regime ($M_p = 2.4\,M_{\rm th}$), similar to Fig. 9 but with a full-3D model. Left: a Cartesian (xy) view of the surface density structure, showing a clear gap around the planet's orbit and a prominent vortex orbiting at its outer edge. Spiral arms permeate the disk. Right: a polar ($R\varphi$) view with a breakdown of most relevant planet-generated features, with annotations that highlight their location and origin. The secondary vortex at $R \approx 0.65\,R_p$ is more easily visible here, and the elliptical shape of vortices in the $R\varphi$ plane becomes clear.

Vortices, in particular, can launch their own spiral arms and interact with the disk as they decay or migrate from the gap edge that they formed at (Paardekooper et al., 2010; Rometsch et al., 2021). As a

result, it is often difficult to isolate the contribution of vortices to the overall disk profile while one or more vortices are present. Examples of such behavior can be found at the outer disk on the right panel of Fig. 36, where planet- and vortex-generated spirals begin overlapping, and near the outer density bump at $R \approx 1.3\,R_\mathrm{p}$ on the bottom right panel of Fig. 38, which is in part exaggerated due to the presence of the vortex. Nevertheless, vortices follow their own life cycle and eventually dissipate over time, as can be seen on Fig. 37 for the vortex in the inner disk, whose lifetime of $\approx 10^3$ local orbits is consistent with the findings of Rometsch et al. (2021) for a disk with $\alpha \sim 10^{-4}$, $\beta = 10^{-2}$ and $h = 0.05$.

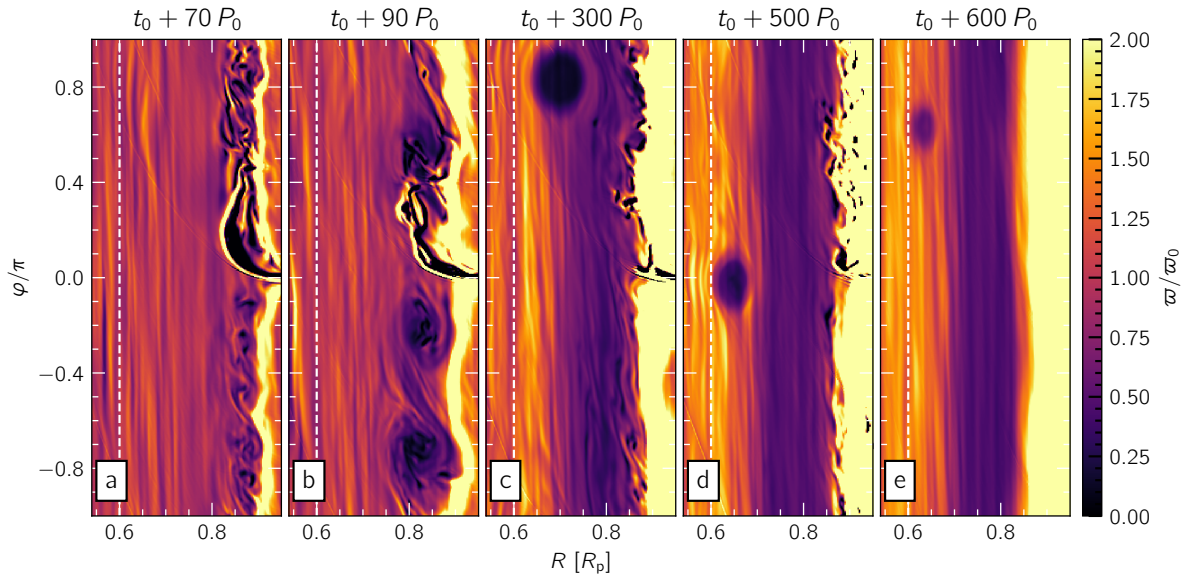

Figure 37. Time evolution of the vortensity $\varpi = \frac{1}{\Sigma}(\nabla \times \boldsymbol{v}) \cdot \hat{z}$ in the inner disk, normalized to the Keplerian vortensity $\varpi_0 = \frac{1}{2\Sigma_0}\Omega_\mathrm{K}$. This quantity is a good tracer of vortex formation, as can be seen by the development of dark spots that form near the planet's inner gap edge (a). As time advances from left to right, the gap edge becomes steeper and the multiple small vortices that form on it (b) eventually merge into a large vortex (c) that migrates inwards as the gap continues to widen. The vortex then dissipates over time (d) as turbulent diffusion and angular momentum transport along its spiral arms drain its mass reservoir, restoring the local vorticity to its Keplerian background $\varpi_0\Sigma_0$ (e). The vertical dashed white line helps guide the eye.

Regarding the ring and gap structure of the disk in the high-mass regime (see Fig. 38), the Saturn-sized planet exceeds the thermal mass for $h = 0.05$ and therefore opens a deep gap around its orbit with $\Sigma/\Sigma_0 < 0.1$ (using the definition of a gap by Crida et al., 2006). The pressure bump formed by the accumulation of gas on the outer gap edge also forms a ring at $R = 1.3\,R_\mathrm{p}$. The combination of a very low viscosity ($\alpha \sim 10^{-5}$–10^{-4}) due to the VSI and a functionally locally isothermal equation of state for $\beta \ll 1$ allow the planet to open a secondary gap at $R = 0.6\,R_\mathrm{p}$ (Miranda and Rafikov, 2020b). The radial position of the secondary gap is consistent with the fit of Zhang et al. (2018) for $\alpha \sim 10^{-5}$–10^{-4}. The carving of this secondary gap also results in the formation of a secondary ring, peaking at $R \approx 0.75\,R_\mathrm{p}$.

On the other hand, the same Saturn-sized planet only amounts to $M_\mathrm{p} = 0.3\,M_\mathrm{th}$ in a disk with $h = 0.1$. As a result, planet–disk interaction falls in the linear regime, as can be seen by the lack of a gap around the planet's orbit and the good agreement between the expected and simulated position of spiral arms on the left panel of Fig. 39. The disk simply continues exhibiting strong VSI activity, and accretes steadily onto the star over time as can be seen on the bottom right panel of that figure.

In both cases, a cross section through φ_p reveals the vertical density stratification of the disk. Planet- and vortex-generated spirals appear as red (overdense) sheets in the top right panel of Figs. 38 and 39 and can be seen covering the full vertical extent. The gas trapped within the planet's Hill sphere can also be seen in those panels.

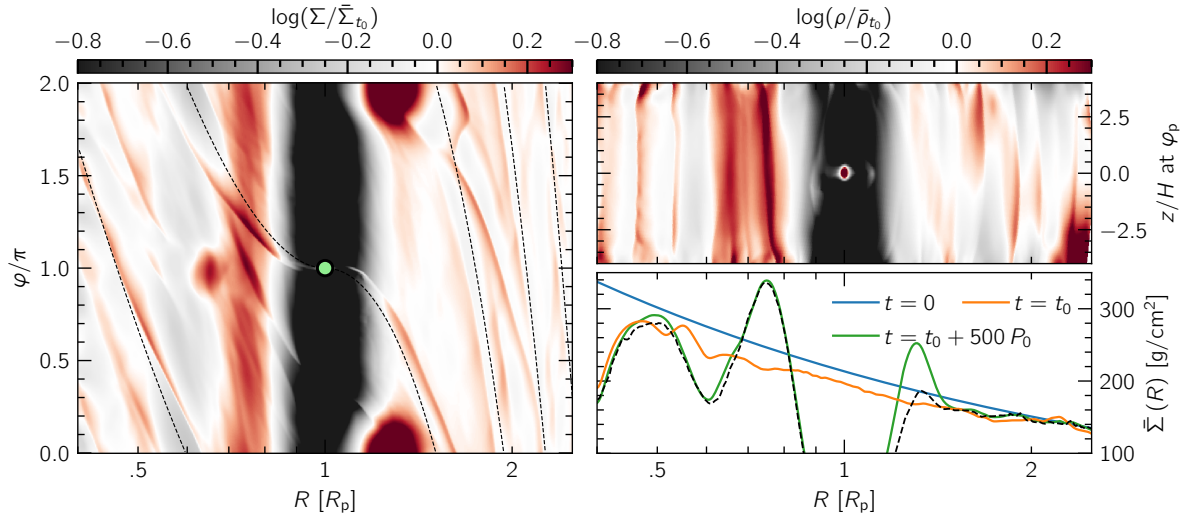

Figure 38. Gas structure in the high-mass regime ($h = 0.05$, $M_\mathrm{p} = 100\,\mathrm{M}_\oplus = 2.4\,M_\mathrm{th}$) after 500 planetary orbits. Left: Σ across the disk, similar to Fig. 36. Dashed black curves mark the position of spiral arms using linear theory (Ogilvie and Lubow, 2002). Top right: vertical density stratification at φ_p, highlighting the vertical structure of spiral arms and the gas accumulation around the planet's Hill radius. Bottom right: the azimuthally averaged disk surface density $\bar{\Sigma}$ with and without a planet (green and orange curves respectively), showing the development of radial structures like gaps and rings. We note that the use of an azimuthal average shows a ring with seemingly high contrast at $R = 1.3\,R_\mathrm{p}$, but most of the gas is instead accumulated around the vortex at that radius. Using the azimuthal median (dashed black line) corrects for this bias.

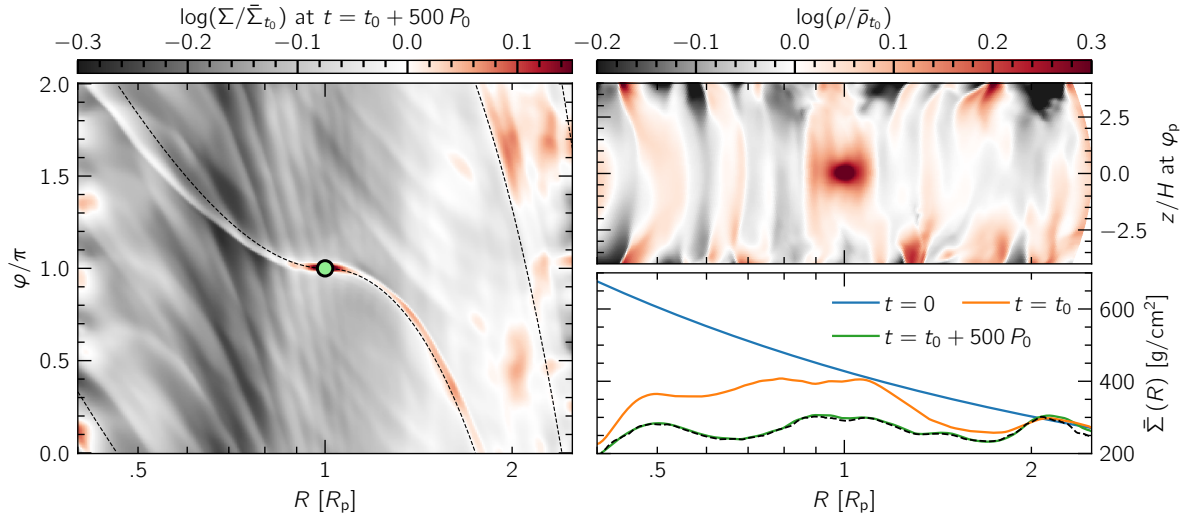

Figure 39. Similar to Fig. 38 for the low-mass regime ($h = 0.1$, $M_\mathrm{p} = 100\,\mathrm{M}_\oplus = 0.3\,M_\mathrm{th}$). In contrast to the previous figure, the planet is unable to open a gap and the disk continues to accrete in an axisymmetric fashion due to the VSI (left and bottom right panels). The planet's Hill sphere is more massive in part due to the lack of a gap (top right panel). Here, the lack of a gap results in the RWI being unable to drive vortex formation.

8.2 VSI activity in the presence of a planet

Planet-generated features excite non-axisymmetric motion in the disk midplane in the form of spiral arms and—in the high-mass regime—vortices. In addition, the gap opening process induces steep radial density gradients. In this section, we examine the extent to which these features can affect the activity of the VSI. We note that the planet's presence quenches the parasitic instability near the inner radial boundary (see Sect. 7.2.2), so the development of VSI modes at $R < 0.7\,R_\mathrm{p}$ might now be possible.

We first carry out our analysis qualitatively, using the vertical velocity at the midplane $u_z(z = 0)$ as a proxy for VSI activity similar to Figs. 19 and 31, where we highlighted that axisymmetric, rapidly-alternating vertical motion as a function of radius is a typical signature of the VSI. We break down our results into four paragraphs, one for each model, and comment on their behavior and notable patterns individually in Sects. 8.2.1 through 8.2.4. These four paragraphs are sorted by the planet's mass in M_th, in ascending order.

8.2.1 Linear regime: $h = 0.1$, $M_\mathrm{p} = 100\,\mathrm{M}_\oplus = 0.3\,M_\mathrm{th}$

In the linear regime, the planet's impact on the overall disk structure is minuscule. The lack of any notable features such as gaps or vortices, coupled with the strong VSI turbulent stress for $h = 0.1$, effectively results in the VSI being largely uninhibited by the planet's presence. The top panels of Fig. 40 show the disk surface density structure at three timestamps during the planet's growth and interaction with the disk. Excluding the planet's weak spirals and the vortex that develops due to the parasitic instability discussed in Sect. 7.2.2, the vertical velocity component of the gas remains practically axisymmetric. The bottom panels on the same figure show that VSI modes are partially perturbed near R_p and otherwise remain axisymmetric, albeit slightly weaker due to the presence of the planet's spiral arms.

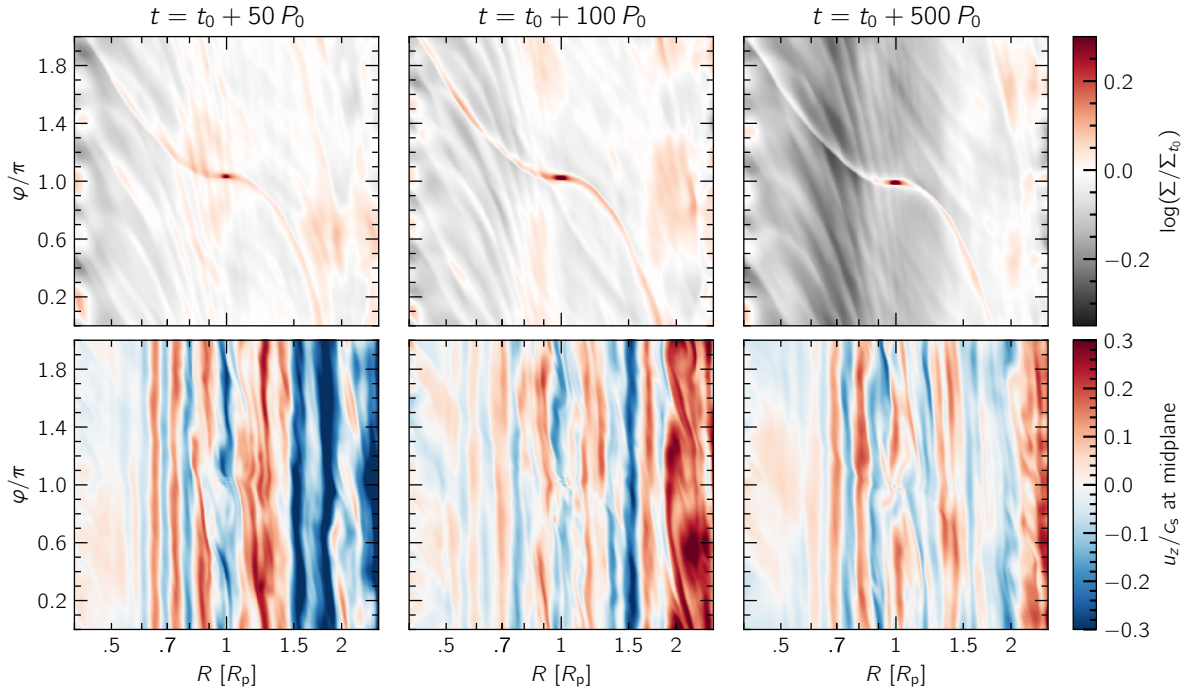

Figure 40. Time evolution of the disk midplane in the linear planet–disk interaction regime, with $h = 0.1$ and $M_\mathrm{p} = 100\,\mathrm{M}_\oplus = 0.3\,M_\mathrm{th}$. The three pairs of panels, from left to right, show the disk during the planet's growth phase, at the end of the growth phase, and in a quasi-equilibrium state. Top: surface density normalized to its profile at t_0, when the planet is introduced. Bottom: the vertical component of gas velocity at the disk midplane, normalized to the local sound speed. VSI signatures appear as vertical red and blue stripes.

Fig. 41 compares the azimuthally averaged vertical profile of u_z for VSI-active disks in a saturated state between our fiducial $r\theta$-2D model, an $r\theta\varphi$-3D model without a planet, and the model where the planet is embedded. The two 3D models—with and without a planet—agree very well qualitatively, with the model including a planet showing slightly weaker vertical activity.

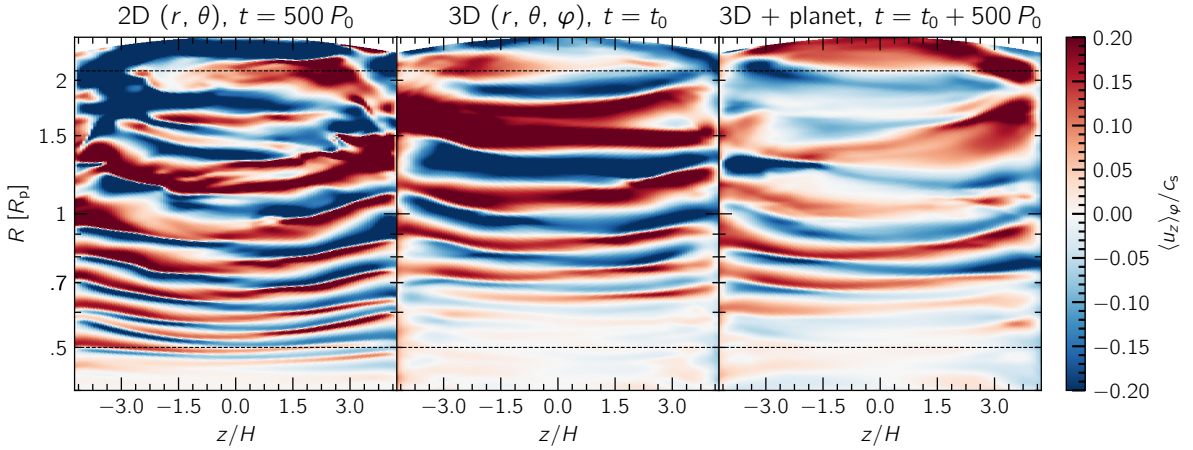

Figure 41. Comparison of a saturated state of the VSI for three different models with identical physical parameters ($h = 0.1$), using u_z as a proxy for VSI activity. Left: an axisymmetric $r\theta$-2D model. Middle: a disk-only $r\theta\varphi$-3D simulation. Right: a 3D model with planet–disk interaction in the linear regime ($M_p = 100\,M_\oplus = 0.3 M_{th}$). The two right panels show qualitatively similar results, due to the planet's inability to develop strong radial or azimuthal features in the disk.

8.2.2 Moderately nonlinear regime: $h = 0.1$, $M_p = 1\,M_J = 1\,M_{th}$

With a planet massive enough to equal the disk's thermal mass, the development of nonlinear effects such as gap opening is to be expected. In this model, the position and shape of spiral arms launched by the Jupiter-sized planet agree very well with the linear regime estimates in Sect. 2.6. However, the planet carves a wide gap around its orbit and, as a result, a vortex forms at the gap's outer edge. We note that the weaker, secondary bump that seemingly forms in the inner disk exists in part due to the outflow condition through our inner radial boundary, as the secondary gap should be centered at $R = 0.4\,R_p$ for this combination of parameters (Zhang et al., 2018) and is therefore subject to our boundary conditions. Our observations are consolidated in the top panels of Fig. 42.

The massive planet has a considerable impact on VSI activity in this regime. As can be seen in the lower panels of Fig. 42, the VSI is weakened significantly in most regions of the disk, for different reasons:

- In the inner disk, the formation of a secondary density bump seems to inhibit VSI activity to the point that the latter vanishes as the secondary gap continues deepening.
- Around the planet's orbit, the horseshoe-like motion of gas disturbs the φ-component of the gas velocity, ultimately destabilizing and destroying VSI modes.
- In the outer disk, the presence of a massive vortex has a similar effect. the combination of a radial density gradient around the outer pressure bump at $R = 1.5\,R_p$ and the vortex's strongly non-axisymmetric perturbations of u_R and u_φ result in a VSI-inactive disk exterior, up to $R \approx 1.8\,R_p$.

The right panel of Fig 43 shows the azimuthally averaged vertical velocity of the disk in a saturated, quasi-equilibrium state after the planet has completed 500 orbits. As shown on that panel, VSI activity is significantly damped near the inner density bump (purple band), the planet's gap region (gray band) and the outer vortex (green band). Outside those regions, which span the majority of the radial extent of the disk, VSI modes continue to exist in a weaker state similar to the linear regime.

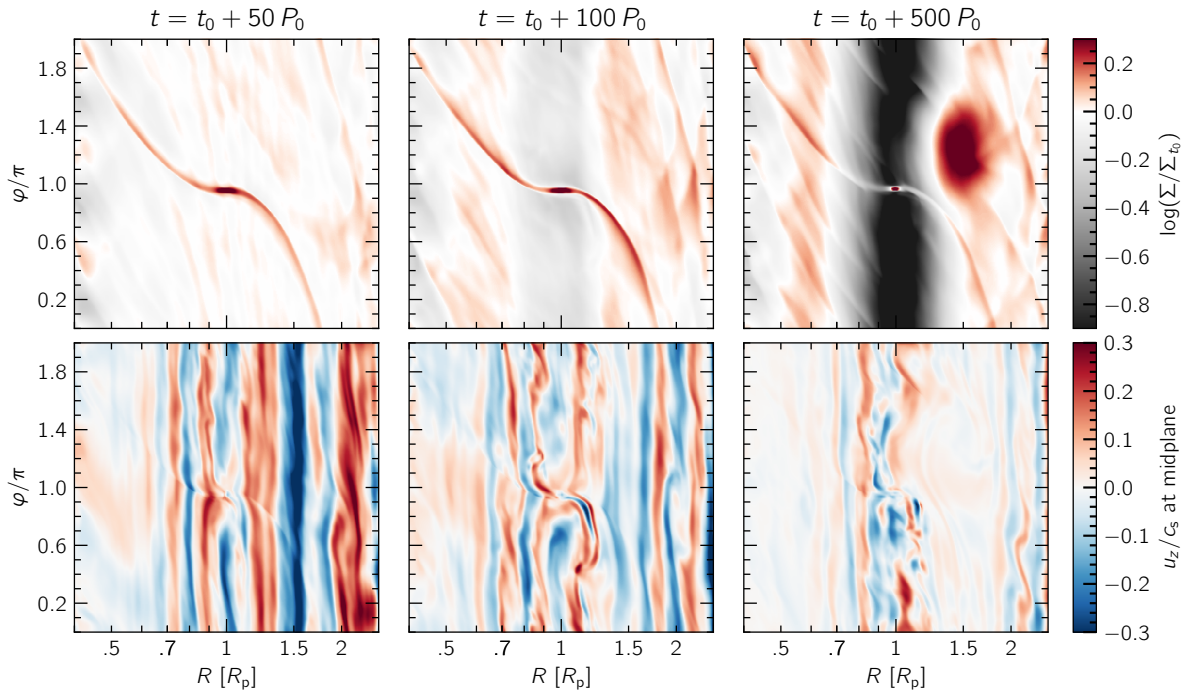

Figure 42. Time evolution of Σ and u_z at the disk midplane similar to Fig. 40, but for $M_\mathrm{p} = 1\,\mathrm{M_J} = 1\,M_\mathrm{th}$. In this regime, nonlinear features such as gaps and vortices develop. As a result, VSI activity is significantly weaker than for the linear regime.

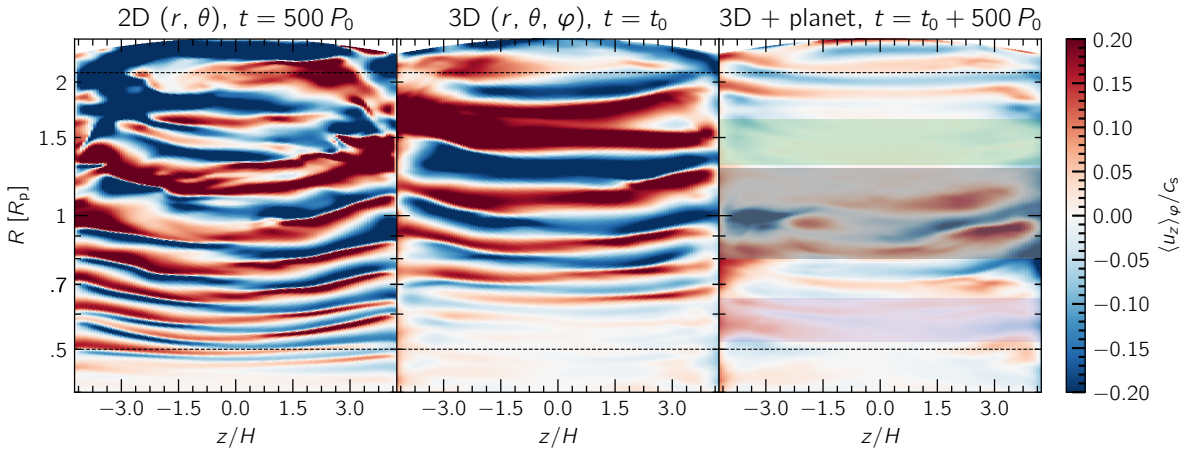

Figure 43. Comparison of the vertical component of the gas velocity in different saturated states similar to Fig. 41, but for $h = 0.01$ and $M_\mathrm{p} = 1\,\mathrm{M_J} = 1\,M_\mathrm{th}$. VSI modes are significantly weaker near the planet's gap region (gray band) and the two bumps in the inner and outer disk (purple and green band, respectively). VSI activity in the outer region is especially damped due to the massive vortex forming on the planet's gap edge.

8.2.3 Nonlinear regime: $h = 0.05$, $M_\mathrm{p} = 100\,\mathrm{M}_\oplus = 2.4\,M_\mathrm{th}$

In this model, which functions as an extension of our fiducial $r\theta$-2D model presented in Sect. 7.1 and analyzed in 3D in Sect. 7.3.3, the mass of the Saturn-sized planet with $M_\mathrm{p} = 100\,\mathrm{M}_\oplus$ amounts to $2.4\,M_\mathrm{th}$, placing the planet well within the nonlinear, high-mass regime. As a result, the VSI is affected in a similar manner to the previous model in Sect. 8.2.2 in the sense that the formation of pressure bumps and large-scale vortices inhibits its activity.

However, the lower aspect ratio of 0.05 instead of 0.1 affects not only the thermal mass, but disk hydrodynamics as well. The primary and secondary gaps carved by the planet are narrower, resulting in radially smaller VSI-inactive zones, but at the same time they are deeper, creating steeper density

gradients and allowing vortices to form even in the inner disk, which can ultimately suppress the VSI locally. In addition, the lower aspect ratio implies a weaker vertical shear rate, but at the same time individual VSI modes are radially thinner (see Fig. 27). This could potentially allow them to develop in-between VSI-inactive zones, where the radially wider modes for $h = 0.1$ would get partially eroded and therefore destroyed.

Fig. 44 shows three snapshots similar to previous models, with an interesting development of events: at the early stage of planet–disk interaction (left panel), the initial launching of planetary spiral arms seems to affect the inner disk more strongly, destabilizing VSI modes momentarily. Since the growth rate of the instability—which is proportional to Ω_K—is much faster in the inner disk, VSI activity resumes over the next 50 orbits (middle panel). However, at the same time, vortex formation around the planet's gap opening region interferes with its development. As can be seen on the same panel, the combination of planet- and vortex-generated spiral arms traveling through the inner disk disturbs the otherwise-axisymmetric nature of the VSI, even though the latter can grow fast enough to survive. In the later stages of planet–disk interaction, once a deep gap has opened and a massive vortex has formed in the outer disk (right panel), VSI turbulence in the outer disk has weakened significantly. On the other hand, vortex activity in the inner disk has subsided and given way to a radially narrow ring at $R = 0.7$–$0.8\,R_p$, allowing the VSI to saturate at levels similar to the 2D setup at $R < 0.7\,R_p$.

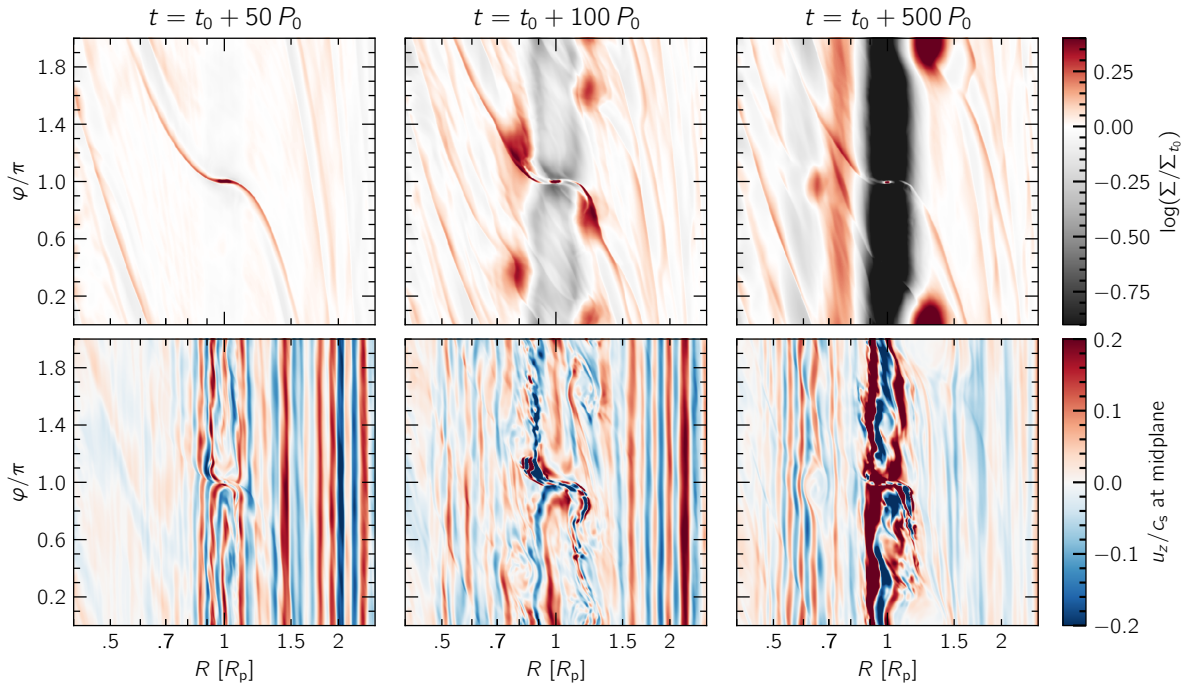

Figure 44. Time evolution of Σ and u_z similar to Fig. 40, with $h = 0.05$ and $M_p = 2.4\,M_{th}$. VSI activity is directly linked to the presence of rings, gaps, and vortices.

Similar to previous sections, the azimuthally averaged profile of $u_z(R, z)$ in a quasi-equilibrium state is shown in Fig. 45. In this saturated state, VSI modes are mainly active in the region $R < 0.7\,R_p$, where they are not disturbed by planet or vortex activity.

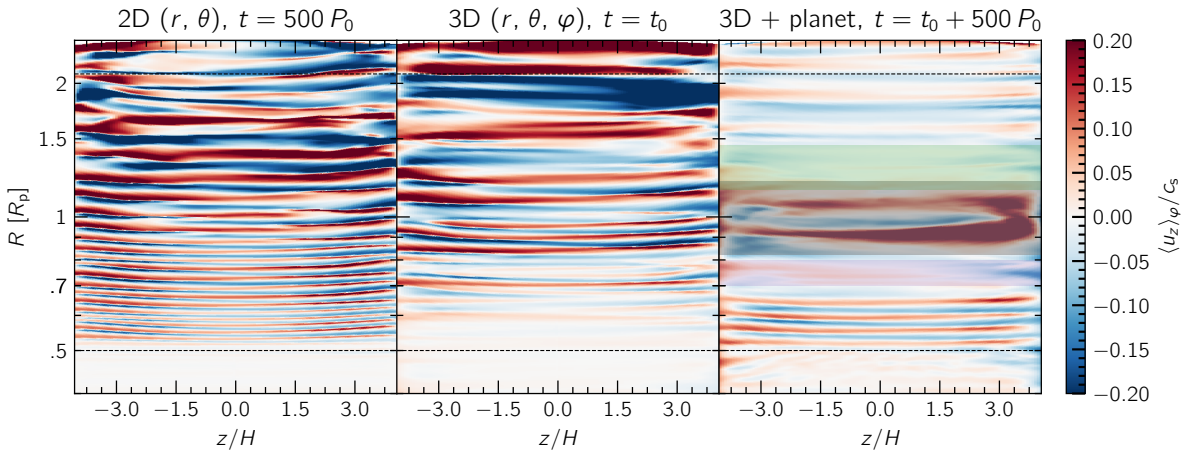

Figure 45. Comparison of the azimuthally averaged $u_z(R, z)$ similar to Fig. 41, for $h = 0.05$ and $M_\mathrm{p} = 100\,\mathrm{M}_\oplus = 2.4\,M_\mathrm{th}$. The secondary pressure bump in the inner disk (purple band) is narrower compared to Fig. 41, and the parasitic instability in the inner boundary has been quenched due to the massive planet's presence, allowing the VSI to develop again for $R < 0.7\,R_\mathrm{p}$.

8.2.4 Strongly nonlinear regime: $h = 0.05$, $M_\mathrm{p} = 1\,\mathrm{M_J} = 8\,M_\mathrm{th}$

For our last model, we increase the planet's mass to $1\,\mathrm{M_J}$ compared to our fiducial run with $M_\mathrm{p} = 100\,\mathrm{M}_\oplus$. The planet now has a mass of 8 times the disk thermal mass, and the disk is dominated by nonlinear effects. This can be seen in Fig. 46 in the form of deep gaps and strong vortices that form earlier than in previous models, and violent perturbations of u_z especially within the planet's corotating region. The picture is similar to the fiducial case, with the difference that the Jupiter-sized planet influences a larger radial extent of the disk, interfering with VSI activity on a larger scale. This becomes clear when looking at the vertical profile of u_z in Fig. 47, with the majority of VSI modes having disappeared in the radial range $0.7 < R/R_\mathrm{p} < 1.5$.

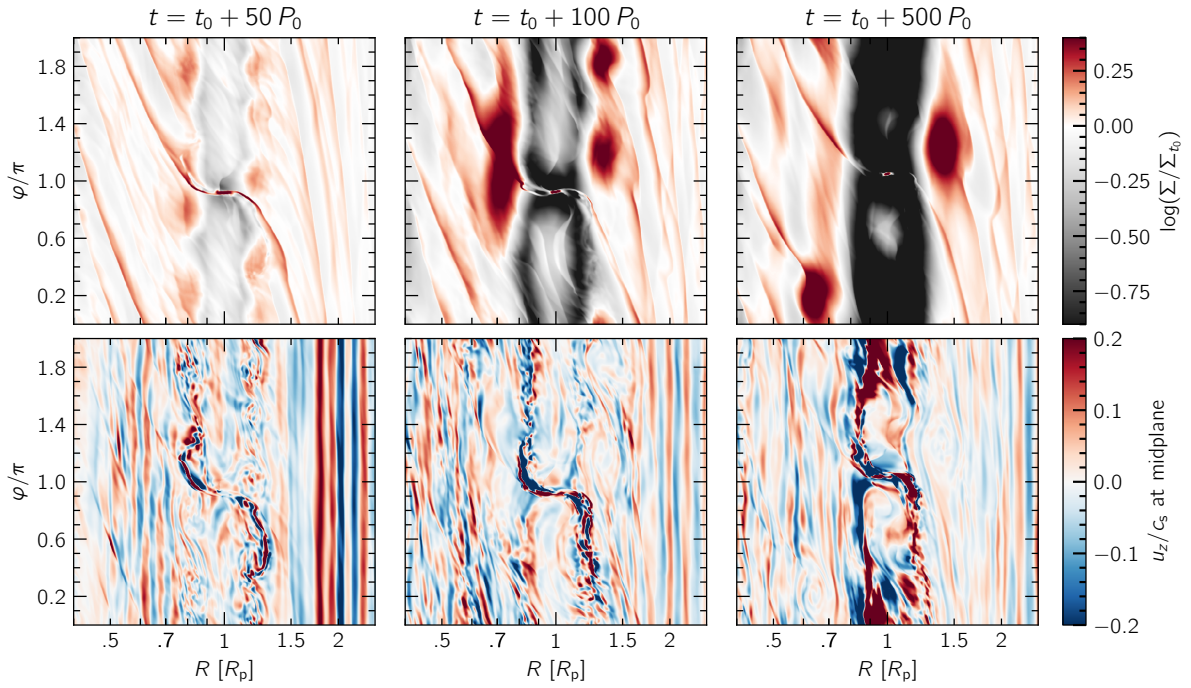

Figure 46. Time evolution of Σ and u_z at the disk midplane similar to Fig. 44, but for $M_\mathrm{p} = 1\,\mathrm{M_J} = 8\,M_\mathrm{th}$. In this strongly nonlinear regime, non-axisymmetric planet-generated features dominate the flow, disrupting VSI activity globally.

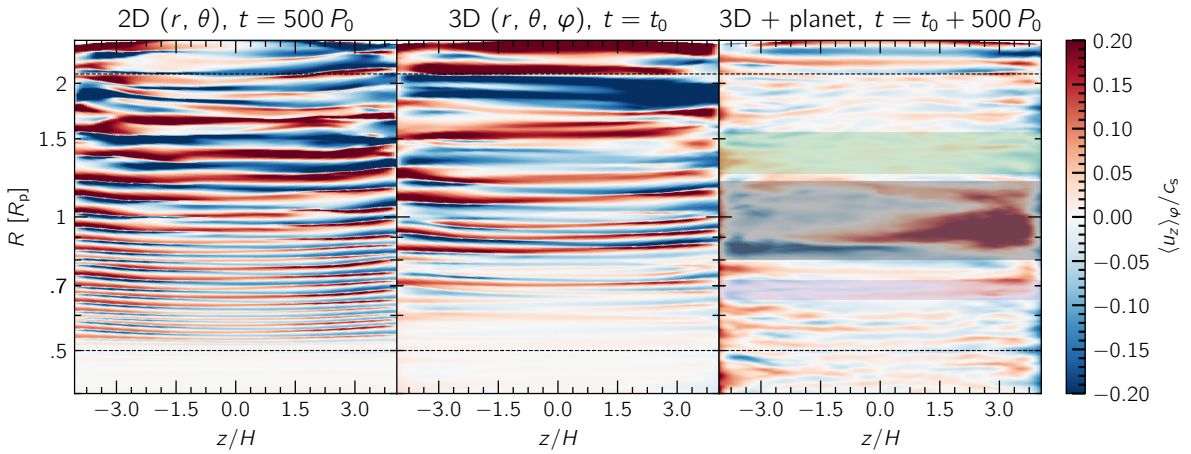

Figure 47. Saturated u_z profiles between different disk models similar to Fig. 45, for $M_\mathrm{p} = 1\,\mathrm{M_J} = 8\,M_\mathrm{th}$ instead. The planet's presence completely shuts down the VSI in the region $0.7 < R/R_\mathrm{p} < 1.5$, and weakens it significantly outside of that zone.

8.3 VSI stress levels

Spiral arms launched by a planet can be a very efficient source of viscosity. Building on Eq. (7.5) by Larson (1990), Goodman and Rafikov (2001) showed that the effective α_acc by planet-generated wakes that are launched by a planet and are damped over a distance smaller than R_p can be estimated as

$$\alpha_\mathrm{eff} \approx 3.2\left(\frac{\gamma+1}{7/5+1}\right)^{-2/5}\frac{R^3 c_\mathrm{s}^2 Z_\mathrm{p}\Sigma}{\mathrm{G}M_\star M_\mathrm{p}}\left(\frac{3\Omega_\mathrm{K}\mathrm{G}M_\mathrm{p}}{2c_\mathrm{s}^3}\right)^{8/5} \approx 1.3\times10^{-3}\left(\frac{h}{0.05}\right)^{1.8}\left(\frac{M_\mathrm{p}}{100\,\mathrm{M}_\oplus}\right)^{8/5}, \tag{8.2}$$

for our choice of parameters. As a result, in order to isolate the effects of the VSI we follow the same strategy as in Sect. 7.3.3, by using the azimuthally averaged velocity field when computing the Reynolds stress tensor $\overline{\mathbf{T}}$. Nevertheless, for comparison, we also compute and plot the total stress, including the planet's contribution. We choose to smooth over 300 orbits using 300 snapshots between $t = t_0 + 200\,P_0$ and $t_0 + 500\,P_0$, to ensure that the planet has fully grown and the disk has reached a quasi-equilibrium state by the time we begin searching for VSI activity.

Our results are consolidated in Fig. 48 for our grid of 4 models. On each panel we overlay shaded areas that correspond to vortex activity (purple and green bands) or gap regions (gray bands), as discussed in Sect. 8.2. The panels are ordered such that the planet mass in units of the disk thermal mass increases along a "Z" shape from top left to bottom right, and the vertical shear rate decreases from top to bottom. Solid lines account for stress due to VSI activity alone, while dashed lines include all turbulent stress that can drive accretion in the disk, which is dominated by spirals throughout the domain. A thick black line on the top right of each panel denotes the radial smoothing length during our averaging process, as a proxy for a resolution limit on visible radial structures due to convolution.

Our results agree with the qualitative analysis in Sect. 8.2 and quantify the impact of the planet's presence on VSI activity. In the linear regime (top left panel of Fig. 48), the planet's spiral arms perturb VSI modes weakly, reducing the effective accretion parameter from $\alpha_\mathrm{acc} \approx 2.5\times10^{-4}$ to 1.5×10^{-4}. Of course, if we were to take into account the effective viscosity due to the planet's wakes following Eq. (8.2), we would find a total $\alpha_\mathrm{eff} \approx 8\times10^{-4}$ that is constant in the inner disk but drops in the outer disk, as the planet's trailing spiral arms damp over distances wider than R_p (Goodman and Rafikov, 2001). For the rest of our analysis, we will focus on VSI stress (solid lines) only.

As the planet's mass increases compared to the disk thermal mass, we find that VSI activity is restricted to very narrow regions that are not directly affected by planet-generated features such as vortices or gap opening. We find an example of this behavior for $M_\mathrm{p} = 1\,M_\mathrm{th}$ on the top right panel of Fig. 48, where we observe visibly weaker VSI stress (6×10^{-5} down from 2.5×10^{-4}) in the few white regions of the panel, and significantly lower within the colored bands. We note that, since the planet opens a

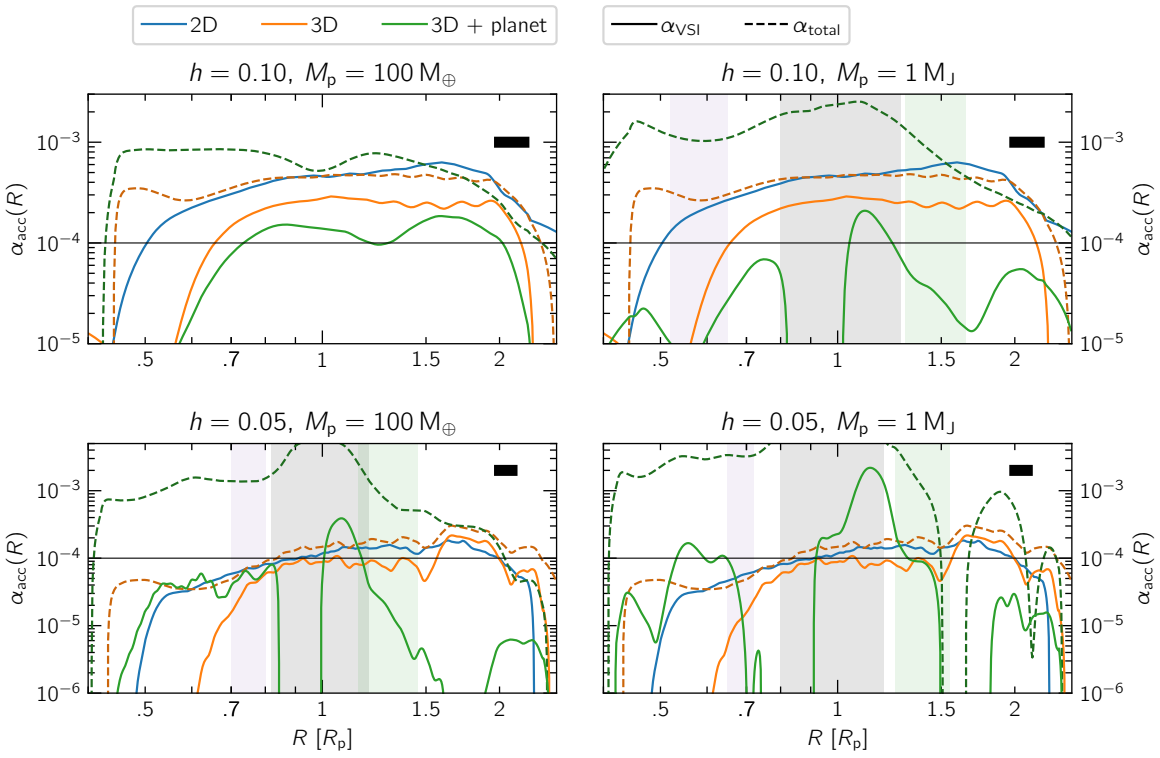

Figure 48. Comparison of the turbulent accretion parameter $\alpha_{\rm acc}(R)$ for four different 3D models with varying aspect ratios h and planet masses $M_{\rm p}$. The ratio $M_{\rm p}/M_{\rm th}$ increases along a "Z" shape from the top left to the bottom right, with the first panel falling in the linear regime ($M_{\rm p} = 0.3 M_{\rm th}$) and the rest in the nonlinear regime ($M_{\rm p} = \{1, 2.4, 8\}\, M_{\rm th}$ respectively). Dashed lines indicate the total turbulent stress following the method in Sect. 3.4, while solid lines attempt to isolate the contribution of the VSI by computing $T_{R\varphi}$ on azimuthally averaged datasets. A solid black line at $\alpha = 10^{-4}$ helps to guide the eye. Thick black lines denote our radial smoothing length of $\Delta R = H(R)$.

gap, the gray-shaded region is strongly affected by the gap profile while also being extremely sensitive to numerical noise, and should therefore not be taken into account on all panels.

The lower panels of Fig. 48 correspond to $h = 0.05$ and build on our fiducial model. As shown in Fig. 27 and discussed in Sect. 8.2.3, while the lower aspect ratio results in a weaker vertical shear rate, the individual VSI modes generated are narrower in the radial direction. This allows the VSI to develop in the narrow, vortex-free region between $0.5 < R/R_{\rm p} < 0.8$, which should contain approximately 3 times more VSI modes for $h = 0.05$ than for $h = 0.1$ according to Fig. 27. Interestingly enough, the higher vertical shear rate in the inner disk, combined with the fact that the planet's presence eliminates the parasitic instability in the inner boundary, allow the VSI to develop in the zone $0.5 < R/R_{\rm p} < 0.7$, which was not possible for the 3D model without a planet (see also Fig. 31). On the other hand, the combination of radially wider VSI modes and a weaker shear rate in the outer disk result in the instability being perturbed and ultimately weakened significantly by the planet, with $\alpha_{\rm acc}$ being reduced by an order of magnitude or more for $R > 1.5\, R_{\rm p}$.

In Sect. 7.3.3 we discussed how the VSI is the primary driver of vertical motion and concluded that spiral arms on the $R\varphi$ plane, while conducive to accretion, perturb the VSI modes in a way that results in weaker turbulent vertical mixing. The picture is different in the presence of an embedded planet, which can drive significant vertical motion via its spiral waves and especially so in the inner disk (Zhu et al., 2015). In this scenario, vertical motion is primarily excited by the VSI in regions where it is active, and by the planet in the other regions. In Fig. 49 we extend the comparison of $\alpha_{\rm mix}(z)$ that was shown in Fig. 34, using the model on the bottom left panel of Fig. 48 and plotting $\alpha_{\rm mix}(z)$ at three representative radii. We find that in the outer disk VSI activity dominates the vertical flow of gas, whereas the inner disk shows strong vertical motion both due to the planet's wakes and the VSI, with the planet contributing more as we approach the inner radial boundary.

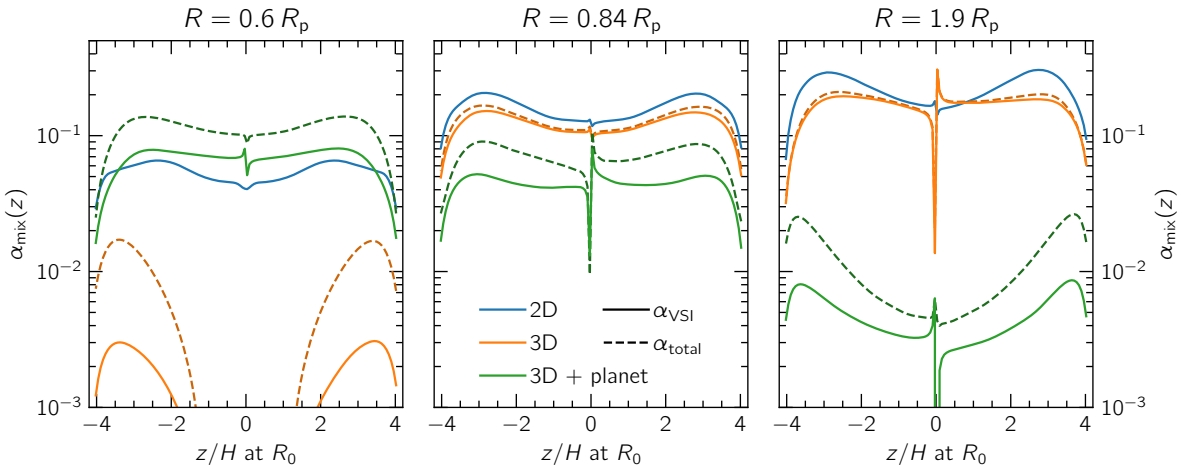

Figure 49. Vertical profiles of α_{mix} at three reference radii similar to Fig. 34, but also including the model with an embedded planet in green. The blue and orange curves refer to the $r\theta$-2D and $r\theta\varphi$-3D models without a planet, respectively. Dashed lines denote the total Reynolds stress $T_{z\varphi}$ including all sources in the disk (VSI, vortices, spiral arms), while the solid lines isolate the contribution of the VSI. Right: in the inner disk, VSI activity recovers in the presence of the planet as the latter eliminates the parasitic instability at the inner rim of the domain. The planet's spiral arms excite additional vertical motion, which adds to the total vertical mixing parameter. Middle: VSI turbulence drops by a factor of 2 near the planet's gap edge, but the instability remains active nevertheless. Right: the VSI has been quenched in the outer disk as the result of the planet's spirals, the presence of a massive vortex at the outer gap edge, and the weak vertical shear rate in the outer disk. The vertical motion excited by the planet's spiral arms is also considerably weaker in the inner disk, and as a result the overall α_{mix} is quite small.

8.4 Planet migration in VSI-active disks

Having investigated the effect of the planet on the development and saturation levels of the VSI, we now allow the disk to exert a torque on the planet and let the latter migrate over 300 orbits measured at R_0. We note that the computational domain is quite restricted, implying that the planet will be subject to boundary effects once it approaches either end of the zone $0.5 < R/R_p < 2.1$. For that reason, and given that the planet can interact with VSI turbulence over a timeframe of 50–100 orbits as shown on Fig. 44, we believe that modeling the migrating planet for 300 orbits is sufficient to offer some insight into the extent to which VSI activity can affect planet migration.

We use ρ_{mid} as an input quantity rather than Σ, such that the latter increases by a factor of 2 for our models with $h = 0.1$. While this might result in a less fair comparison of migration tracks between equally-sized planets in disks with different h, it provides a more reasonable surface density estimate for regions of importance to planet formation and is therefore a good choice in terms of interpreting our results from an observational point of view. Namely, in the outer disk where h would be of the order of 0.1 (Flock et al., 2020), our choice of parameters results in a surface density of 4.61 $(h/0.1)\,(M_\star/M_\odot)$ and 1.15 $(h/0.1)\,(M_\star/M_\odot)$ at 50 and 100 au, respectively. Table III lists several ALMA targets and compares the estimated surface density at several radii where the VSI could be active against our estimated Σ, rescaled appropriately to those distances.

Our results are consolidated in Fig. 50 for all models, showing the time evolution of the semimajor axis a, the eccentricity e and the inclination i of planets from the moment when we allow the planets to migrate ($t_0^{mig} = t_0 + 500\,P_0 = 1700\,P_0$). The top panel of Fig. 50 shows that the 100-Earth-mass planet migrates inwards in the model with $h = 0.1$ and all other models show outward migration, with quite different migration timescales between models. The migration tracks of planets in all models are shown separately in Fig. 51 to better distinguish their individual behavior.

System	$M_\star$ [$M_\odot$]	R_0 [au]	h_0	Σ_0^{Z+} [g/cm^2]	suggested model	Σ_0^{est} [g/cm^2]
HD 163296	2.04	48	0.056	10–30	$h = 0.05$; $M_p = 1\,M_J$	5.21
HD 163296	2.04	48	0.056	10–30	$h = 0.1$; $M_p = 1\,M_J$	10.42
HD 163296	2.04	86	0.066	3–30	$h = 0.05$; $M_p = 1\,M_J$	2.91
HD 163296	2.04	86	0.066	3–30	$h = 0.1$; $M_p = 1\,M_J$	5.82
AS 209	0.83	99	0.082	3–30	$h = 0.1$; $M_p = 1\,M_J$	0.97
HD 143006	2.04	51	0.051	3–30	$h = 0.05$; $M_p = 100\,M_\oplus$	3.94
Elias 24	0.78	57	0.090	10–100	$h = 0.1$	2.75
GW Lup	0.46	74	0.086	3–10	$h = 0.1$; $M_p = 100\,M_\oplus$	0.96

Table III. Comparison of the surface density of several target sources of the DSHARP study (Andrews et al., 2018) at different reference radii R_0 against our estimated Σ_0^{est} for our choice of parameters after appropriate rescaling. The aspect ratio h_0 is calculated assuming a simple irradiation model using Eq. (2.18), and Σ_0^{Z+} denotes the estimates of Zhang et al. (2018) (Table 3 therein).

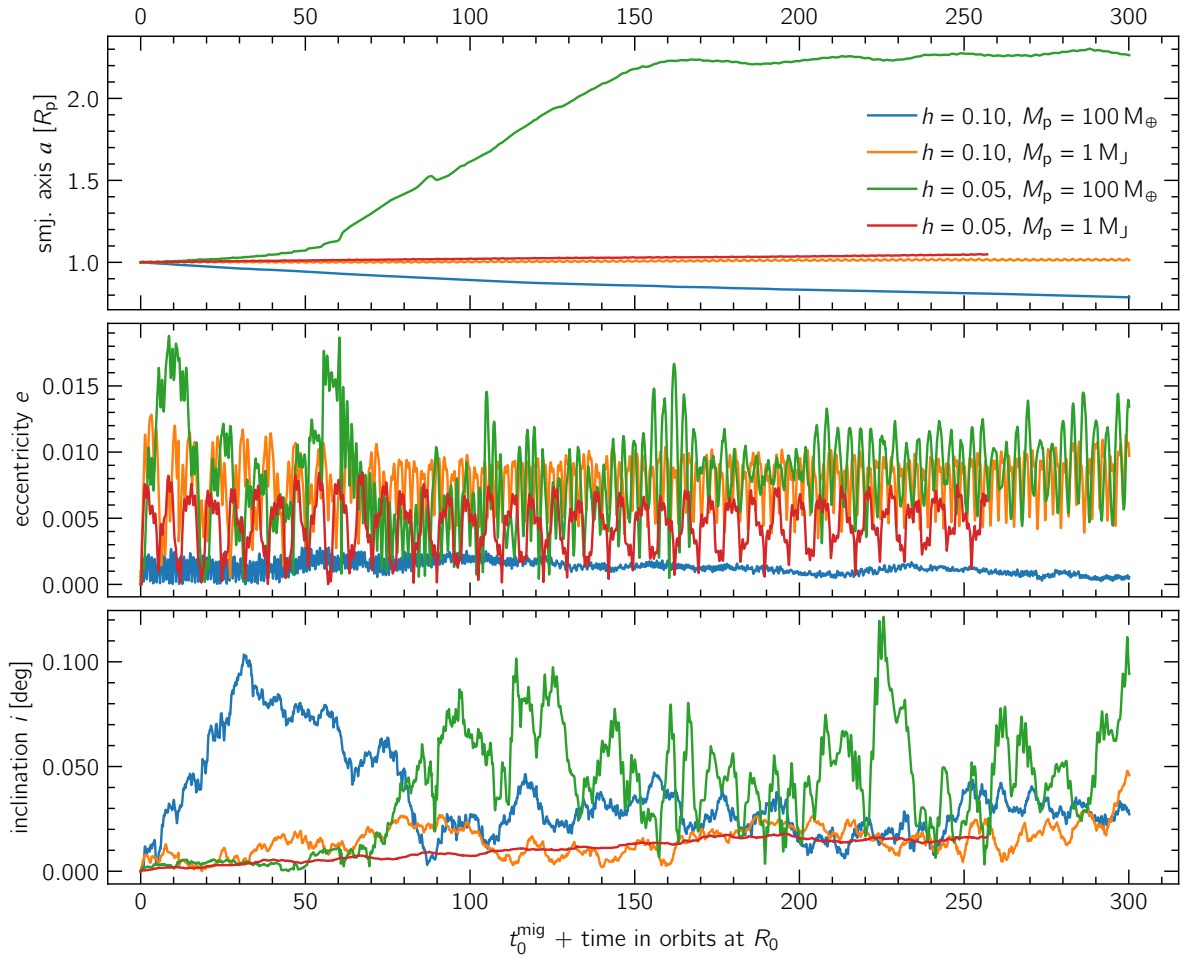

Figure 50. Time evolution of the orbital elements of migrating planets in our simulations. Top: the semimajor axis a increases for models with $h = 0.05$, indicating outward migration, while the planets migrate inward for $h = 0.1$. Middle: the eccentricity of all planets remains relatively low, with oscillations excited by planet–vortex interaction. The planet in the linear regime (blue curve) shows little to no eccentricity excitation. Bottom: inclination of the orbits of all planets with respect to the disk midplane. It is likely that these oscillations, while vanishingly small in amplitude, are caused by a combination of vortex and VSI activity. Here, $t_0^{mig} = t_0 + 500\,P_0$ denotes the moment when planets begin to migrate.

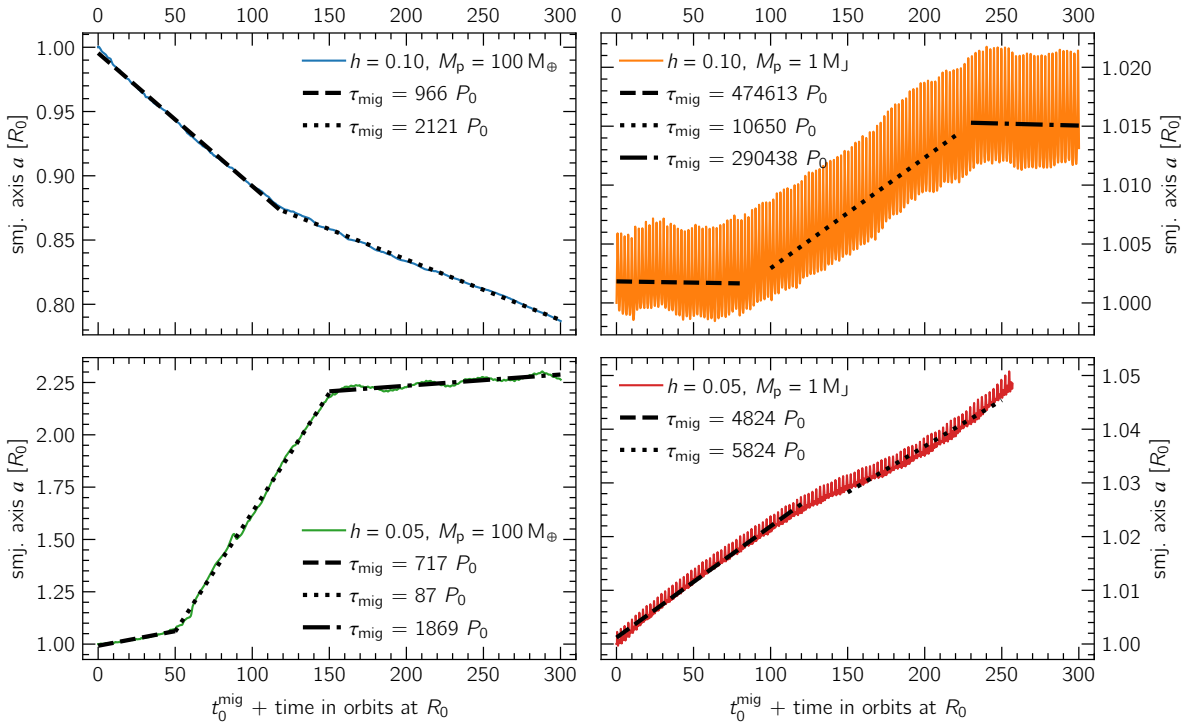

Figure 51. Time evolution of the semimajor axis of migrating planets for our four models. The dashed, dotted, and dashed-dotted black lines mark linear fits to each leg of the planets' migration tracks, and the slope of each fit is translated into a migration timescale $\tau_{\rm mig} = a/\dot{a}$ and added to the legend.

The low-mass planet in the model with $h = 0.1$ and $M_{\rm p} = 100\,{\rm M}_\oplus$ (blue curve) does not open a gap, and follows classic type-I migration (Kley and Nelson, 2012) until it reaches $R = 0.87\,R_0$, where it slows down but continues migrating inwards. The reason why it slows down is that an artificial "gap" is formed near $R = 0.7\,R_0$, due to the combination of two effects. For one, the radial gradient of $\alpha_{\rm acc}$ imposed by the VSI-inactive inner disk ($R < 0.65\,R_0$, see solid green curve on top left panel of Fig 48) causes an artificial accumulation of material near the inner boundary as the accretion rate decreases close to it. Our wave damping boundary condition then enhances this effect by slowly replenishing the gas mass near $R \approx 0.5\,R_0$. As a result, the planet reaches a high-density zone which strengthens the positive torque contribution of the inner disk and therefore slows the planet's migration speed by a factor of 2, as can be seen on the top left panel of Fig. 51.

The top panels of Fig. 52 show the surface density distribution in the disk for the model with $h = 0.1$ and $M_{\rm p} = 100\,{\rm M}_\oplus$ at three timestamps: soon after the planet begins migrating (left panel), while it travels down the edge of the artificial gap (middle), and after it reaches the center (trough) of the gap. The bottom panels, using u_z as a proxy for VSI activity, show that the VSI remains active during the entire process due to the lack of strong perturbations in the form of vortices in the disk.

In the remaining three cases, for planets equal to or exceeding the disk thermal mass, migration is substantially more unpredictable due to the presence of massive vortices at the outer gap edge at the time when migration starts and the nearly-inviscid disk background (Lega et al., 2021). Here, due to the gap carved by the planet, migration typically falls in the type-II regime (Kley and Nelson, 2012). In all three cases, the planets initially migrate outwards with migration timescales $\tau_{\rm mig} = a/\dot{a} \sim 0.5$–$1 \times 10^4\,P_0$. The Jupiter-sized planets then continue drifting outwards slowly, always remaining within their gap region, with the planet in the model with $h = 0.1$ stalling its migration at $t = t_0^{\rm mig} + 220\,P_0$.

Contrary to the type-II migration observed for $M_{\rm p} = 1\,{\rm M}_{\rm J}$, the planet in the model with $h = 0.05$ and $M_{\rm p} = 100\,{\rm M}_\oplus$ transitions to rapid type-III migration at $t \approx t_0^{\rm mig} + 50\,P_0$. This is the result of the planet migrating to its gap edge while interacting with the vortices present in the disk, where it receives a strongly positive torque that triggers rapid outward migration (Pepliński et al., 2008). We speculate

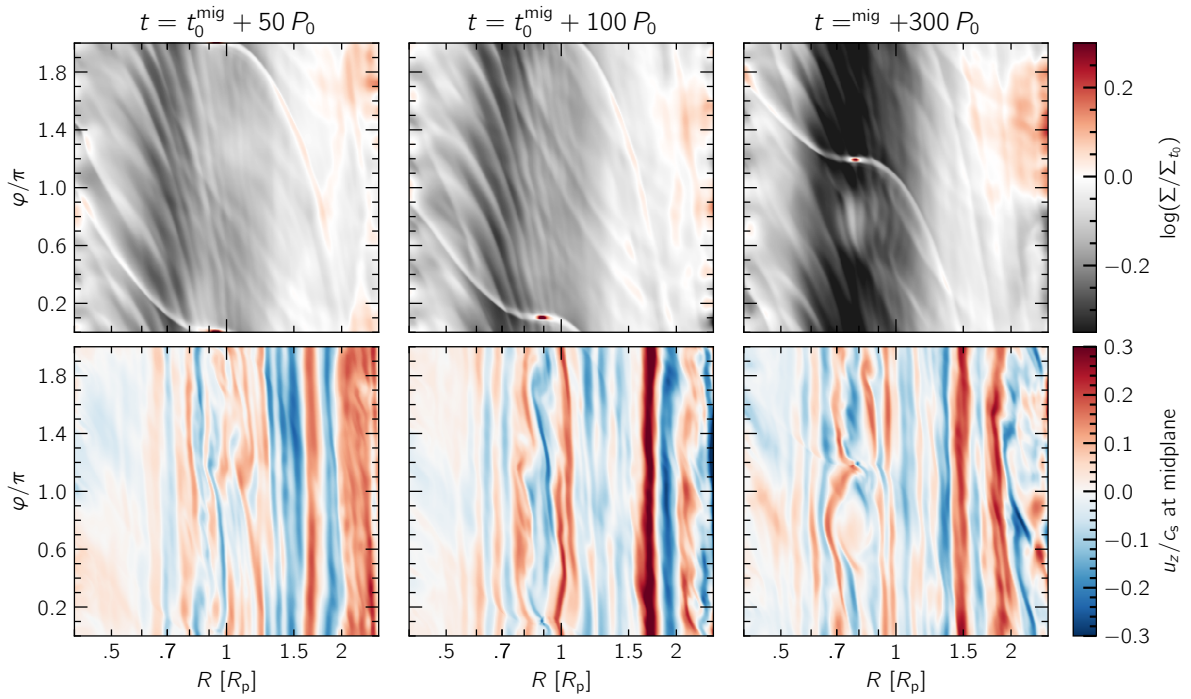

Figure 52. The surface density distribution of the gas and its vertical velocity component at the midplane for the model with $h = 0.1$ and $M_{\rm p} = 100\,{\rm M}_\oplus$ at three snapshots after the planet has started migrating. As we progress from left to right, the planet migrates inwards towards the artificial gap caused by the inactivity of the VSI and our wave damping boundary condition near the inner rim of our domain. Eventually the planet reaches the center of the gap, and its migration speed slows down by a factor of 2.

that the subsequent halt at $R \approx 2.2\,R_0$ is in part artificial, due to our wave-damping boundary conditions at $R > 2.1\,R_0$.

Regarding eccentricity excitations (middle panel of Fig. 50), we find that all models with gap-opening planets ($M_{\rm p} \geq M_{\rm th}$) generate massive vortices at their outer gap edge, causing periodic fluctuations in each planet's eccentricity as the planet interacts gravitationally with the vortex. Conversely, the lack of a gap for the model where $M_{\rm p} = 0.3\,M_{\rm th}$ (blue curve) results in a vortex-free disk, and the planet's orbit remains circular. In all cases, the—axisymmetric by nature—VSI does not affect the planet's eccentricity. We highlight the eccentricity spike at approximately $t_0^{\rm mig} + 60\,P_0$ for the rapidly-outward-migrating planet (green curve), which corresponds to the time when the planet begins interacting with the outer gap edge and type-III migration is triggered.

Finally, the inclinations of all planets remain very close to their initial value of zero, with very slight fluctuations of the order of 0.1° about the midplane (bottom panel of Fig. 50). Once again, it is unlikely that the VSI can drive inclination excitation in models with massive planets, as planet–vortex interaction has significantly more noticeable effects. In the case of the low-mass planet (blue curve), where vortices are not present, it appears that the planet's inclination fluctuations might be slightly stronger due to the VSI during its first 100 orbits of migration, but this effect subsides soon after.

9 Conclusions and discussion

In the broader scope of this project, we carried out two- and three-dimensional hydrodynamics simulations of accretion disks with and without embedded planets. We highlighted the importance of the equation of state and radiative effects in planet–disk interaction in Sects. 4 and 5, and explored the behavior of planet-generated vortices in disks with different physical parameters in Sect. 6. Then, we quantified the effects of turbulence excited by the vertical shear instability (VSI) in protoplanetary disks and its properties as a function of various physical and numerical parameters in Sect. 7. In the context of planet–disk interaction, we measured the impact of the planet's presence on the development of the VSI and vice versa in Sect. 8 using high-resolution 3D models with a simplified but appropriately motivated treatment of radiative effects.

In this section we outline our results on the VSI and planet–disk interaction in VSI-active disks. We highlight the main takeaway messages of this project, and discuss their implications individually.

9.1 Physical properties of the VSI

We explored the behavior of the VSI by monitoring its growth and saturation phase, analyzing its impact on the disk once the latter reaches a quasi-equilibrium state, and measuring its potential to drive turbulent accretion, mixing, and heating. We then investigated the impact of different physical parameters and frameworks on our observations.

Our results are consistent with and enrich previous studies, providing a comprehensive overview of the physical and numerical parameters considered in the literature. Our estimates for the growth phase and turbulent stress levels of the VSI agree both qualitatively and quantitatively with most studies on the subject (e.g., Nelson et al., 2013; Stoll and Kley, 2014; Stoll et al., 2017; Flock et al., 2017; Manger et al., 2020), and our preliminary resolution study ensures the robustness of our results.

VSI-driven turbulence is compatible with observed accretion rates and turbulent diffusion
We showed that VSI-active disks develop an accretion channel with a width of $\pm H$ about the midplane, regardless of the radial temperature profile of the disk. This implies gas accretion within this channel and outward movement outside of it, with a vertically-integrated turbulent stress that corresponds to an accretion parameter $\alpha_{\rm acc} \approx 1.5 \times 10^{-4}$ for a non-flared disk with $h = 0.05$, and an accretion rate $\dot{M} \sim 10^{-9}$–$10^{-7}\,{\rm M}_\odot$/yr depending on disk parameters. While we used a non-flared disk model ($q = -1$) to keep our models scale-free, our results are applicable in the ALMA-observable regions of the disk ($R > 50$ au) where the combination of $h \sim 0.1$ and $q = -0.5$ produces similar levels of $\alpha_{\rm acc} \approx 10^{-4}$ (e.g., Flock et al., 2020).

We found that VSI motion also drives turbulent heating that translates to an effective $\alpha_{\rm heat} \approx 2.5 \times 10^{-4}$, a value very similar to our estimate of $\alpha_{\rm acc}$. This finding is in line with the passive, irradiated disk model of Chiang and Goldreich (1997), ruling out the VSI as a primary source of heating in the regions where it is active ($R \gtrsim 30$ au, Flock et al., 2020).

Based on our estimates of $\alpha_{\rm acc}$ and $\alpha_{\rm heat}$, it is not unlikely that the VSI can drive radial mixing with an effective turbulent diffusion parameter of similar magnitude $\alpha_{\rm turb} \sim 1$–2×10^{-4}, which would be consistent with the inferred lower limit of $\alpha_{\rm turb} \gtrsim 10^{-4}$ through the analysis of turbulent spreading of dust rings in observed systems (Dullemond et al., 2018). In addition, this relatively weak level of turbulence is necessary to model the ring structures of ALMA-observed systems via the planet formation scenario, which require $\alpha \lesssim 10^{-4}$ (e.g., Zhang et al., 2018).

The above statements suggest that the VSI is a competitive candidate in interpreting observed accretion rates and turbulent diffusion in protoplanetary disks.

The equation of state does not affect stress levels in a fully saturated VSI-active disk
While a locally isothermal equation of state creates optimal conditions for the development of the VSI by essentially allowing the vertical shear to ignore the restoring effects of buoyancy, a reasonably short cooling timescale yields the same results. We found that, as long as the cooling timescale β is shorter by at least a factor of 10 than the $\beta_{\rm crit}$ given by Lin and Youdin (2015) (≈ 0.1 for our models), the VSI can function uninhibited and produces turbulent stress comparable to that of a locally isothermal model. In other words, as long as $\beta \ll \beta_{\rm crit}$, the cooling timescale only determines whether the instability will be active, and not the stress it can generate. For $\beta \lesssim \beta_{\rm crit}$ a sigmoid-like correlation between α and β is expected, as highlighted in Manger et al. (2020) and confirmed qualitatively by our results.

It should be noted, however, that in principle β can vary with distance and height. It is therefore crucial to calculate β self-consistently in realistic models by treating radiation transport, if one would like to examine the onset radius of the instability (e.g., Flock et al., 2020). Taking into consideration the decoupling of dust and gas at the surface layers of the disk can also translate to a cooling timescale $\beta(z)$ that increases with z, constraining VSI activity to a few scale heights about the midplane (Pfeil and Klahr, 2021). Nevertheless, for disk parameters chosen such that $\beta \ll \beta_{\rm crit}$ throughout the domain, we found that a fully radiative model behaves no differently from a locally isothermal model in terms of VSI activity.

Concerning the assumption of a vertically isothermal relaxation temperature profile, we showed that as long as we restrict the vertical extent of our domain to within a few pressure scale heights (below the $\tau_{\rm abs} = 1$ surface, or $\pm 4H$), the disk can be approximated as vertically isothermal very well. As a result, a complex model of an irradiated disk was not necessary in our domain of interest.

As we executed most models using a constant $\beta = 10^{-2}$, our results are relevant in the region where the cooling timescale is short enough to allow the VSI to fully develop, or $R \gtrsim 50$ au for typical stellar and disk parameters. This region can be resolved very well using ALMA instruments (Andrews et al., 2018), and is relevant in modeling the radial structure of observed systems via the planet formation scenario (e.g., Zhang et al., 2018; Ziampras et al., 2020b).

VSI-generated turbulence can excite vortices, which might inhibit vertical mixing
While it is primarily an axisymmetric instability that operates on the Rz plane, the perturbations induced on the azimuthal velocity by the VSI can break the axisymmetry of the disk and facilitate the formation of vortices (Stoll and Kley, 2016; Manger and Klahr, 2018). These vortices, while small enough to not affect the activity of the instability on a large scale significantly, generate spiral arms that can slightly increase the turbulent $\alpha_{\rm acc}$. In addition, the interface between the VSI-stable and -unstable regions of the disk—at $R \sim 30$ au for typical disk parameters—effectively creates a steep gradient in $\alpha_{\rm acc}$ and therefore the accretion rate, resulting in a radial surface density gradient that can be subject to the Rossby-wave instability and forming large-scale vortices (Flock et al., 2020). Such non-axisymmetric features can help facilitate mechanisms such as the streaming instability (Youdin and Goodman, 2005), supporting the formation of planetesimals (Schäfer et al., 2020).

It should be noted that the increase in $\alpha_{\rm acc}$ is a result of the damping of these vortex-generated spiral shocks (Larson, 1990), while the subsequent perturbation of VSI modes reduces the contribution of the VSI itself to the total stress in absolute value. As a result, vertical mixing can be weaker in the presence of vortices and spiral arms if predominantly driven by the VSI, as the latter is subject to perturbations by such non-axisymmetric features on the $R\varphi$ plane.

Resolving the VSI appropriately is important, but also sensitive to parasitic instabilities
During our resolution study, we noted the necessity of a resolution of roughly 20 cps to capture the fastest-growing VSI mode. While failing to capture this mode can affect the development of the instability during its growth phase, it does not necessarily imply that the VSI will saturate at different stress levels. Our physically motivated resolution study is consistent with the numerical models of (Manger et al., 2020). We found, however, that increasing the resolution further results in slightly lower stress levels. This has been observed for the MRI as well (Fromang and Papaloizou, 2007), but we are uncertain

whether stress levels converge at higher resolution.

In addition, high-resolution models are subject to a parasitic instability in the inner radial boundary of the domain. This instability develops in the form of two vertically-rotating vortices that are symmetric about the disk midplane and grow radially and vertically outwards, destabilizing and destroying VSI modes in the process. The origin of this instability is likely physical, as it grows faster at a higher resolution and aspect ratio h, but we expect its development to be sensitive to numerical parameters such as the radial boundary condition. A three-dimensional domain, a finite cooling timescale, an outflow boundary condition, and weaker wave damping near the inner radial boundary inhibit the development of the instability and limit it to weaker saturation levels, containing it within the innermost $0.2\,R_0$ of the domain.

9.2 Planet–disk interaction in VSI-active disks

We simulated models with planets that are growing while embedded in VSI-active disks, for planet masses that corresponded from a small fraction up to a significant multiple of the disk thermal mass $M_{\rm th}$. We identified key planet-generated features in the form of spiral arms, gaps and rings, and analyzed qualitatively and then quantitatively the impact of such features on VSI turbulence while letting the planet–disk system reach a quasi-equilibrium state. We then allowed the planet to migrate through the disk and examined its migration track to gain insight into the extent to which VSI activity can affect planet migration.

Low-mass planets can weaken VSI activity to an extent

The continuous perturbations of the velocity field by an embedded planet in the linear regime ($M_{\rm p} = 0.3\,M_{\rm th}$) result in slightly weaker VSI-driven accretion by a factor of 1.5–2 in our model, similar to the effect of vortices and spiral arms on the $R\varphi$ plane. This can result in weaker vertical mixing, as planetary spiral wakes damp VSI activity and therefore $\alpha_{\rm mix}$.

Wakes by massive planets dominate accretion and can stir vertical motion in the inner disk

The effective viscosity by planet-launched spiral wakes far exceeds the capability of the VSI to drive accretion, by approximately an order of magnitude for $h = 0.05$ and a factor of order 3–6 for $h = 0.1$ in our models. This is especially relevant for the leading spirals of a massive planet in the inner disk, which damp over very short distances compared to the planet's semimajor axis (Goodman and Rafikov, 2001) while also exciting significant vertical motion in the gas (Zhu et al., 2015). This combination of phenomena can compete with or coexist alongside the VSI depending on disk parameters. For our models, a lower aspect ratio $h = 0.05$ allowed the radially thinner VSI modes to sustain themselves more easily in the inner disk once vortex activity had subsided. In contrast to that, the VSI was only weakly active in our model with $h = 0.1$, presumably because of the sensitivity of its radially wider modes to planet-generated features as both compete for "space" in the disk.

Planet-generated vortices often weaken VSI activity significantly in the outer disk

In contrast to the inner disk, the vertical shear rate is considerably weaker in the outer disk as it scales with $\Omega_{\rm K} \propto R^{-3/2}$. Combined with the tendency of massive planets to form long-living, large-scale vortices on their outer gap edge (Paardekooper et al., 2010; Lega et al., 2021; Rometsch et al., 2021), this renders the outer disk unsuitable for the development of the VSI, possibly until these vortices have dissipated. Given that vortex lifetime could be of the order of 10^3–10^4 planetary orbits (Rometsch et al., 2021) for our disk parameters, it is unlikely that the VSI can coexist with a massive planet in the 50–100 au range. It is nevertheless a possibility for VSI activity to precede the planet and coexist with it during its growth phase, driving accretion from the outer disk as well as facilitating planetesimal-forming instabilities to support its growth (e.g., the streaming instability, Schäfer et al., 2020).

VSI signatures in disks with low-mass planets could be faintly observable in the future
Barraza-Alfaro et al. (2021) showed that VSI signatures could be detectable for typical disk parameters in the radial range that is observable by ALMA, for reasonably close-by systems and with a favorable viewing angle. At that range, however, the shallower temperature profile with $q = -1/2$ would weaken VSI activity, while the larger aspect ratio of the order of $h \approx 0.1$ would give rise to radially wide VSI modes, which are susceptible to planetary features based on the result of our analysis. Given that massive planets can generate an abundance of such features and especially so in low-viscosity flared disks (Zhang et al., 2018), we believe that it is unlikely to detect the combination of a massive planet and VSI signatures at ALMA ranges. Nevertheless, it remains possible to observe VSI signatures in systems with low-mass planets during their early, "coexistence" stage.

The VSI does not affect the migration track of a planet noticeably
In the nonlinear regime, gap-opening planets generate massive vortices at their outer gap edge. As it has been shown before, the interplay between planet and vortex plays a key role in the planet's migration track (Lega et al., 2021). Compared to the effects of planet–vortex interaction, we found that the VSI has no noticeable effect on planet migration. In the linear regime, however, the low-density zone outside of our high-density inner boundary—due to a radial gradient of $\alpha_{\rm acc}$—resulted in the formation of an artificial partial gap, which slowed the migration speed of the low-mass planet that would otherwise migrate in the type-I regime by factor of 2. We speculate that this effect can slow down or even stall migration at the interface of VSI activity ($R \sim 30$ au) in a realistic disk for low-mass planets, while also promoting the growth of said planets near that zone.

10 Acknowledgments

I acknowledge and am grateful for the support of the High Performance Computing cluster BinAC, which I used to execute the majority of my simulations. Without their computational resources, this project would literally not have been possible.

This project has been an adventure not only for me, but also for the people around me that kept me company, gave me advice, reminded me about lunch time, listened to me ramble about what I enjoy doing, or pointed out the light at the end of the tunnel. Here, I want to take a moment and thank those people.

First and foremost, my omnipresent and omnipotent family, who gave me complete freedom to choose my path in life, continue to support all of my silly decisions and encourage me to follow what I enjoy. Without the support of you guys I would probably not be making a career out of what I'm doing today.

My friends who I know since I remember myself, and those that I met over the internet and spent the last couple of years with during the pandemic. Between hanging out all day and spending all night on call online, they have always been there for me to let loose and to remind me that no number of degrees will help me suck any less in strategy games.

My late thesis supervisor and mentor, Willy Kley, for his (usually gentle) pushes in the right direction, for offering me the opportunity to be part of this amazing department for the last 4 years, and for having me teach some of my favorite subjects as part of the job. I enjoyed teaching as much as working with others, and showing my students what our group can do has been giving me life these last few years. He was the common denominator in everyone's work, growth and enjoyment in the group, and I owe to him most of the skills I learned during my studies that shaped who I am today. His personality, teachings and scientific legacy have left a lasting impression on me, the department and the scientific community.

All of my office colleagues in the CPT group in Tübingen for a great time both in and out of the office. Especially, my office mate Anna Penzlin who fueled my perpetual sugar rush under the guise of snacks, Will Béthune who suggested roughly as many improvements to my code and methods as video game titles, André Oliva who was always present for the impromptu beer on a Friday evening, Thomas Rometsch with whom I probably spent the last semester working day and night on our paper together so we could actually graduate, and my student Prakruti Sudarshan who decided to proofread my thesis for fun and *still* found mistakes. I'm also grateful to the secretary and patron saint of our group, Heike Fricke, who has always been there to help me maneuver around all sorts of bureaucracy during every stage of my studies in Tübingen.

My colleagues in the GEPARD collaboration between Tübingen and Nice, who provided endless amounts of feedback that significantly improved the quality of my work, as well as some of the best profiteroles I've ever had. Special thanks to my co-supervisor and member of this collaboration, Richard Nelson, with whom I hope to produce some great research over the next few years.

Finally, I'd like to thank my high school physics teacher, Prof. Dimitrios Pantos, who implanted in me the idea of "Dr. Ziampras" a good decade ago and left it up to me to make that idea a reality.

References

Alexiades, V., Amiez, G., and Gremaud, P.-A. (1996). Super-time-stepping acceleration of explicit schemes for parabolic problems. *Communications in numerical methods in engineering*, 12(1):31–42.

ALMA Partnership, Brogan, C. L., Pérez, L. M., Hunter, T. R., Dent, W. R. F., Hales, A. S., Hills, R. E., Corder, S., Fomalont, E. B., Vlahakis, C., Asaki, Y., Barkats, D., Hirota, A., Hodge, J. A., Impellizzeri, C. M. V., Kneissl, R., Liuzzo, E., Lucas, R., Marcelino, N., Matsushita, S., Nakanishi, K., Phillips, N., Richards, A. M. S., Toledo, I., Aladro, R., Broguiere, D., Cortes, J. R., Cortes, P. C., Espada, D., Galarza, F., Garcia-Appadoo, D., Guzman-Ramirez, L., Humphreys, E. M., Jung, T., Kameno, S., Laing, R. A., Leon, S., Marconi, G., Mignano, A., Nikolic, B., Nyman, L. A., Radiszcz, M., Remijan, A., Rodón, J. A., Sawada, T., Takahashi, S., Tilanus, R. P. J., Vila Vilaro, B., Watson, L. C., Wiklind, T., Akiyama, E., Chapillon, E., de Gregorio-Monsalvo, I., Di Francesco, J., Gueth, F., Kawamura, A., Lee, C. F., Nguyen Luong, Q., Mangum, J., Pietu, V., Sanhueza, P., Saigo, K., Takakuwa, S., Ubach, C., van Kempen, T., Wootten, A., Castro-Carrizo, A., Francke, H., Gallardo, J., Garcia, J., Gonzalez, S., Hill, T., Kaminski, T., Kurono, Y., Liu, H. Y., Lopez, C., Morales, F., Plarre, K., Schieven, G., Testi, L., Videla, L., Villard, E., Andreani, P., Hibbard, J. E., and Tatematsu, K. (2015). The 2014 ALMA Long Baseline Campaign: First Results from High Angular Resolution Observations toward the HL Tau Region. *ApJ*, 808(1):L3.

Andrews, S. M., Huang, J., Pérez, L. M., Isella, A., Dullemond, C. P., Kurtovic, N. T., Guzmán, V. V., Carpenter, J. M., Wilner, D. J., Zhang, S., Zhu, Z., Birnstiel, T., Bai, X.-N., Benisty, M., Hughes, A. M., Öberg, K. I., and Ricci, L. (2018). The Disk Substructures at High Angular Resolution Project (DSHARP). I. Motivation, Sample, Calibration, and Overview. *ApJ*, 869(2):L41.

Armitage, P. J. (2009). *Astrophysics of Planet Formation*.

Balay, S., Abhyankar, S., Adams, M. F., Brown, J., Brune, P., Buschelman, K., Dalcin, L., Dener, A., Eijkhout, V., Gropp, W. D., Kaushik, D., Knepley, M. G., May, D. A., McInnes, L. C., Mills, R. T., Munson, T., Rupp, K., Sanan, P., Smith, B. F., Zampini, S., Zhang, H., and Zhang, H. (2018). PETSc Web page. http://www.mcs.anl.gov/petsc.

Balbus, S. A. and Hawley, J. F. (1991). A Powerful Local Shear Instability in Weakly Magnetized Disks. I. Linear Analysis. *ApJ*, 376:214.

Balbus, S. A. and Papaloizou, J. C. B. (1999). On the Dynamical Foundations of α Disks. *ApJ*, 521(2):650–658.

Barraza-Alfaro, M., Flock, M., Marino, S., and Pérez, S. (2021). Observability of the vertical shear instability in protoplanetary disk CO kinematics. *A&A*, 653:A113.

Bell, K. R. and Lin, D. N. C. (1994). Using FU Orionis outbursts to constrain self-regulated protostellar disk models. *ApJ*, 427:987–1004.

Birnstiel, T., Dullemond, C. P., Zhu, Z., Andrews, S. M., Bai, X.-N., Wilner, D. J., Carpenter, J. M., Huang, J., Isella, A., Benisty, M., Pérez, L. M., and Zhang, S. (2018). The Disk Substructures at High Angular Resolution Project (DSHARP). V. Interpreting ALMA Maps of Protoplanetary Disks in Terms of a Dust Model. *The Astrophysical Journal*, 869(2):L45.

Casasola, V. (2008). *The small scale physical evolution of molecular gas in nearby galaxies*. PhD thesis, Université Pierre et Marie Curie-Paris VI.

Chiang, E. I. and Goldreich, P. (1997). Spectral Energy Distributions of T Tauri Stars with Passive Circumstellar Disks. *ApJ*, 490(1):368–376.

Crida, A., Morbidelli, A., and Masset, F. (2006). On the width and shape of gaps in protoplanetary disks. *Icarus*, 181(2):587–604.

Cui, C. and Bai, X.-N. (2020). Global Simulations of the Vertical Shear Instability with Nonideal Magnetohydrodynamic Effects. *ApJ*, 891(1):30.

de Val-Borro, M., Edgar, R. G., Artymowicz, P., Ciecielag, P., Cresswell, P., D'Angelo, G., Delgado-Donate, E. J., Dirksen, G., Fromang, S., Gawryszczak, A., Klahr, H., Kley, W., Lyra, W., Masset, F., Mellema, G., Nelson, R. P., Paardekooper, S.-J., Peplinski, A., Pierens, A., Plewa, T., Rice, K., Schäfer, C., and Speith, R. (2006). A comparative study of disc-planet interaction. *MNRAS*, 370:529–558.

Dullemond, C. P., Birnstiel, T., Huang, J., Kurtovic, N. T., Andrews, S. M., Guzmán, V. V., Pérez, L. M., Isella, A., Zhu, Z., Benisty, M., Wilner, D. J., Bai, X.-N., Carpenter, J. M., Zhang, S., and Ricci, L. (2018). The Disk Substructures at High Angular Resolution Project (DSHARP). VI. Dust Trapping in Thin-ringed Protoplanetary Disks. *ApJ*, 869(2):L46.

ESO (2019). Exoplanet detection methods. https://www.esa.int/ESA_Multimedia/Images/2019/12/Exoplanet_detection_methods. Accessed: 2021-10-18.

Flock, M., Nelson, R. P., Turner, N. J., Bertrang, G. H. M., Carrasco-González, C., Henning, T., Lyra, W., and Teague, R. (2017). Radiation Hydrodynamical Turbulence in Protoplanetary Disks: Numerical Models and Observational Constraints. *ApJ*, 850(2):131.

Flock, M., Turner, N. J., Nelson, R. P., Lyra, W., Manger, N., and Klahr, H. (2020). Gas and Dust Dynamics in Starlight-heated Protoplanetary Disks. *ApJ*, 897(2):155.

Fricke, K. (1968). Instabilität stationärer Rotation in Sternen. *ZAp*, 68:317.

Fromang, S. and Papaloizou, J. (2007). MHD simulations of the magnetorotational instability in a shearing box with zero net flux. I. The issue of convergence. *A&A*, 476(3):1113–1122.

Fung, J. and Ono, T. (2021). Cooling-induced Vortex Decay in Keplerian Disks. *ApJ*, 922(1):13.

Gammie, C. F. (1996). Layered Accretion in T Tauri Disks. *ApJ*, 457:355.

Gammie, C. F. (2001). Nonlinear Outcome of Gravitational Instability in Cooling, Gaseous Disks. *ApJ*, 553:174–183.

Goldreich, P. and Schubert, G. (1967). Differential Rotation in Stars. *ApJ*, 150:571.

Goodman, J. and Rafikov, R. R. (2001). Planetary Torques as the Viscosity of Protoplanetary Disks. *ApJ*, 552(2):793–802.

Haffert, S. Y., Bohn, A. J., de Boer, J., Snellen, I. A. G., Brinchmann, J., Girard, J. H., Keller, C. U., and Bacon, R. (2019). Two accreting protoplanets around the young star PDS 70. *Nature Astronomy*, 3:749–754.

Haisch, Karl E., J., Lada, E. A., and Lada, C. J. (2001). Disk Frequencies and Lifetimes in Young Clusters. *ApJ*, 553(2):L153–L156.

Hammer, M., Pinilla, P., Kratter, K. M., and Lin, M.-K. (2019). Observational diagnostics of elongated planet-induced vortices with realistic planet formation time-scales. *MNRAS*, 482:3609–3621.

Harten, A. (1983). High Resolution Schemes for Hyperbolic Conservation Laws. *Journal of Computational Physics*, 49(3):357–393.

Hartmann, L., Calvet, N., Gullbring, E., and D'Alessio, P. (1998). Accretion and the Evolution of T Tauri Disks. *ApJ*, 495(1):385–400.

Hawley, J. F., Gammie, C. F., and Balbus, S. A. (1995). Local Three-dimensional Magnetohydrodynamic Simulations of Accretion Disks. *ApJ*, 440:742.

Huang, J., Andrews, S. M., Dullemond, C. P., Isella, A., Pérez, L. M., Guzmán, V. V., Öberg, K. I., Zhu, Z., Zhang, S., Bai, X.-N., Benisty, M., Birnstiel, T., Carpenter, J. M., Hughes, A. M., Ricci, L., Weaver, E., and Wilner, D. J. (2018a). The Disk Substructures at High Angular Resolution Project (DSHARP). II. Characteristics of Annular Substructures. *ApJ*, 869(2):L42.

Huang, J., Andrews, S. M., Pérez, L. M., Zhu, Z., Dullemond, C. P., Isella, A., Benisty, M., Bai, X.-N., Birnstiel, T., Carpenter, J. M., Guzmán, V. V., Hughes, A. M., Öberg, K. I., Ricci, L., Wilner, D. J., and Zhang, S. (2018b). The Disk Substructures at High Angular Resolution Project (DSHARP). III. Spiral Structures in the Millimeter Continuum of the Elias 27, IM Lup, and WaOph 6 Disks. *ApJ*, 869(2):L43.

Hubeny, I. (1990). Vertical structure of accretion disks - A simplified analytical model. *ApJ*, 351:632–641.

Isella, A., Huang, J., Andrews, S. M., Dullemond, C. P., Birnstiel, T., Zhang, S., Zhu, Z., Guzmán, V. V., Pérez, L. M., Bai, X.-N., Benisty, M., Carpenter, J. M., Ricci, L., and Wilner, D. J. (2018). The Disk Substructures at High Angular Resolution Project (DSHARP). IX. A High-definition Study of the HD 163296 Planet-forming Disk. *ApJ*, 869(2):L49.

Kant, I. (1755). *Allgemeine Naturgeschichte und Theorie des Himmels*.

Keppler, M., Benisty, M., Müller, A., Henning, T., van Boekel, R., Cantalloube, F., Ginski, C., van Holstein, R. G., Maire, A. L., and Pohl, A. (2018). Discovery of a planetary-mass companion within the gap of the transition disk around PDS 70. *A&A*, 617:A44.

Klahr, H. and Kley, W. (2006). 3D-radiation hydro simulations of disk-planet interactions. I. Numerical algorithm and test cases. *A&A*, 445(2):747–758.

Kley, W. (1989). Radiation hydrodynamics of the boundary layer in accretion disks. I - Numerical methods. *A&A*, 208(1-2):98–110.

Kley, W., Bitsch, B., and Klahr, H. (2009). Planet migration in three-dimensional radiative discs. *A&A*, 506(2):971–987.

Kley, W. and Nelson, R. P. (2012). Planet-Disk Interaction and Orbital Evolution. *ARA&A*, 50:211–249.

Kolb, S. M., Stute, M., Kley, W., and Mignone, A. (2013). Radiation hydrodynamics integrated in the PLUTO code. *A&A*, 559:A80.

Laplace, P. S. d. (1796). *Exposition du système du monde*, volume 1. Mme. Ve. Courcier.

Larson, R. B. (1990). Non-linear acoustic waves in discs. *MNRAS*, 243:588–592.

Lega, E., Nelson, R. P., Morbidelli, A., Kley, W., Béthune, W., Crida, A., Kloster, D., Méheut, H., Rometsch, T., and Ziampras, A. (2021). Migration of Jupiter-mass planets in low-viscosity discs. *A&A*, 646:A166.

Levermore, C. D. and Pomraning, G. C. (1981). A flux-limited diffusion theory. *ApJ*, 248:321–334.

Lin, D. N. C. and Papaloizou, J. (1985). On the dynamical origin of the solar system. In Black, D. C. and Matthews, M. S., editors, *Protostars and Planets II*, pages 981–1072.

Lin, D. N. C. and Papaloizou, J. C. B. (1993). On the Tidal Interaction Between Protostellar Disks and Companions. In Levy, E. H. and Lunine, J. I., editors, *Protostars and Planets III*, page 749.

Lin, M.-K. and Youdin, A. N. (2015). Cooling Requirements for the Vertical Shear Instability in Protoplanetary Disks. *ApJ*, 811(1):17.

Lovelace, R. V. E., Li, H., Colgate, S. A., and Nelson, A. F. (1999). Rossby Wave Instability of Keplerian Accretion Disks. *ApJ*, 513:805–810.

Lyra, W. and Umurhan, O. M. (2019). The Initial Conditions for Planet Formation: Turbulence Driven by Hydrodynamical Instabilities in Disks around Young Stars. *PASP*, 131(1001):072001.

Manger, N. and Klahr, H. (2018). Vortex formation and survival in protoplanetary discs subject to vertical shear instability. *MNRAS*, 480(2):2125–2136.

Manger, N., Klahr, H., Kley, W., and Flock, M. (2020). High resolution parameter study of the vertical shear instability. *MNRAS*, 499(2):1841–1853.

Masset, F. (2000). FARGO: A fast eulerian transport algorithm for differentially rotating disks. *A&AS*, 141:165–173.

Mathis, J. S., Rumpl, W., and Nordsieck, K. H. (1977). The size distribution of interstellar grains. *ApJ*, 217:425–433.

Mayor, M., Queloz, D., Marcy, G., Butler, P., Noyes, R., Korzennik, S., Krockenberger, M., Nisenson, P., Brown, T., Kennelly, T., Rowland, C., Horner, S., Burki, G., Burnet, M., and Kunzli, M. (1995). 51 Pegasi. *IAU Circ.*, 6251:1.

Menou, K. and Goodman, J. (2004). Low-Mass Protoplanet Migration in T Tauri α-Disks. *ApJ*, 606(1):520–531.

Mignone, A., Bodo, G., Massaglia, S., Matsakos, T., Tesileanu, O., Zanni, C., and Ferrari, A. (2007). PLUTO: A Numerical Code for Computational Astrophysics. *The Astrophysical Journal Supplement Series*, 170:228–242.

Mignone, A., Flock, M., Stute, M., Kolb, S. M., and Muscianisi, G. (2012). A conservative orbital advection scheme for simulations of magnetized shear flows with the PLUTO code. *A&A*, 545:A152.

Miranda, R. and Rafikov, R. R. (2019). On the Planetary Interpretation of Multiple Gaps and Rings in Protoplanetary Disks Seen By ALMA. *ApJ*, 878(1):L9.

Miranda, R. and Rafikov, R. R. (2020a). Planet-Disk Interaction in Disks with Cooling: Basic Theory. *ApJ*, 892(1):65.

Miranda, R. and Rafikov, R. R. (2020b). Gaps and Rings in Protoplanetary Disks with Realistic Thermodynamics: The Critical Role of In-plane Radiation Transport. *ApJ*, 904(2):121.

Molyarova, T., Akimkin, V., Semenov, D., Henning, T., Vasyunin, A., and Wiebe, D. (2017). Gas Mass Tracers in Protoplanetary Disks: CO is Still the Best. *ApJ*, 849(2):130.

Müller, T. W. A., Kley, W., and Meru, F. (2012). Treating gravity in thin-disk simulations. *A&A*, 541:A123.

Nelson, R. P., Gressel, O., and Umurhan, O. M. (2013). Linear and non-linear evolution of the vertical shear instability in accretion discs. *MNRAS*, 435(3):2610–2632.

Ogilvie, G. I. and Lubow, S. H. (2002). On the wake generated by a planet in a disc. *MNRAS*, 330(4):950–954.

Paardekooper, S.-J., Lesur, G., and Papaloizou, J. C. B. (2010). Vortex Migration in Protoplanetary Disks. *ApJ*, 725(1):146–158.

Pepliński, A., Artymowicz, P., and Mellema, G. (2008). Numerical simulations of type III planetary migration - III. Outward migration of massive planets. *MNRAS*, 387(3):1063–1079.

Pfeil, T. and Klahr, H. (2021). The Sandwich Mode for Vertical Shear Instability in Protoplanetary Disks. *ApJ*, 915(2):130.

Rab, C., Baldovin-Saavedra, C., Dionatos, O., Vorobyov, E., and Güdel, M. (2016). The Gas Disk: Evolution and Chemistry. *Space Sci. Rev.*, 205(1-4):3–40.

Rafikov, R. R. (2002). Nonlinear Propagation of Planet-generated Tidal Waves. *ApJ*, 569(2):997–1008.

Rafikov, R. R. (2016). Protoplanetary Disk Heating and Evolution Driven by Spiral Density Waves. *ApJ*, 831:122.

Ribas, Á., Merín, B., Bouy, H., and Maud, L. T. (2014). Disk evolution in the solar neighborhood. I. Disk frequencies from 1 to 100 Myr. *A&A*, 561:A54.

Rodenkirch, P. J., Rometsch, T., Dullemond, C. P., Weber, P., and Kley, W. (2021). Modeling the nonaxisymmetric structure in the HD 163296 disk with planet-disk interaction. *A&A*, 647:A174.

Rometsch, T., Ziampras, A., Kley, W., and Béthune, W. (2021). Survival of planet-induced vortices in 2D disks. *A&A*, None:None.

Sachs, A. (1974). Babylonian Observational Astronomy. *Philosophical Transactions of the Royal Society of London Series A*, 276(1257):43–50.

Schäfer, U., Johansen, A., and Banerjee, R. (2020). The coexistence of the streaming instability and the vertical shear instability in protoplanetary disks. *A&A*, 635:A190.

Semenov, D., Henning, T., Helling, C., Ilgner, M., and Sedlmayr, E. (2003). Rosseland and Planck mean opacities for protoplanetary discs. *A&A*, 410:611–621.

Shakura, N. I. and Sunyaev, R. A. (1973). Reprint of 1973A&A....24..337S. Black holes in binary systems. Observational appearance. *A&A*, 500:33–51.

Stoll, M. H. R. and Kley, W. (2014). Vertical shear instability in accretion disc models with radiation transport. *A&A*, 572:A77.

Stoll, M. H. R. and Kley, W. (2016). Particle dynamics in discs with turbulence generated by the vertical shear instability. *A&A*, 594:A57.

Stoll, M. H. R., Kley, W., and Picogna, G. (2017). Anisotropic hydrodynamic turbulence in accretion disks. *A&A*, 599:L6.

Tassoul, J.-L. (1978). *Theory of rotating stars*.

Teague, R., Bae, J., Bergin, E. A., Birnstiel, T., and Foreman-Mackey, D. (2018). A Kinematical Detection of Two Embedded Jupiter-mass Planets in HD 163296. *ApJ*, 860(1):L12.

Thun, D. and Kley, W. (2018). Migration of planets in circumbinary discs. *A&A*, 616:A47.

Toomre, A. (1964). On the gravitational stability of a disk of stars. *ApJ*, 139:1217–1238.

Toro, E. F., Spruce, M., and Speares, W. (1994). Restoration of the contact surface in the HLL-Riemann solver. *Shock waves*, 4(1):25–34.

Urpin, V. and Brandenburg, A. (1998). Magnetic and vertical shear instabilities in accretion discs. *MNRAS*, 294(3):399–406.

Van Leer, B. (1974). Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme. *Journal of computational physics*, 14(4):361–370.

Youdin, A. N. and Goodman, J. (2005). Streaming Instabilities in Protoplanetary Disks. *ApJ*, 620(1):459–469.

Zhang, S. and Zhu, Z. (2020). The effects of disc self-gravity and radiative cooling on the formation of gaps and spirals by young planets. *MNRAS*, 493(2):2287–2305.

Zhang, S., Zhu, Z., Huang, J., Guzmán, V. V., Andrews, S. M., Birnstiel, T., Dullemond, C. P., Carpenter, J. M., Isella, A., Pérez, L. M., Benisty, M., Wilner, D. J., Baruteau, C., Bai, X.-N., and Ricci, L. (2018). The Disk Substructures at High Angular Resolution Project (DSHARP). VII. The Planet-Disk Interactions Interpretation. *ApJ*, 869(2):L47.

Zhu, Z., Dong, R., Stone, J. M., and Rafikov, R. R. (2015). The Structure of Spiral Shocks Excited by Planetary-mass Companions. *ApJ*, 813(2):88.

Ziampras, A., Ataiee, S., Kley, W., Dullemond, C. P., and Baruteau, C. (2020a). The impact of planet wakes on the location and shape of the water ice line in a protoplanetary disk. *A&A*, 633:A29.

Ziampras, A., Kley, W., and Dullemond, C. P. (2020b). Importance of radiative effects in gap opening by planets in protoplanetary disks. *A&A*, 637:A50.

A Implementation of FLD in 3D

In collaboration with Dr. William Béthune (henceforth referred to as WB), we implemented radiation transport into PLUTO using the flux-limited diffusion approach (FLD) similar to Kolb et al. (2013). We integrate Eq. (2.6) by first solving for the radiation energy $E_{\rm rad}$ in the next timestep (denoted $E'_{\rm rad}$) using an implicit scheme:

$$\frac{\partial E_{\rm rad}}{\partial t} + \nabla \cdot \boldsymbol{F}_{\rm rad} = \kappa_{\rm P}\rho c(a_{\rm R}T^4 - E_{\rm rad}) \Rightarrow \frac{E'_{\rm rad} - E_{\rm rad}}{\Delta t} \approx \kappa_{\rm P}\rho c(a_{\rm R}T'^4 - E'_{\rm rad}) - \nabla \cdot \boldsymbol{F}'_{\rm rad} \tag{A.1}$$

This can be represented as a linear equation of the form $\overline{\mathbf{M}}_{\rm rad} \,.\, E'_{\rm rad} = E_{\rm rad}$, where $\overline{\mathbf{M}}_{\rm rad}$ is the radiation matrix and $E'_{\rm rad}$ is a vector of size $N_{\rm cells}$ that contains the solution after a timestep Δt. Since $\overline{\mathbf{M}}_{\rm rad}$ consists of mostly zeroes (the number of nonzero elements is approximately $(1+2D)/N_{\rm cells}$ where D is the number of dimensions), it can be inverted reasonably cheaply using sparse matrix inversion techniques. To that end, we mainly use the PetSc library (Balay et al., 2018) but have also implemented a Successive Over-Relaxation (SOR) scheme that produces the same results.

After obtaining $E'_{\rm rad}$, we update the gas specific internal energy ϵ following (Kolb et al., 2013):

$$\epsilon' = c_{\rm v}\frac{\kappa_{\rm P}c\left(3a_{\rm R}T^4 + E'_{\rm rad}\right)\Delta t + c_{\rm v}T}{c_{\rm v} + 4\kappa_{\rm P}ca_{\rm R}T^3\Delta t} \tag{A.2}$$

A.1 Motivation

Our FLD module offers several upgrades compared to Kolb et al. (2013), even though the numerical methods are the same. Namely:

- it supports all PLUTO geometry options (Cartesian, cylindrical, spherical) instead of only spherical, making it usable for physical problems in non-spherical geometries;
- it supports 1–3 dimensions instead of only 3D, alleviating the need for full-3D, "pseudo-1–2D" setups and gaining a significant speedup in the process;
- discretization has been rewritten to closely follow PLUTO's grid structures, resulting in more accurate and faster calculations;
- boundary conditions have been corrected such that the required boundary condition (Dirichlet, von Neumann, custom) is applied precisely on the boundary wall instead of in the first ghost cell;
- it has been written for `PLUTO 4.3`, making it compatible with the latest versions of the code (also `4.4` and higher).

Overall, our implementation was much more robust when comparing against the module by Kolb et al. (2013), with the latter often crashing at very high grid resolutions. When using our new model, we also observed a typical speedup factor of up to 4–5 in $r\theta$-2D models depending on grid resolution and parallelization level. Due to the option to use a "true" 2D setup as opposed to a "pseudo-2D" one (which would require 2–4 cells with periodic boundaries in the azimuthal direction), this speedup factor can increase to 20–30 depending on the setup.

A.2 Testing

We tested our module by carrying out two sets of tests, namely *static* and *dynamic*. For our *static* tests, we verified the implementation of our physics for all geometry–dimension combinations by running models that converge to an analytical steady state. Then, with our *dynamic* tests we verified the temporal accuracy of our solver by exciting sound waves, simulating their radiative damping, and comparing to theoretical estimates. All quantities are quoted in code units unless otherwise specified.

A.2.1 Static tests

Our static tests aim to reach an equilibrium state by solving Eq. (2.6) repeatedly until convergence, without evolving the gas according to the hydrodynamics equations in Eq. (2.1). This is achieved by using a fixed gas density profile, setting the gas velocity to zero during the initialization step, and disabling PLUTO's hydro update step.

Our first static test setup aims to verify the implementation of our equations and boundary conditions. The setup is as follows:

- Physics: FLD only, no hydro evolution.
- Grid: 2D Cartesian, $x, y \in [0, 1]$ with 100 cells each, uniformly spaced.
- Boundary conditions: $E_{\rm rad}|_{x=0} = E_0$, $E_{\rm rad}|_{x=1} = 2E_0$, periodic in y.
- $\rho, \kappa_{\rm R}$ are constant, $\lambda = 1/3$.

We are therefore effectively solving the equation

$$\frac{\partial^2 E_{\rm rad}}{\partial x^2} = 0 \Rightarrow E_{\rm an}(x, y) = E_0(1 + x/x_0). \tag{A.3}$$

The results of our test are shown in Fig. 53. The relative error levels combined with the model being symmetric in the y direction support the validity of our implementation of boundary conditions and grid discretization.

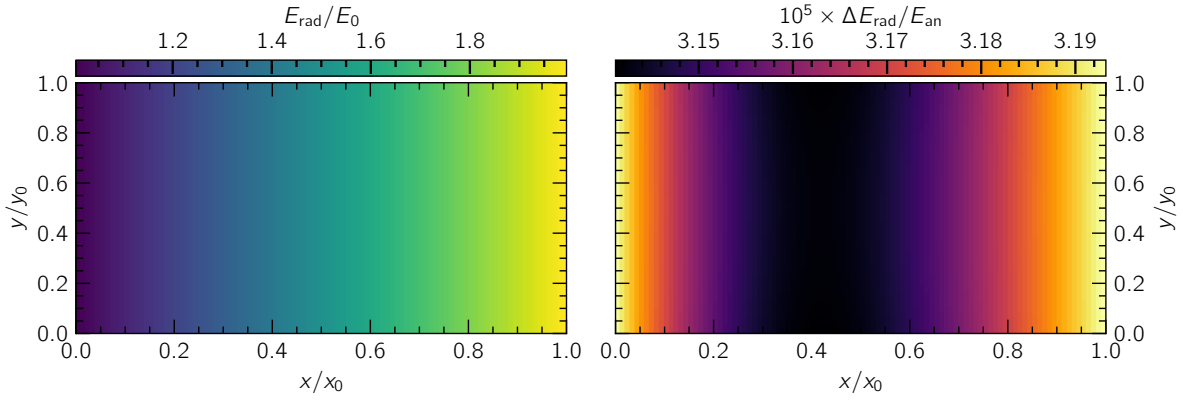

Figure 53. Results of our 2D Cartesian static test. Left: the equilibrium solution. Right: relative error between our solution and the analytical prediction.

For our second test, we verified our implementation in different geometries and grid combinations. As an example, we design a 1D setup as follows:

- Physics: FLD only, with a constant heating source term $Q_+ = \frac{1}{\gamma - 1}$. No hydro evolution.
- Grid: 1D spherical, $r \in [1, 5]$, composite grid.

- Boundary conditions: $E_{\rm rad}|_{r=1} = E_0$, $\left.\frac{\partial E_{\rm rad}}{\partial r}\right|_{r=5} = 0$.
- $\rho, \kappa_{\rm R}$ are constant, $\lambda = 1/3$.

We use a composite grid in r that is "stretched" between 1–2 with 80 cells, logarithmic between 2–4 with 50 cells, and uniform between 4–5 with 70 cells, for a total of 200 cells. This is done in order to test whether our discretization is compatible with PLUTO's different grid options. A schematic of our grid is shown in Fig. 54.

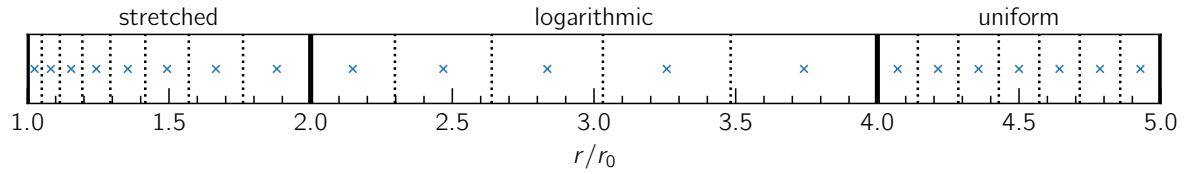

Figure 54. A schematic of the radial grid used in the 1D spherical static test. The crosses and dotted lines denote cell centers and interfaces respectively.

This time, the analytical solution is given by:

$$\nabla^2 E_{\rm rad} = Q_+ \Rightarrow E_{\rm an} = -\frac{ar^2}{3} - \frac{c_1}{r} + c_2, \qquad a = \frac{Q_+}{2}\frac{\kappa_{\rm R}\rho}{\lambda c}, \tag{A.4}$$

where $c_1 = \frac{250}{3}a$ and $c_2 = E_0 + \frac{251}{3}a$. The addition of a heating term forces the solution to depend on physical quantities through a, which also lets us test whether unit conversions within the code are handled correctly. The results of our test, which show a very good agreement between model and prediction down to our tolerance levels, are shown in Fig. 55.

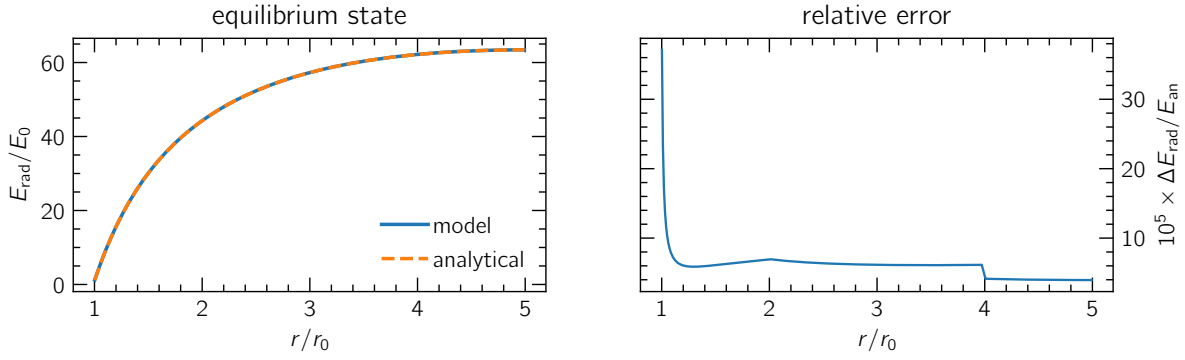

Figure 55. Results of our 1D spherical static test. Left: the equilibrium solution. Right: relative error between our solution and the analytical prediction. The error behaves differently at different grid patches, but remains very small overall.

A.2.2 Dynamic tests

Having verified that our module's physics are implemented correctly, we now test the temporal accuracy of our solver. This is done by exciting an eigenmode of the 1D radiative hydrodynamics equations in a periodic domain. Essentially, this corresponds to a standing acoustic wave that oscillates with a frequency $\omega_{\rm rad}$ and decays over a timescale $\tau_{\rm damp}$ as it interacts with the radiation field. An example of such a configuration and its time evolution is shown in Fig. 56. We carry out a series of tests for two different densities $\rho = 10^{-12}, 10^{-13}$ g/cm^3 and a wide range of opacities $\kappa = 10^{-2}$–10^3 cm$^2/g_{\rm gas}$ to cover both the optically thin and thick regimes. We then log the amplitude of the pressure $P(t)$ at the crest of each wave as a function of time and fit it with an exponentially-decaying sine curve:

$$P_{\rm fit} = P_0 \sin(\omega_{\rm rad} t + \varphi_0)\, e^{-t/\tau_{\rm damp}} \tag{A.5}$$

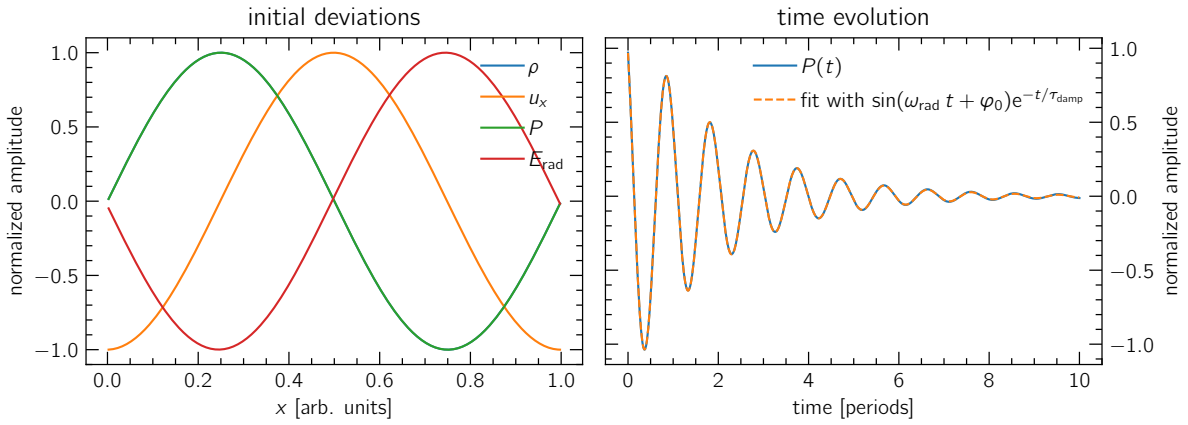

Figure 56. A sample configuration of our dynamic test. Left: our initial conditions for all quantities, the amplitudes and phases of which were provided by WB. Right: time evolution of the pressure $P(t)$ at the crest of the wave, fit with a decaying sine function.

Finally, we compare the results of our fit with theoretical estimates of $\omega_{\rm rad}$ and $\tau_{\rm damp}$, provided by WB. Fig. 57 shows that we recover the predicted eigenvalues to better than 10^{-2} accuracy overall in both the optically thin and thick limits.

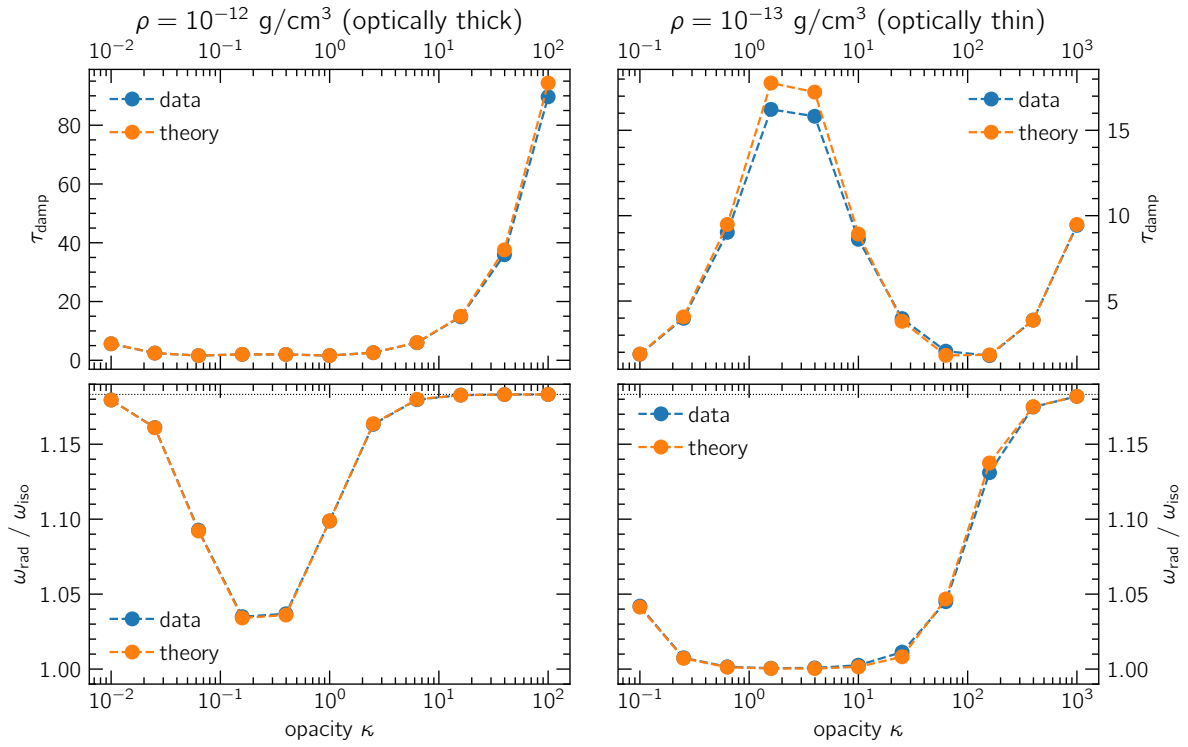

Figure 57. The oscillation frequency and damping timescale of the eigenmodes excited in our dynamic tests for both optically thin and thick configurations. Frequencies are normalized to the isothermal $\omega_{\rm iso} = 2\pi$ for our setup. The horizontal dotted line corresponds to $\omega_{\rm adb} = \sqrt{\gamma}\omega_{\rm iso}$.

B Ray-traced stellar irradiation

Our implementation of ray-traced irradiation, again in collaboration with WB, follows the methods of Kolb et al. (2013) but was motivated by reasons similar to those presented in Sect. A.1. While this module has only been implemented for a spherical geometry due to the ease of ray-tracing for such a configuration, our version supports 1–3 dimensions, follows PLUTO's grid discretization, and can be used as a standalone module (i.e., it is not necessarily coupled to the FLD module). In this section we describe the discretization process in detail, present a simple verification test for the module, and note possible caveats related to its implementation.

B.1 Discretization

Similar to our approach in Sect. 2.2.2, we consider a cell at radius r from the star. This cell has a radial width δr such that its left (inner) and right (outer) interfaces are at $r_{\rm L} = r - \delta r/2$ and $r_{\rm R} = r + \delta r/2$ respectively. The incoming and outgoing irradiation fluxes on the left and right interfaces of the cell $L_{\rm in}$ and $L_{\rm out}$ are then given by Eqs. (2.10) and (2.11) such that the absorbed energy density is

$$S_{\rm irrad} = \frac{L_{\rm in} - L_{\rm out}}{V}, \qquad L_{\rm in} = A_{\rm L}\frac{L_\star}{4\pi r_{\rm L}^2}e^{-\tau_{\rm abs}(r_{\rm L})}, \qquad L_{\rm out} = A_{\rm R}\frac{L_\star}{4\pi r_{\rm R}^2}e^{-\tau_{\rm abs}(r_{\rm R})}, \tag{B.1}$$

where $\tau_{\rm abs}$ is given by Eq. (2.9). The cell volume V and cell surface at both interfaces $A_{\rm L}$, $A_{\rm R}$ are provided by PLUTO's grid structures, so it is not necessary to recompute them during each timestep. Computation of the radially-integrated $\tau_{\rm abs}$ is done by first evaluating the optical depth contribution of each cell

$$\delta\tau_{\rm abs}(r) = \int_{r_{\rm L}}^{r_{\rm R}} \kappa_{\rm P}\rho {\rm d}r' = \kappa_{\rm P}\rho\delta r, \tag{B.2}$$

and then summing over all contributions radially outwards. This summation, which was implemented by WB, is handled by a radial MPI communicator that only communicates with cells along the same "ray" (i.e., for a constant θ and φ index). This method provides a fast and clean implementation of ray-tracing and is generally the most computationally expensive substep within the irradiation module.

We note that, in principle, the inner radial boundary $r_{\rm in}$ of our domain does not correspond to the physical inner rim of the disk. As a result, an additional contribution to the absorption optical depth from the "inner disk" between the stellar surface and $r_{\rm in}$ has to be considered. Therefore we introduce an optical depth offset τ_0 such that

$$\tau_{\rm abs}(r) = \int_{R_\star}^{r} \kappa_{\rm P}\,\rho\,{\rm d}r' = \int_{R_\star}^{r_{\rm in}} \kappa_{\rm P}\,\rho\,{\rm d}r' + \int_{r_{\rm in}}^{r} \kappa_{\rm P}\,\rho\,{\rm d}r' = \tau_0 + \int_{r_{\rm in}}^{r} \kappa_{\rm P}\,\rho\,{\rm d}r' \tag{B.3}$$

Following Flock et al. (2017), we define $\tau_0(\theta,\varphi) = \kappa_{\rm P}\,\rho(r_{\rm in})\,(R_{\rm in} - 3R_\star)$. This results in the disk being heated primarily through its surface layers with a thin, hot buffer zone near the inner radial boundary for $T_{\rm rad}$.

After $\tau_{\rm abs}(r)$ and $S_{\rm irrad}$ have been computed, the gas is heated with a simple Euler step

$$\frac{\partial(\rho\epsilon)}{\partial t} = S_{\rm irrad} \Rightarrow P' \approx P + (\gamma - 1)\,S_{\rm irrad}\,\Delta t \tag{B.4}$$

B.2 Testing

Our test consists of a simple 1D model aiming to verify our implementation of the physics, similar to our 1D static test in Sect. A.2.1. The setup is as follows:

- Physics: FLD and irradiation, no hydro evolution.
- Grid: 1D spherical, $r \in [1, 5]$ with 200 cells, uniformly spaced.
- Boundary conditions: $E_{\rm rad}|_{r=1} = E_0$, $\left.\frac{\partial E_{\rm rad}}{\partial r}\right|_{r=5} = 0$.
- $\rho, \kappa_{\rm R} = \kappa_{\rm P} = \kappa$ are constant, $\lambda = 1/3$, $\tau_0 = 5$.

To calculate an analytical solution we start from Eq. (2.6), adding the irradiation flux of Eq. (2.12) on the RHS:

$$\nabla \cdot \boldsymbol{F}_{\rm rad} = S_{\rm irrad}. \tag{B.5}$$

To compute $S_{\rm irrad}$ we start from Eq. (B.1), and using that $A = r^2\Delta\Omega$ and $V \approx r^2\delta r\Delta\Omega$ we write

$$S_{\rm irrad} \approx \frac{L_\star}{4\pi r^2 \delta r}\left(e^{-\tau_{\rm abs}(r_{\rm L})} - e^{-\tau_{\rm abs}(r_{\rm R})}\right), \tag{B.6}$$

where $\tau_{\rm abs}(r_{\rm L}) = \tau_0 + \kappa\rho(r - \delta r/2 - r_{\rm in})$ and $\tau_{\rm abs}(r_{\rm R}) = \tau_{\rm abs}(r_{\rm L}) + \kappa\rho\delta r$. Since δr is constant, we finally have

$$S_{\rm irrad} \approx \frac{L_\star}{4\pi r^2\delta r}\, e^{-\tau_0 - \kappa\rho(r_{\rm in} + \delta r/2)}\left(1 - e^{-\kappa\rho\delta r}\right) e^{-\kappa\rho r} \Rightarrow S_{\rm irrad} \approx A\frac{e^{-\kappa\rho r}}{r^2}, \quad A = \text{const.} \tag{B.7}$$

We can now integrate Eq. (B.5) to obtain the analytical solution for $E_{\rm rad}$

$$\frac{\lambda c}{\kappa\rho}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial E_{\rm rad}}{\partial r}\right) = A\frac{e^{-\kappa\rho r}}{r^2} \Rightarrow E_{\rm rad} = a\kappa\rho\,\mathrm{Ei}(-\kappa\rho r) - a\frac{e^{-\kappa\rho r}}{r} - \frac{c_1}{r} + c_2, \tag{B.8}$$

where Ei is the exponential integral, $a = A/\lambda c$, and c_1, c_2 are integration constants that satisfy our boundary conditions.

Fig. 58 shows a comparison between our numerical model and the analytical solution calculated above. We note that, due to the approximation used in calculating the analytical solution ($V \approx r^2\delta r\Delta\Omega$), the two curves will not agree exactly but the agreement will improve for larger r as $\delta r/r$ decreases. Nevertheless, our prediction follows the model very well.

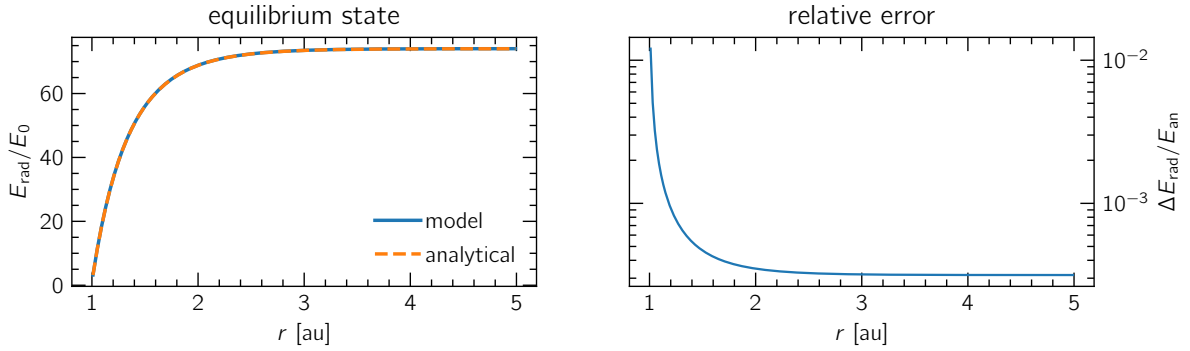

Figure 58. Results of our irradiation test. Left: the equilibrium solution. Right: relative error between our solution and the analytical prediction.

B.3 Caveats

In the previous section we verified that the irradiation module produces accurate results by comparing the numerical output of a 1D model with its analytical solution. Nevertheless, we found that the Euler scheme which we use during the pressure update step is extremely sensitive to the timestep Δt, and as a result so is the equilibrium temperature we obtain in the regions of the disk that are directly exposed to starlight ($\tau_{\rm abs} < 1$). Fortunately, the temperature profile remains consistent regardless of timestep beneath the disk surface layers ($\tau_{\rm abs} > 1$).

To showcase this behavior, we ran an $r\theta$-2D model of a protoplanetary disk in hydrodynamic equilibrium (see Eq. (2.3)) and with typical disk parameters: $p = -2.25$, $q = -0.5$, $h = 5\%$ at 5.2 au). We do not evolve the disk hydrodynamically, but instead allow it to reach thermal equilibrium between stellar irradiation and radiation transport with a given timestep. Our results are shown in Fig. 59. The midplane temperature $T_{\rm mid}$ follows a typical radial profile for an irradiated disk and remains consistent for all timesteps. We highlight the very large differences in $T_{\rm surface}$ for different timesteps, but also note that a typical timestep for our problem would be in the range 10^{-4}–10^{-3} P_0, which should alleviate this problem.

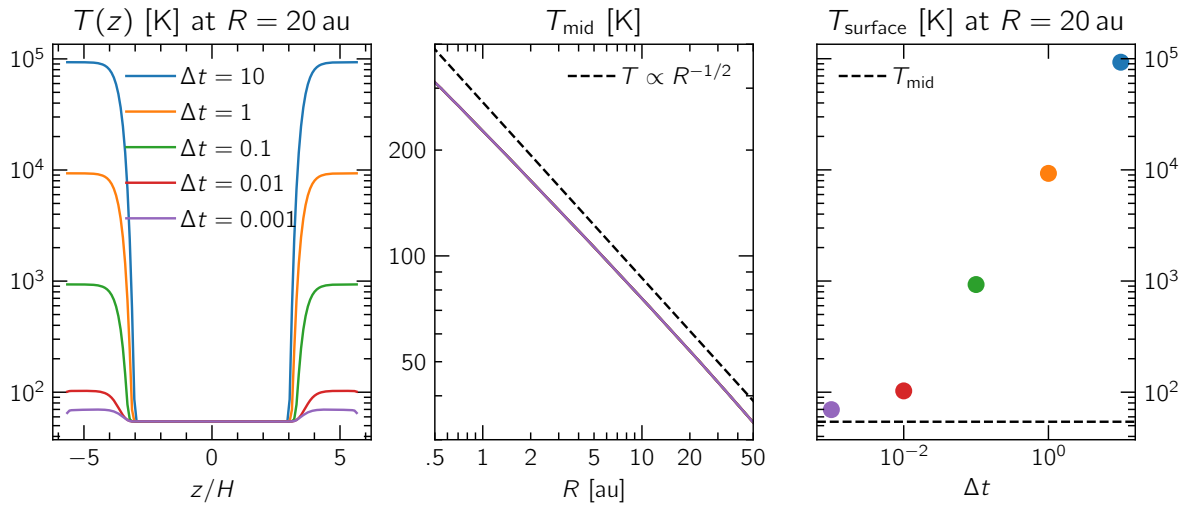

Figure 59. Radial and vertical temperature profiles as a function of timestep (quoted in orbital periods at $R = 1$ au) for a test with irradiation and radiation transport using our ray-tracing and FLD modules. The temperature at the disk surface depends strongly on the timestep, but the midplane temperature remains consistent and in agreement with an expected profile for an irradiated disk ($T_{\rm mid} \propto R^{-1/2}$).

In their module, Kolb et al. (2013) instead opt to update the gas internal energy by incorporating $S_{\rm irrad}$ into the FLD update step (see Eq. (A.2)). We tried this method, but found that the results did not agree with the analytical solution of the 1D model in the previous section. As a result, we chose to keep using the Euler update step. While this will likely result in inaccurate temperatures at the disk corona as discussed above, our area of interest (the vertically isothermal disk interior below the $\tau_{\rm abs} = 1$ surface) is completely unaffected.

C Publications

Here we list our papers that were published over the course of this project, and describe our contribution to each work. The papers listed here are appended below.

1. Ziampras et al. (2020a)
Title: The impact of planet wakes on the location and shape of the water ice line in a protoplanetary disk
Authors: Alexandros Ziampras, Sareh Ataiee, Wilhelm Kley, Cornelis P. Dullemond, Clément Baruteau
Journal: Astronomy & Astrophysics
Status: published
DOI: 10.1051/0004-6361/201936495
Contribution: Implementation of numerical setups; execution of simulations; analysis of results; construction of Figs. 4–10 as well as figures in the Appendix, adaptation of remaining figures; authorship of Sects. 4–7 and Appendix, partial authorship of abstract and Sects. 1, 2, 6; development of FLD module; correspondence with referee.
The publication can be found in Sect. C.1, and an overview is provided in Sect. 4.

2. Ziampras et al. (2020b)
Title: Importance of radiative effects in gap opening by planets in protoplanetary disks
Authors: Alexandros Ziampras, Wilhelm Kley, Cornelis P. Dullemond
Journal: Astronomy & Astrophysics
Status: published
DOI: 10.1051/0004-6361/201937048
Contribution: Selection of physical parameters; implementation of numerical setups; execution of simulations; analysis of results; construction of figures; authorship of Sects. 1–3, 5 and Appendix, partial authorship of abstract and Sect. 4; correspondence with referee.
The publication can be found in Sect. C.2, and an overview is provided in Sect. 5.

3. Rometsch et al. (2021)
Title: Survival of planet-induced vortices in 2D disks
Authors: Thomas Rometsch, Alexandros Ziampras, Wilhelm Kley, William Béthune
Journal: Astronomy & Astrophysics
Status: published
DOI: 10.1051/0004-6361/202142105
Contribution: Suggestions towards the selection of physical parameters; implementation of PLUTO numerical setups; execution of PLUTO simulations; partial analysis of results; construction of Figs. 1, 3, 9, A.1, suggestions towards the construction of remaining figures; authorship of Sects. 2, 5 and Appendix A, partial authorship of abstract and Sects. 1, 3, 4, 6, 7.
The publication can be found in Sect. C.3, and an overview is provided in Sect. 6.

C.1 Ziampras et al. (2020a)

A&A 633, A29 (2020)
https://doi.org/10.1051/0004-6361/201936495

Astronomy & Astrophysics

The impact of planet wakes on the location and shape of the water ice line in a protoplanetary disk

Alexandros Ziampras[1], Sareh Ataiee[1], Wilhelm Kley[1], Cornelis P. Dullemond[2], and Clément Baruteau[3]

[1] Institut für Astronomie und Astrophysik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany
e-mail: alexandros.ziampras@uni-tuebingen.de
[2] Institute for Theoretical Astrophysics, Zentrum für Astronomie, Heidelberg University, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
[3] Research Institute in Astrophysics and Planetology, University of Toulouse, 14 Avenue Edouard Belin, 31400 Toulouse, France

Received 10 August 2019 / Accepted 17 October 2019

ABSTRACT

Context. Planets in accretion disks can excite spiral shocks and if these planets are massive enough, they can even open gaps in their vicinity. Both of these effects can influence the overall thermal structure of the disk.
Aims. We model planets of different masses and semimajor axes in disks of various viscosities and accretion rates to examine their impact on disk thermodynamics and to highlight the mutable, non-axisymmetric nature of ice lines in systems with massive planets.
Methods. We conducted a parameter study using numerical hydrodynamics simulations where we treated viscous heating, thermal cooling, and stellar irradiation as additional source terms in the energy equation, with some runs including radiative diffusion. Our parameter space consists of a grid containing different combinations of planet and disk parameters.
Results. Both gap opening and shock heating can displace the ice line, with the effects amplified for massive planets in optically thick disks. The gap region can split an initially hot ($T > 170$ K) disk into a hot inner disk and a hot ring just outside of the planet's location, while shock heating can reshape the originally axisymmetric ice line into water-poor islands along spirals. We also find that radiative diffusion does not alter the picture significantly in this context.
Conclusions. Shock heating and gap opening by a planet can effectively heat up optically thick disks and, in general, they can move or reshape the water ice line. This can affect the gap structure and migration torques. It can also produce azimuthal features that follow the trajectory of spiral arms, creating hot zones which lead to "islands" of vapor and ice around spirals that could affect the accretion or growth of icy aggregates.

Key words. protoplanetary disks – planet-disk interactions – planets and satellites: formation – hydrodynamics

1. Introduction

Protostellar disks are the birth sites of all sorts of planets. Several observations, such as the discovery of PDS 70b (Keppler et al. 2018) and the recent DSHARP survey (Andrews et al. 2018) have spatially resolved such disks, providing valuable constraints on their composition, structure, and the possible planets they might harbor. Dust continuum observations reveal annular structures and non-axisymmetric features, such as spirals, crescents, or blobs, all of which are consistent with the planet formation scenario (Zhang et al. 2018). According to this scenario, a sufficiently massive planet can trap dust particles by forming pressure maxima (e.g., Ataiee et al. 2018) at radii close to its semimajor axis as it launches density waves in the form of spiral arms that permeate the disk (Ogilvie & Lubow 2002). These pressure traps can allow dust particles to become concentrated enough for their emission to be observable, and also provide an environment for them to collide and grow.

Dust growth is expected to be further facilitated around opacity transition regions (Drążkowska & Alibert 2017; Zhang et al. 2015). Common dust opacity models are, in principle, density-, and temperature-dependent, with boundaries defined at conditions where aggregates of certain composition change phase to the extent that crossing between two opacity regimes can change the absorption and emission properties of the disk. For example, the water content of ice-coated particles sublimates at the so-called water ice line around $T_{\rm ice} \approx 170$ K (Lin & Papaloizou 1985), with small variations depending on model assumptions. This temperature marks the first opacity transition threshold that particles will cross as they drift inwards according to several opacity models (e.g., Bell & Lin 1994; Semenov et al. 2003). Since water can only be found on particles outside of this ice line, its location can provide insight and constraints on the origin of water content for planetesimals and young planets in an evolving protostellar disk, depending on the disk's temperature profile (Bitsch et al. 2019).

The disk's thermal structure depends on the balance between heating and cooling terms. Kley & Crida (2008) show that accounting for radiation transport, instead of treating the disk as locally isothermal, can have significant effects on the migration of super-Earths by slowing down or even reversing the migration rate. Additionally, Rafikov (2016) shows that shock heating due to planet-induced spirals can be a significant heat source in the inner few au of the disk. Evidently, the optically thick region near the star can reach high densities and temperatures, which could lead to an important contribution by shock heating to the energy content of the disk.

In this study, we investigate the conditions under which planet shock heating can significantly raise temperatures and the degree to which spirals can affect the location and shape of the

A&A 633, A29 (2020)

water ice line. Based on our findings, we speculate about the possible implications for dust and planetesimal growth around the ice line.

In Sect. 2, we calculate an estimate for the amount of heat a planet can pump into the disk through shock heating. Our physical framework, as well as the numerical setup, is described in Sect. 3. In Sects. 4 and 5, we present our results regarding the disk structure and ice line shape, respectively. In Sect. 6, we comment on our findings and discuss their potential implications, while Sect. 7 contains a summary of our work, along with our conclusions.

2. Shock heating

Planet-induced spiral arms form as a result of density waves shearing in the Keplerian disk flow as they propagate away from the planet (Kley & Nelson 2012). They are overdensities with respect to the disk "background" (azimuthally-averaged) profile that can steepen into shocks as they travel through the disk (Goodman & Rafikov 2001). In an adiabatic framework, we can expect a pressure jump at the location of the shock, which can lie close to the planet (Zhu et al. 2015). This pressure jump can generate heat near the planet, potentially affecting the temperature profile near the corotating region. Thus, the question is how important this shock heating can be when compared to other heat sources in the disk (e.g., viscosity and stellar irradiation).

In order to get an idea of the prominence of spiral shocks as a heat-generating mechanism, it is worthwhile to first estimate their contribution theoretically and compare it to other sources of heat in the disk. We follow a line of thought similar to that of Rafikov (2016) and calculate the heating by an adiabatic spiral shock for an assumed density jump at the shock.

Heating by a spiral shock can be considered as a three-phase process: (1) heating by the shocks; (2) decompression; and (3) settling to the pre-shock density. In this subsection, we refer to the pre-shock quantities (before phase 1) by the subscript 1, to the decompression phase by subscript 2, and to the post-shock state by subscript 3. The following calculation is performed in the shock's comoving frame. The shock heating rate can be estimated by calculating the specific internal energy difference between phase 1 and phase 3 for each passage of the shock and then dividing it by the time between two passages. Knowing the pressure p and surface density Σ in each of the three phases, we can calculate the specific internal energy via $e = p/(\Sigma(\gamma-1))$. The classical jump condition and equation of state can give us the values for all necessary quantities. In our calculations, we use the surface density $\Sigma = \int_{-\infty}^{+\infty} \rho dz$ instead of the density ρ. Note that the jump conditions are also valid if ρ is replaced by Σ because during the shock, the disk does not have enough time to expand vertically and change the local density. This allows us to use two-dimensional (2D) hydrodynamic simulations to test the predictions of this analytical model.

When the shock hits the pre-shock gas between phase 1 and phase 2, the density and pressure at the second phase can be given by the Rankine–Hugoniot jump condition that reads

$$\frac{\Sigma_2}{\Sigma_1} = \frac{(\gamma+1)\mathcal{M}_1^2}{(\gamma-1)\mathcal{M}_1^2+2}, \tag{1}$$

$$\frac{p_2}{p_1} = \frac{2\gamma\mathcal{M}_1^2-(\gamma-1)}{\gamma+1} \tag{2}$$

where γ is the adiabatic index and $\mathcal{M}$ denotes the Mach number.

The decompression between phase 2 and phase 3 can be either adiabatic or isothermal depending on disk cooling. If the shocks are in the optically thin part of the disk, the inserted energy can be easily radiated away, the post-shock temperature returns to its pre-shock value rapidly (except in a very narrow region of the shock itself), and the decompression would be isothermal. Conversely, in the optically thick part of the disk where the energy cannot quickly escape, the decompression is adiabatic. Because we are interested in the cases where shocks heat the disk up and change its temperature structure, we choose the adiabatic decompression. Therefore, the pressure and density before and after decompression can be given by an adiabatic equation of state as

$$\frac{p_3}{p_2} = \left(\frac{\Sigma_3}{\Sigma_2}\right)^\gamma. \tag{3}$$

Because the gas density will return to its pre-shock value, we can replace Σ_3 with Σ_1 so that:

$$p_3 = p_2\left(\frac{\Sigma_1}{\Sigma_2}\right)^\gamma. \tag{4}$$

Let us assume that the time between two passages of a shock through a specific location of radius r in the disk is $t_{\rm pass} = 2\pi/|\Omega_{\rm K}(r) - \Omega_{\rm K}(r_{\rm p})|$, where $\Omega_{\rm K} = \sqrt{{\rm G}(M_* + M_{\rm p})/r^3}$ is the Keplerian frequency, G the gravitational constant, M_* and $M_{\rm p}$ the masses of the star and planet respectively, and $r_{\rm p}$ the planet's semimajor axis. The amount of heat per unit time (averaged over many passages) is then:

$$Q_{\rm sh} = \frac{\Delta(\Sigma e)}{\Delta t} = \frac{e_3-e_1}{t_{\rm pass}}\Sigma_1 = \frac{1}{t_{\rm pass}(\gamma-1)}\left[p_2\left(\frac{\Sigma_1}{\Sigma_2}\right)^\gamma - p_1\right]. \tag{5}$$

Expressing $\mathcal{M}_1$ from Eq. (1) we obtain:

$$\mathcal{M}_1^2 = \frac{2(\Sigma_2/\Sigma_1)}{(\gamma+1)-(\gamma-1)(\Sigma_2/\Sigma_1)}. \tag{6}$$

Inserting this into Eq. (2), we can remove p_2 from the above equation and obtain:

$$Q_{\rm sh} = \frac{p_1}{t_{\rm pass}(\gamma-1)} \times \left[\left(\frac{1}{\sigma}\right)^\gamma \frac{(\gamma+1)\sigma-(\gamma-1)}{(\gamma+1)-(\gamma-1)\sigma} - 1\right], \tag{7}$$

where $\sigma := \Sigma_2/\Sigma_1$ is the "shock strength". This equation is identical to Eq. (16) of Rafikov (2016) assuming a one-armed spiral. In the literature, there is no straightforward way to find the strength of planetary spirals. Following Rafikov (2016), we take σ as a free parameter and compare the shock heating rate with the viscous and the irradiation heating rates. This comparison is shown in Fig. 1, where we artificially damp σ exponentially with distance from the planet's gap opening region to avoid overestimating shock heating far from the planet:

$$\sigma(r) = \begin{cases} \sigma, & |r-r_{\rm p}| \le 2.5\,{\rm R_{Hill}}, \\ 1+(\sigma-1)e^{-4(|r-r_{\rm p}|-2.5\,{\rm R_{Hill}})/r_{\rm p}}, & \text{otherwise}, \end{cases} \tag{8}$$

where ${\rm R_{Hill}} \equiv \sqrt[3]{M_{\rm p}/3M_*}\, r_{\rm p}$.

For $\mathcal{M}_1^2 \to \infty$, Eq. (1) gives an upper limit to shock strength for adiabatic shocks as $\sigma \to \frac{\gamma+1}{\gamma-1} = 6$ for $\gamma = 7/5$. We should note that this upper limit is not strict if additional thermal mechanisms (such as cooling) are included in the models.

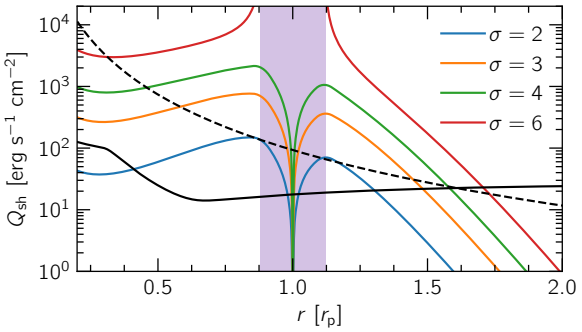

Fig. 1. Shock heating rate by planet's spirals estimated by method in Sect. 2 and damped using Eq. (8) for different shock strengths ($\sigma = \Sigma_2/\Sigma_1$, $\gamma = 7/5$) compared to viscous and irradiative heating rates (dashed and solid black lines; see Sect. 3.1). Lilac band indicates the corotating region, in which the estimates are not valid due to potential gap opening. The model used for this plot assumes $M_p = 100\,M_\oplus$, $\dot{M} = 10^{-8}\,M_\odot\,\mathrm{yr}^{-1}$, $\alpha = 10^{-3}$, $r_p = 4\,\mathrm{au}$ (see Sects. 3.1 and 3.2).

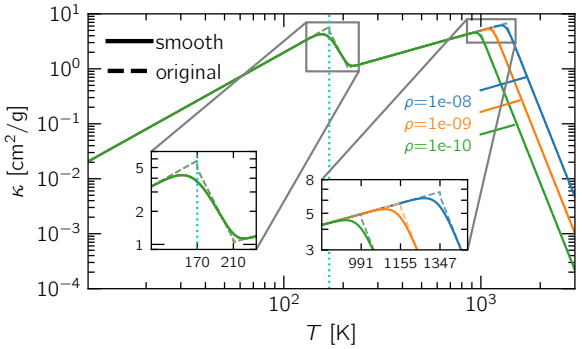

Fig. 2. First four opacity regimes according to Lin & Papaloizou (1985). Dotted teal line marks the water ice line ($T_{ice} = 170\,\mathrm{K}$). Different branches are patched together by interpolation.

This estimate shows that shock heating by a planet can overcome the other two heating sources if the planet is massive enough to produce strong shocks, and the disk opacity is large enough to prevent heat from quickly escaping from the midplane. This extra heating raises the temperature in the disk up to the location where the shocks damp greatly. Because the disk opacity also depends on temperature (see Fig. 2), spiral heating by a planet in the vicinity of an ice line (either via migration or in-situ formation) might displace the latter. In the following sections, we study this problem for a more realistic model with shocks that are not necessarily adiabatic and examine how much and under what conditions the location and shape of an ice line can change.

3. Model setup

In this section, we present the physical framework that we utilize in our planet–disk modeling. We describe the relevant equations and the assumptions behind them, along with our numerical setup as far as our parameter space, initial and boundary conditions, and grid structures are concerned.

3.1. Physics

We solve the vertically integrated Navier–Stokes equations for a disk with surface density Σ, velocity vector $\boldsymbol{v}$ and vertically integrated specific internal energy e on a polar coordinate system $\{r, \phi\}$ centered around the star. For a perfect gas, the equations read

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = -\Sigma\nabla\cdot\boldsymbol{v},$$
$$\Sigma\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\nabla p - \Sigma\nabla\Phi + \nabla\cdot\overline{\boldsymbol{\Sigma}},$$
$$\frac{\mathrm{d}\Sigma e}{\mathrm{d}t} = -\gamma\Sigma e\nabla\cdot\boldsymbol{v} + Q_{\mathrm{visc}} + Q_{\mathrm{irr}} - Q_{\mathrm{cool}}, \tag{9}$$

where $\gamma = 7/5$ is the adiabatic index, $p = (\gamma - 1)\Sigma e$ is the vertically integrated pressure and $\overline{\boldsymbol{\Sigma}}$ denotes the viscous stress tensor.

The adiabatic and isothermal sound speeds c_s, c_{siso} are related as:

$$c_{\mathrm{siso}} = c_s/\sqrt{\gamma} = H\Omega_K = hv_K = \sqrt{RT/\mu}, \tag{10}$$

where H is the pressure scale height, $h = H/r$ is the aspect ratio and $v_K = \Omega_K r = \sqrt{GM/r}$ is the Keplerian azimuthal velocity. The universal gas constant and mean molecular weight are denoted by R and $\mu = 2.353$, respectively.

Source terms Q_{visc}, Q_{irr}, and Q_{cool} in the energy equation correspond to viscous heating, stellar irradiation, and thermal cooling, respectively:

$$Q_{\mathrm{visc}} = \frac{1}{2\nu\Sigma}\left(\sigma_{rr}^2 + 2\sigma_{r\phi}^2 + \sigma_{\phi\phi}^2\right) + \frac{2\nu\Sigma}{9}(\nabla\cdot\boldsymbol{v})^2,$$
$$Q_{\mathrm{irr}} = 2\frac{L_*}{4\pi r^2}(1-\epsilon)\left(\frac{\mathrm{dlog}\,H}{\mathrm{dlog}\,r} - 1\right)h\frac{1}{\tau_{\mathrm{eff}}},$$
$$Q_{\mathrm{cool}} = 2\sigma_{\mathrm{SB}}\frac{T^4}{\tau_{\mathrm{eff}}}, \tag{11}$$

where $\nu = \alpha c_s H$ is the kinematic viscosity according to the α-viscosity model of Shakura & Sunyaev (1973), σ_{SB} is the Stefan–Boltzmann constant, and τ_{eff} is an effective optical depth following Hubeny (1990):

$$\tau_{\mathrm{eff}} = \frac{3\tau}{8} + \frac{\sqrt{3}}{4} + \frac{1}{4\tau}, \qquad \tau = \int_0^\infty \kappa\rho\mathrm{d}z \approx c_1\kappa\rho_{\mathrm{mid}}H, \tag{12}$$

with the Rosseland mean opacity $\kappa(\rho, T)$ defined according to Lin & Papaloizou (1985), shown in Fig. 2. The correction factor $c_1 = 1/2$ is added following Müller & Kley (2012) to account for the drop in opacity with height. We assume a Gaussian vertical density profile so that $\Sigma = \sqrt{2\pi}\rho_{\mathrm{mid}}H$.

As far as irradiation is concerned, we assume a star of solar luminosity $L_* = L_\odot$ and a disk albedo of $\epsilon = 1/2$. Following Menou & Goodman (2004), the factor $\frac{\mathrm{dlog}\,H}{\mathrm{dlog}\,r}$ is assumed to be constant and equal to $9/7$ (i.e., disk self-shadowing is not considered).

3.2. Numerics

We utilized the numerical MHD code PLUTO (Mignone et al. 2007) for our simulations, along with the FARGO algorithm described by Masset (2000) and implemented as a library into PLUTO by Mignone et al. (2012). To enable radiative diffusion, we implemented a separate module that is briefly described in Appendix D. Simulations with an embedded planet were run on a polar $\{r, \phi\}$ grid, logarithmically spaced in the radial direction.

Our parameter space is shown in Table 1. It contains the planet mass M_p, the planet's semimajor axis r_p (fixed, circular

A&A 633, A29 (2020)

Table 1. Parameter space: r_p is quoted in au and $\dot{M}$ in $M_\odot\,\mathrm{yr}^{-1}$.

Parameter	Values
M_p	$10\,M_\oplus$, $100\,M_\oplus$, $1\,M_J$, $3\,M_J$
r_p	1, 4, 10
$\dot{M}$	10^{-9}, 10^{-8}, 10^{-7}
α	10^{-4}, 10^{-3}, 10^{-2}

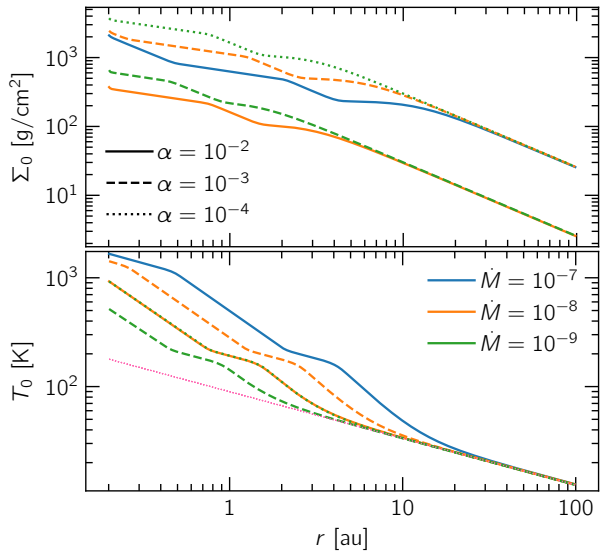

Fig. 3. Initial profiles for five disk models with different $\dot{M}$ and α used throughout this study. Dotted pink line refers to an inviscid disk where $Q_{\rm cool} = Q_{\rm irr}$ (i.e., the irradiation temperature) and functions as our effective temperature floor. It becomes clear that viscous heating is strongest in the inner disk, while irradiation dominates its outer parts.

orbits), the viscosity parameter α and the initial disk mass accretion rate $\dot{M}$, which is constant throughout the disk in viscous equilibrium such that $\dot{M} = 3\pi\nu\Sigma$ (Lodato 2007). By selecting an α value and a constant accretion rate we can then construct well-defined equilibrium states.

To generate our initial conditions, we prepared 1D models that satisfy viscous and thermal equilibrium conditions:

$$\begin{aligned} \dot{M} &= 3\pi\nu\Sigma \qquad \text{(viscous equilibrium)},\\ Q_{\rm cool} &= Q_{\rm visc} + Q_{\rm irr} \qquad \text{(thermal equilibrium)}, \end{aligned} \tag{13}$$

and ruled out very cold disks or gravitationally unstable ones, for which the Toomre parameter Q_T (Toomre 1964), defined as:

$$Q_T \equiv \frac{c_s \Omega_K}{\pi G \Sigma}, \tag{14}$$

is less than unity. The initial profiles we used are plotted in Fig. 3.

We then embedded planets in each configuration and ran each model until the disk reached, roughly, its viscous and thermal equilibrium or a maximum simulated time of $t_{\rm max} = 10^5$ years elapses. To ensure a constant $\dot{M}$ through the boundaries, Σ, $\boldsymbol{v}$ were damped to the initial profiles according to de Val-Borro et al. (2006) over a timescale of 0.3 boundary orbital periods.

Table 2. Grid setup.

$\log\left(\frac{\dot{M}}{M_\odot/\mathrm{yr}}\right)$	$\log\alpha$	r_p [au]	$M_{\rm disk}\left[\frac{M_*}{100}\right]$	N_r	N_ϕ
−7	−2	1	2.58	435	849
−7	−2	4	2.57	383	748
−7	−2	10	8.12	441	861
−8	−2	1	0.08	699	1364
−8	−2	4	0.38	651	1271
−8	−2	10	0.94	527	1029
−8	−3	1	0.49	579	1130
−8	**−3**	**4**	**3.40**	**531**	**1037**
−8	−3	10	9.10	515	1005
−9	−3	1	0.10	813	1587
−9	−3	4	0.40	685	1337
−9	−3	10	0.96	529	1033
−9	−4	1	0.77	699	1364
−9	−4	4	3.84	655	1279
−9	−4	10	9.51	529	1033

Notes. Our fiducial model is shown in bold, and the same setup was used for runs with radiative diffusion.

The gravitational forces read

$$\begin{aligned} \boldsymbol{g}_{\rm grav} &= \boldsymbol{g}_* + \boldsymbol{g}_p + \boldsymbol{g}_{\rm in} = -\nabla\Phi \\ &= -\frac{GM_*}{r^3}\boldsymbol{r} - \frac{GM_p}{(r_e^2+\epsilon^2)^{3/2}}\boldsymbol{r}_e - \frac{GM_p}{r_p^3}\boldsymbol{r}_p, \quad \boldsymbol{r}_e = \boldsymbol{r} - \boldsymbol{r}_p, \end{aligned} \tag{15}$$

where $\boldsymbol{g}_*, \boldsymbol{g}_p, \boldsymbol{g}_{\rm in}$ refer to the gravitational acceleration by the star, the planet, and the indirect term that arises due to the star–planet system orbiting around their mutual barycenter. The planet is on a fixed orbit and we neglected the backreaction of the disk onto star and planet. For the softening length, we used $\epsilon = 0.6H$ to prevent singularities around the planet's location. The value 0.6 was selected according to Müller et al. (2012) as it provides very similar results to 3D models. We note that ϵ was evaluated using the local H at each cell.

4. Disk profiles

Having described our physics and numerical methods, we proceeded to execute our simulations. The grid setup for each model is shown in Table 2. A cross-code comparison as well as a resolution test for the verification of our numerical setup is provided in Appendix A. We constructed the numerical grid such that the pressure scale height H is resolved by at least six grid cells at the planet's location (see Appendix B for more details).

We first investigated the thermal input of a planet onto the disk and the structure of the gap that planet might possibly carve. In order to do so, we first presented some comparisons of the gap width and depth across different models for both surface density and temperature. We then highlighted the influence of disk aspect ratio and viscosity on the planet's ability to open a gap. We then determined to show temperature profiles for various disks with identical initial temperatures so that the planet's impact would become more apparent.

Afterwards, we took a closer look at the structure of spiral arms by tracking the same quantities along their crests and compared them against the azimuthally averaged disk profiles. This would give us an estimation for both the temperature contrast between the spirals and the disk, as well as the shock strength along those spirals.

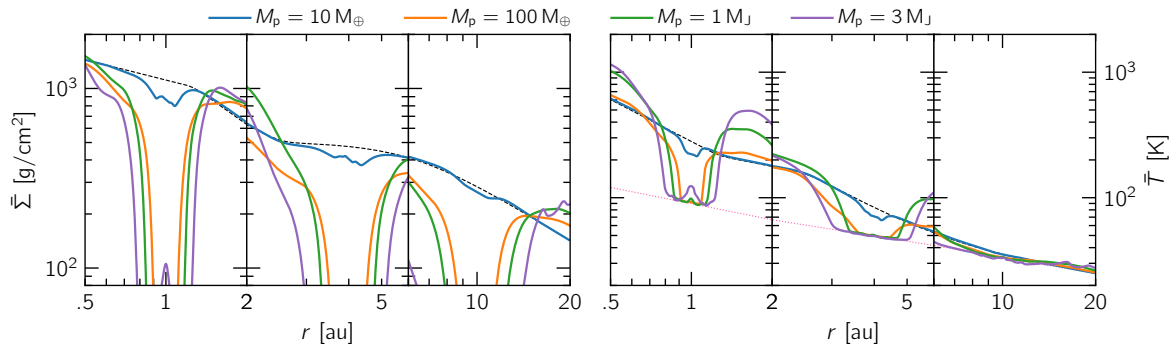

Fig. 4. Azimuthally-averaged profiles of surface density (*left*) and temperature (*right*) across planet masses and locations for our models. More massive planets open deeper and wider gaps but the temperature inside the gap region is not necessarily lower in the outer, irradiation-dominated disk due to stellar irradiation. The dotted pink line marks the disk irradiation temperature ($Q_{\rm cool} = Q_{\rm irr}$).

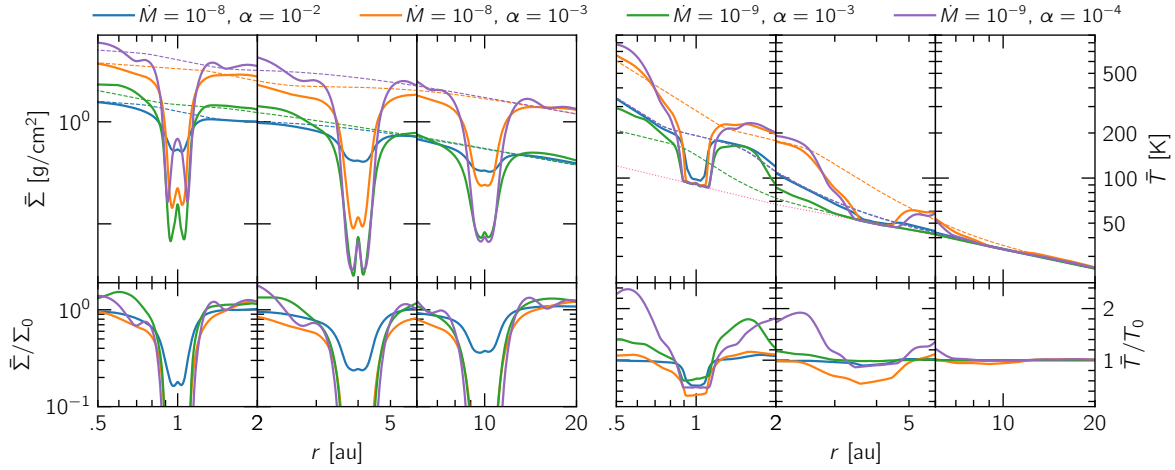

Fig. 5. Azimuthally-averaged profiles of surface density (*left*) and temperature (*right*) around planet's location across different disk models for $M_p = 100\,M_\oplus$. These snapshots are taken once each model has reached a quasi-equilibrium state, meaning that the gap depth is not well-defined for very low viscosities. However, it only takes a few hundred orbits for the overall disk structure to equilibrate.

4.1. Gap opening capabilities of a planet

In Fig. 4, we compare the gap-opening capabilities of planets of different masses for our fiducial model ($\dot{M} = 10^{-8}$, $\alpha = 10^{-3}$). While the least massive planet in these models (10 $M_\oplus$) does not open a gap, the rest are sufficiently massive to show a clear trend between planet mass and gap width, with more massive planets opening deeper and wider gaps.

However, we also find that there is a lower limit to the temperatures inside the gap. This arises due to stellar irradiation, which provides enough heat to form an effective temperature floor where $Q_{\rm irr} = Q_{\rm cool}$. This term overpowers other heating effects with increasing radii and, as a result, a temperature gap is not visible in the outer disk regardless of planet mass.

Then, for a given planet mass of $M_p = 100\,M_\oplus$, we carried out the same comparison across models with different disk parameters. The results are shown on Fig. 5, where a similar behavior is visible for temperatures inside the gap.

A key point in these findings is that we observed shallower gaps for higher values of α (for a given $\dot{M}$) or $\dot{M}$ (for a given α). This can be understood by looking at the two main mechanisms determining the gap edge, as shown by Crida et al. (2006): viscosity and pressure gradients. Before adding a planet to a disk of a given $\dot{M}$ and α, one can show that $\dot{M} \propto \nu\Sigma \propto \alpha p r^{3/2}$, such that pressure gradients are stronger in disks with a higher $\dot{M}$ or lower α. This, combined with the fact that viscous dissipation is scaled with α, allows for easier gap opening in disks with either a higher $\dot{M}$ or lower α. This is nothing new, as it has been pointed out by Crida et al. (2006), Zhang et al. (2018) as well as in previous studies, that disks with a lower viscosity or aspect ratio support the gap opening process.

4.2. Spiral shock heating by a planet

On the right panels of Figs. 4 and 5, apart from the depth (or lack thereof) of a gap, we noted a temperature increase on both sides of the planet's vicinity, sometimes by a factor of 1.5–2 compared to the initial profiles. This heat excess is scaled with planet mass and it can lead to quite high temperatures in the inner disk. This pattern is in agreement with our theoretical estimates, given in Sect. 2, based on which we should expect that the optically thick inner disk is more susceptible to heating by spiral shocks.

In the attempt to compare the individual effect of each of our four parameters – planet mass, accretion rate, viscosity parameter, and planet radius – we plotted pairs of models where three out of four parameters are the same, allowing us to quantify the influence of the fourth. The trends we find with this method are clear enough that a comparison between our fiducial model

A&A 633, A29 (2020)

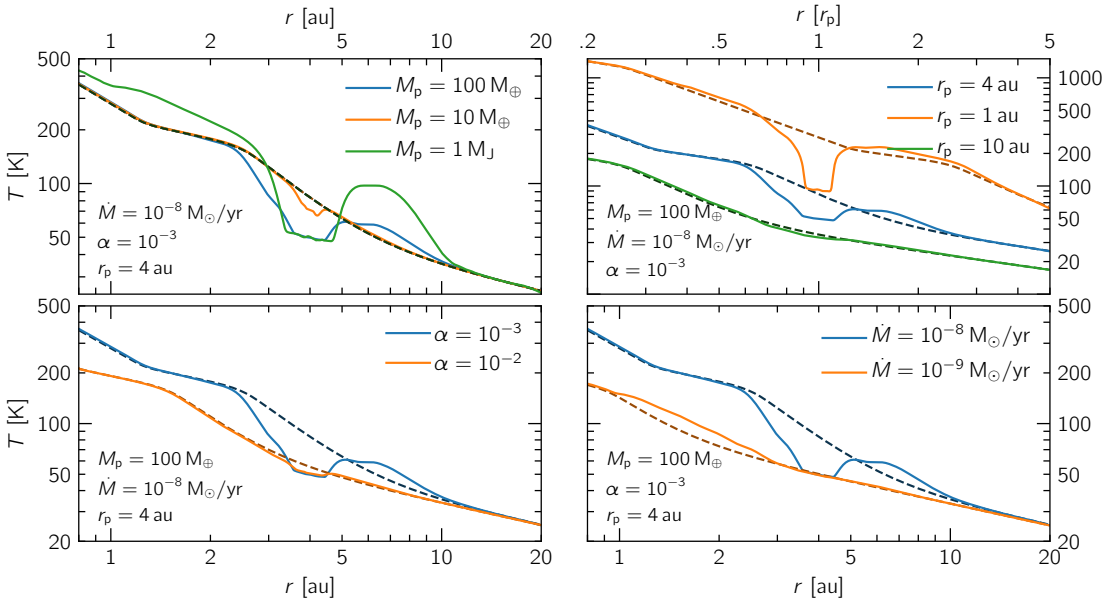

Fig. 6. Azimuthally-averaged temperature profiles for four pairs of models, showcasing influence of each of four variables in our parameter space (M_p, $\dot{M}$, α, r_p) on spiral shock heating. Our fiducial model ($M_\mathrm{p} = 100\,M_\oplus$, $\dot{M} = 10^{-8}, \alpha = 10^{-3}, r_\mathrm{p} = 4$) is shown in blue on every panel, while dashed lines depict initial profiles of the model of corresponding color. In all cases, we noted stronger shock heating in optically thicker regions of the disks. This behavior can also be seen in the rest of our suite of models.

and four other models suffices to convey the general picture. This comparison is shown in Fig. 6.

The common denominator for these four panels is the cooling timescale of different disks and the regions within them. We can get a rough estimate of this timescale by focusing on the cooling term in Eq. (11) and writing

$$\frac{\partial \Sigma e}{\partial t} \sim \frac{\Sigma e}{\tau_\mathrm{cool}} \sim Q_\mathrm{cool} = \sigma_\mathrm{SB}\frac{T^4}{\tau_\mathrm{eff}} \Rightarrow \tau_\mathrm{cool} \approx \frac{\tau_\mathrm{eff} R\Sigma}{\mu(\gamma-1)\sigma_\mathrm{SB}T^3}, \quad (16)$$

which further backs the assumption that the deciding factor in determining the contribution of shock heating to the thermal budget of the disk is the optical depth. Finally, we compared two models where $\{\dot{M} = 10^{-8}, \alpha = 10^{-2}\}$ and $\{\dot{M} = 10^{-9}, \alpha = 10^{-4}\}$, respectively. These two models happen to have identical initial temperature profiles but they show the lowest and highest optical depths in our suite of simulations, respectively. This comparison is plotted in Fig. 7 and clearly shows the effect of optical depth on the contribution of shock heating. It should be noted that even though the optically thinnest model in that figure shows only small traces of excess heat due to shocks, the cooling timescale is still more than 10% of the orbital period at 10 au and, therefore, radiative effects of the disk should still be treated self-consistently to get a correct picture of its evolution.

From our results, we can conclude that shock heating is, in principle, important for all of our models and that it sometimes dominates when the cooling timescale of the disk is sufficiently long. This implies that planets with semimajor axes in the range of 1–10 au can noticeably heat up their environment through spiral shocks and, as a result, an adiabatic equation of state is necessary when modeling planet–disk interaction in this regime.

4.3. Spiral arm structure and shock strength

In the previous section, we discuss the effect of shock heating by comparing azimuthally-averaged profiles in simulations with

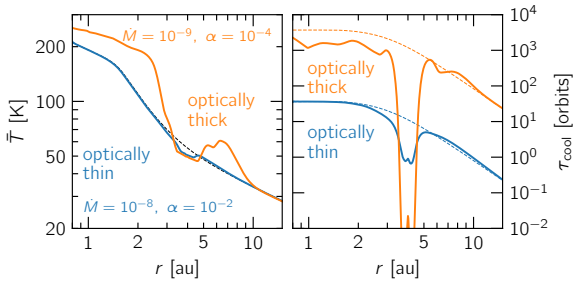

Fig. 7. Azimuthally-averaged temperature and cooling timescale for two models that share the same initial temperature profile but have very different optical depth profiles. We see that shock heating is significantly stronger for the optically thicker model, while the optically thinner one looks almost unchanged.

and without planets. While this is a useful approximation to form a general image of planet–disk interaction, it cannot isolate the contribution of individual spirals or their properties. Inspired by the approach of Zhu et al. (2015) we wrote a script that can trace spiral arms as they propagate away from the planet and logged their coordinates, as well as Σ_arm and T_arm along their crests. We then used this data to estimate a proxy for the shock strength along those spirals as $\Sigma_\mathrm{arm}/\bar{\Sigma}$ (shown in Sect. 2), as well as their pitch angles β defined as $\tan\beta = \mathrm{d}\log r_\mathrm{arm}/\mathrm{d}\phi_\mathrm{arm}$.

As in the previous section, trends among models are clear enough such that we do not need to present results for our entire library of simulations. Instead, we take into account that the contribution of shock heating is scaled with the optical depth and show results for three regimes: the optically thinnest model ($\dot{M} = 10^{-8}$, $\alpha = 10^{-2}$, $r_\mathrm{p} = 10$), the optically thickest one ($\dot{M} = 10^{-9}$, $\alpha = 10^{-4}$, $r_\mathrm{p} = 1$), as well as our fiducial model ($\dot{M} = 10^{-8}$, $\alpha = 10^{-3}$, $r_\mathrm{p} = 4$), which also happens to

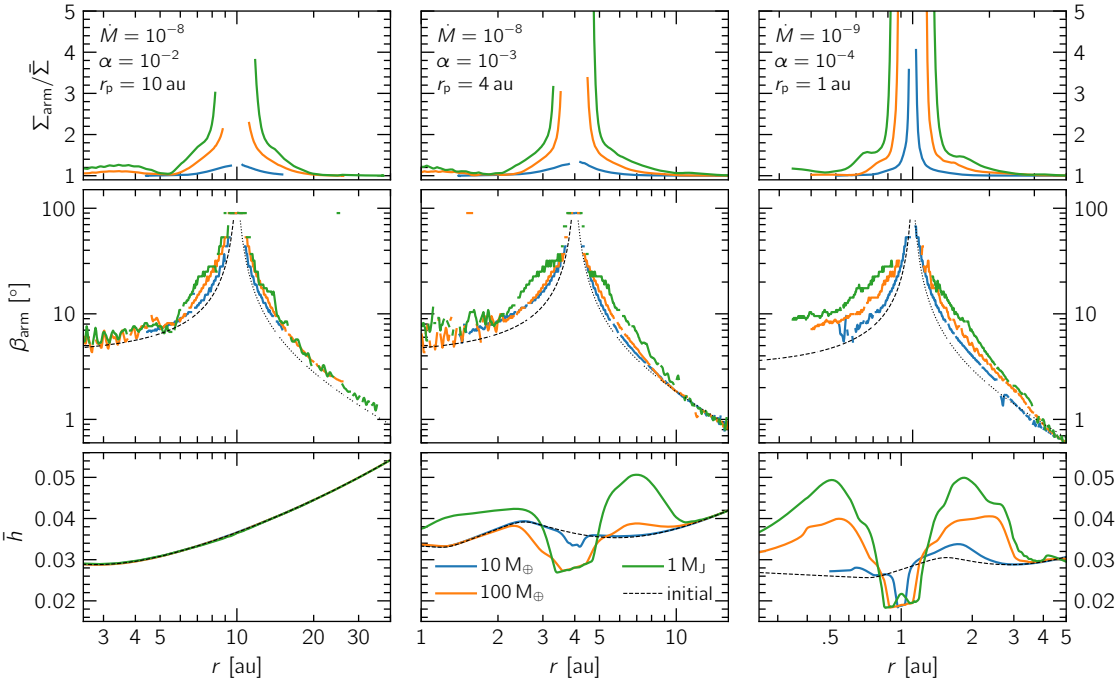

Fig. 8. Shock strength, pitch angles, and azimuthally-averaged aspect ratios for three representative models. Optical depth and cooling timescale increase from left to right. *Top*: shocks are stronger for more massive planets but their dependence on disk parameters is not clear. To filter out unphysical results, we mask points that lie within the gap region ($|r - r_p| \leq 2.5\,R_{Hill}$) or the corotating region (see Eq. (17)). *Middle*: pitch angles are roughly the same regardless of planet mass for the optically thin case but deviate with increasing optical depth. Black lines denote the expected values using Eq. (19) for $h = h_p$ (dashed) or Eq. (20) for $h = h_p r^{2/7}$ (dotted). *Bottom*: a power law fit of aspect ratio is only possible for optically thin case or low mass planets but fidelity for such a fit breaks down even for 10-Earth-mass planets in an optically thick disk. Black dashed lines refer to initial aspect ratio profiles.

lie somewhere in the middle. For each model, we calculate the shock strength and pitch angles of primary spirals (i.e., those that connect to the planet). Since the pitch angle is scaled with h far from the launching point according to linear theory, we also plotted the azimuthally-averaged aspect ratio $\bar{h}$.

In an attempt to filter out unphysical shock strength values inside the low-density ring around the planet, we calculated the half-width of the horseshoe region as shown in Paardekooper et al. (2010):

$$x_h = 1.1 r_p \sqrt{\frac{1}{h_p}\frac{M_p}{M_*}} \left(\frac{0.4 H_p}{\gamma \epsilon_p}\right)^{1/4}, \tag{17}$$

as well as the shock length following Dong et al. (2011):

$$x_s = 0.8 \left(\frac{\gamma+1}{12/5}\frac{M_p}{M_{th}}\right)^{-2/5} H, \quad M_{th} \equiv \frac{c_s^3}{G\Omega_p}, \tag{18}$$

where $M_{th} \approx 1\,M_J \left(\frac{h_p}{0.1}\right)^3 \left(\frac{M_*}{M_\odot}\right)$ is the disk thermal mass. If $M_p > M_{th}$, then $x_s = 0$ (spirals shock immediately upon launch). For more massive planets, where a gap opens, we set a cut-off where $|r - r_p| \leq 2.5\,R_{Hill}$. We then excluded data from within any of those three regions.

Our results are summarized in Fig. 8. We see that the shock strength of spirals typically lies between 1.5–3 for massive planets but rarely exceeds 1.5 for 10 $M_\oplus$ models. Looking at Fig. 1, we can see that such shocks produce competitive heating when compared to either viscosity or stellar irradiation for $r_p \leq 4$. In the $r_p = 10$ case (see the left panels of Fig. 8), however, the planet is embedded in an optically thin, irradiation-dominated region and, as a result, shock heating is overcome by stellar irradiation, which eventually sets the overall profile.

As far as pitch angles are concerned, we attempted to fit their curves with analytical formulas that assume an aspect ratio profile. In the inner disk, due to the temperature being defined by different power laws depending on the opacity regime, it is easier to assume that the aspect ratio is roughly constant and use the formula by Ogilvie & Lubow (2002) to calculate the location of spirals

$$\phi_{arm} = -\mathrm{sgn}(r - r_p)\frac{2}{3h}\left[\left(\frac{r}{r_p}\right)^{3/2} - \frac{3}{2}\ln\left(\frac{r}{r_p}\right) - 1\right]. \tag{19}$$

On the other hand, for the irradiation-dominated outer disk (where $Q_{irr} \cong Q_{cool}$), we can utilize the formula by Muto et al. (2012):

$$\begin{aligned}\phi_{arm} = &-\frac{\mathrm{sgn}(r - r_p)}{h_p} \\ &\times\left[\left(\frac{r}{r_p}\right)^{1+\eta}\left\{\frac{1}{1+\eta} - \frac{1}{1-\zeta+\eta}\left(\frac{r}{r_p}\right)^{-\zeta}\right\}\right. \\ &\left.-\left(\frac{1}{1+\eta} - \frac{1}{1-\zeta+\eta}\right)\right],\end{aligned} \tag{20}$$

where $\Omega_K \propto r^{-\zeta}$ and $h \propto r^{0.5-\eta}$. For $h \propto r^{2/7}$ (see Eq. (11)), we have $\zeta = 3/2$ and $\eta = 3/14$.

A&A 633, A29 (2020)

Even when utilizing both formulas, we find that it is difficult to get a good fit and that we always underestimate the pitch angles. This shows that the heating generated by the planets' spirals can change the disk aspect ratio such that it cannot be accurately approximated with a power law. The fit completely breaks down for massive planets or optically thick disks, where we find that pitch angles are inflated around the location of planets. This makes sense since shock heating peaks at these locations (e.g., see Figs. 1 and 5).

5. Location and shape of the water ice line

Using the analytical estimates in Sect. 2, we show that if a massive planet is located in the optically thick part of the disk, the shocks are adiabatic and spiral heating can increase the temperature of the disk. In Sect. 4, we present our results on the heating potential of these shocks through our numerical simulations, confirming our estimates. In this section, we investigate how much a planet can displace ice lines (e.g., the water ice line) in a disk. Our motivation to do so lies in quantifying the possibility that a planet can starve itself or the inner disk of water as it forms, as well as the possibility that it can impact the environment in which planetesimals could grow (Drążkowska & Alibert 2017).

From Fig. 2 it becomes clear that the Rosseland mean opacity is independent of density up to $T \approx 10^3$ K such that the water ice line effectively represents the temperature where ice sublimates. This implies that the location of the ice line does not explicitly depend on the density jump across shocks but instead on the temperature they can reach. In this study, we follow Lin & Papaloizou (1985) and define the water ice line $r_{\rm ice}$ as the point where the temperature reaches $T_{\rm ice} = 170$ K (see opacity transition at this point in Fig. 2).

5.1. Location of the azimuthally-averaged water ice line

In the absence of a planet and assuming an axisymmetric disk, the equation $r = r_{\rm ice}$ defines a circle with radius $r_{\rm ice}$ from the star, within which water can only be found in the form of vapor. For now, let us assume that the presence of the planet does not significantly perturb the ice line in the azimuthal direction but instead moves it uniformly towards or away from the star, such that the new location of the ice line is $\bar{r}_{\rm ice} = \langle r_{\rm ice} \rangle_\phi$ once equilibrium is reached.

Depending on its initial location, the ice line is susceptible to two planet-induced phenomena: the closer it lies next to the planet, the more it is exposed to shock heating from the latter, leading to a larger outward displacement as the planet heats up its surroundings. In the extreme case that the ice line initially lies directly next to the planet, it can be pushed away by a factor of 1.5 or even more (see Fig. 9a). The impact on the location of the ice line is scaled with the mass of the planet, with Jupiter-sized planets increasing $\bar{r}_{\rm ice}$ by a factor of 2.

However, in the case that the planet is massive enough to open a gap, embedding it too close to the ice line such that the latter overlaps with the optically thin gap region results in a recession of the ice line towards the inner gap edge (see Fig. 9b). In this case, the final location of the ice line depends on the width of the gap, which is also scaled with planet mass. Of course, for planets of sufficiently low mass, a gap will not open and, therefore, the ice line location will essentially remain intact.

Finally, it is possible that gap opening and shock heating can compete for the determination of the location of the ice line. This behavior is shown on Fig. 9c: shock heating initially moves the ice line outwards, "pulling" it closer to the planet. However, the slower gap opening process eventually "catches up" and the steep temperature gradient near the inner gap edge extends the region affected by the gap down to around 2.5 au ($0.6\,r_{\rm p}$), returning the ice line to a location similar to its initial one once the gap has fully opened.

5.2. Azimuthal structure of the ice line

In the previous section, we consider the ice line as an axisymmetric line – a circle with radius $\bar{r}_{\rm ice}$ around the star – by tracking its location using azimuthally-averaged temperature profiles. However, shock heating is a strongly non-axisymmetric process as it follows the trajectories of spiral arms. Therefore, its influence on the location of the ice line should introduce azimuthal features on the latter. In other words, the full picture is 2D – at least within the scope of this project.

To examine the azimuthal structure of the ice line, we plot temperature maps of our models at equilibrium and draw contours at $T(r, \phi) = T_{\rm ice}$. We summarize the results for some of our representative models in Fig. 10.

In general, a few key behaviors can be observed in our simulation results. The ice line tends to move outwards in optically thick disks (as shown above) and deform it in such a way that it follows the trajectories of spirals when strong shocks are present. Therefore, we can distinguish a few distinct "extremes." For a low-mass planet in an optically thin disk, the location or shape of the ice line do not change. In an optically thick disk, the same planet might slightly move the ice line outwards or perturb it along the azimuthal axis.

In both of these cases, the analysis in Sect. 5.1 still applies with good accuracy. However, a massive planet launches strong shocks and opens a gap which can halt the outward movement of the ice line or even cause it to recede to the inner gap edge. Therefore, for a high-mass planet in an optically thin disk (where typically $r_{\rm ice}^{t=0} < r_{\rm p}$), we see hot spirals form in the inner disk (such that $T_{\rm arm} > T_{\rm ice}$) but we see little to no radial displacement of the ice line (see Fig. 10b). If initially $r_{\rm ice}^{t=0} \sim r_{\rm p}$, the ice line will recede to the inner gap edge, in addition to forming hot spirals in the outer disk (see Fig. 10c).

On the other hand, for a massive planet in an optically thick disk, shock heating is strong enough to displace the ice line to the outer disk to the point where spiral pitch angles are small and the spirals are very tightly wound, heating the disk uniformly in azimuth. In this case, the domain is split into a hot inner disk, a cold gap, and a hot ring in the outer disk (see Fig. 10a). If initially $r_{\rm ice}^{t=0} \sim r_{\rm p}$, the ice line will again recede to the inner gap edge while possibly forming hot spirals or a hot ring in the outer disk, depending on the optical depth.

Of course, if $r_{\rm ice}^{t=0} \gg r_{\rm p}$, far out at the irradiation-dominated outer disk, the ice line will not change in shape or location but a cold ring can still form inside the gap region. However, the optical depth rapidly increases at small distances from the star and, as such, the pile-up of inner spiral arms by a Jupiter-sized planet can still cause substantial heating, moving the ice line outwards even in optically thinner models (as shown in Fig. 9d).

6. Discussion

In this section we discuss our findings with respect to their possible impact on the growth and change of orbital elements of the planets, and on the structure of the disk.

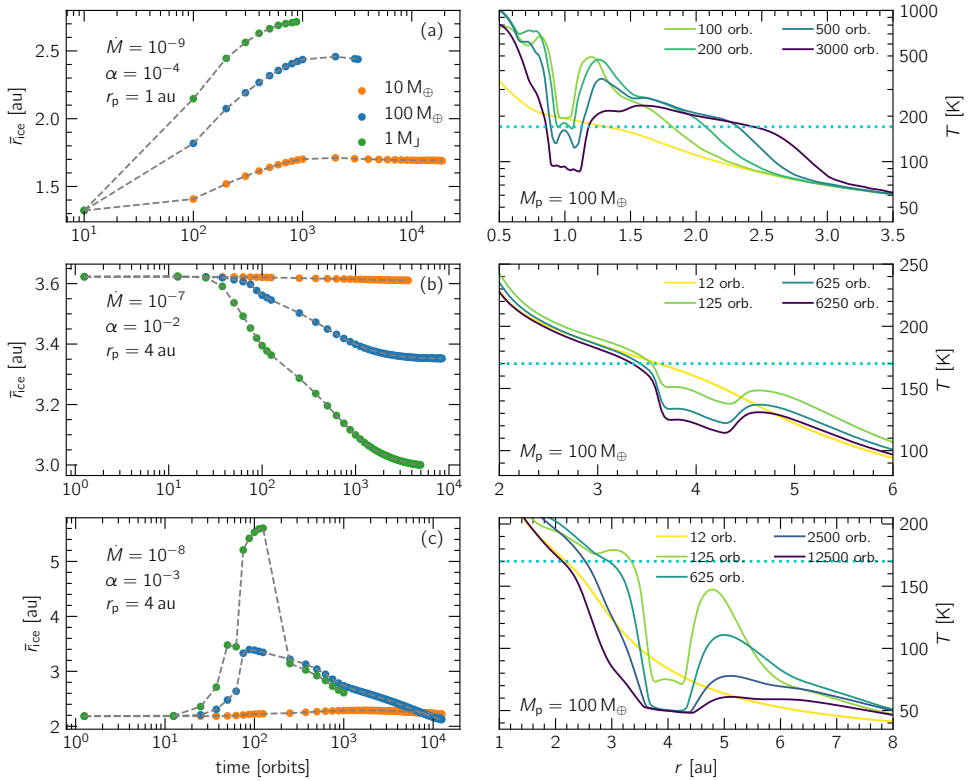

Fig. 9. Azimuthally-averaged ice line location for three sample models, showcasing three possible evolution scenarios, *panel a*: spiral heating pushes ice line outwards; *panel b*: ice line recedes to the inner gap edge; *panel c*: shock heating initially pushes ice line outwards but eventually a gap is carved and ice line recedes inwards. These effects are amplified for more massive planets. It should be noted that in the case of a cold gap, ice can recondense within the gap region (*panel a*).

6.1. A shift of the ice line

Under certain conditions (low $\dot{M}$, low α), a planet located at 1 au could push the ice line outwards by a few au. This leads to a reduction of icy solid material present in the disk in terms of size and number. This lowers the accretion of solid material onto the planet as more matter is in gaseous form, which has a lower (viscous) drift than embedded particles. On the other hand, when the snow line moves just a little beyond the planet, icy aggregates that may fall apart can release tiny silicate grains (Schoonenberg et al. 2018) and tiny dust may be less well trapped in the outer edge of the planet gap, possibly affecting the accretion of dust onto the planet. The net effect will be an enhancement of dry over wet particle accretion onto the planet and a reduction of the water content in the inner regions of the disk.

6.2. Slush islands

In our simulations, we locate regions where the conditions within the disk are such that the temperature along the spirals is above the ice sublimation threshold and drops below it between spiral crests (see Fig. 10). Ice sublimates around the peak of the shock but condenses again further away from it, such that along the boundary of the spiral (as defined by the ice line) one might find a mixture of ice and water vapor with a "slushy" consistency, hence, we call them "slush islands." These may occur inside as well as outside of the planet's location. The repetitive sublimation and condensation will slow down dust growth as growing particles will periodically be reduced in size.

6.3. Migration torques

The disk heating of an embedded planet will change the torques acting on it and, hence, affect its migration rate. As the torques are scaled inversely to the disk's scale height (Kley & Nelson 2012), it is expected that the planets slow down due to the heating they produce. This possibility has already been explored and discussed by Hallam & Paardekooper (2018), who show that even a simplistic prescription of gap edge illumination can result in the slowing down or even reversing of the migration rate of planets. In light of our results concerning the planets' ability to heat up their vicinity through spiral shocks, our modeling supports the findings of that study in that regard.

6.4. Temperature within the gap

In our simulations the temperature within the gap is lower that the environment because of the reduced optical depth. For deep gaps, the irradiation temperature was reached. Previous works look at the gas temperature in a planet's gap (via radiative transfer calculations) while considering the three dimensional vertical extent of the disk (Jang-Condell & Turner 2012). In that particular study, the gap's temperature is determined by shadowing and illumination effects, which are not included in our 2D treatment of irradiation. It was also found that the outer gap edge, which is directly exposed to starlight, can heat the interior of the gap to the extent that the temperature can rise even higher than the ambient temperature.

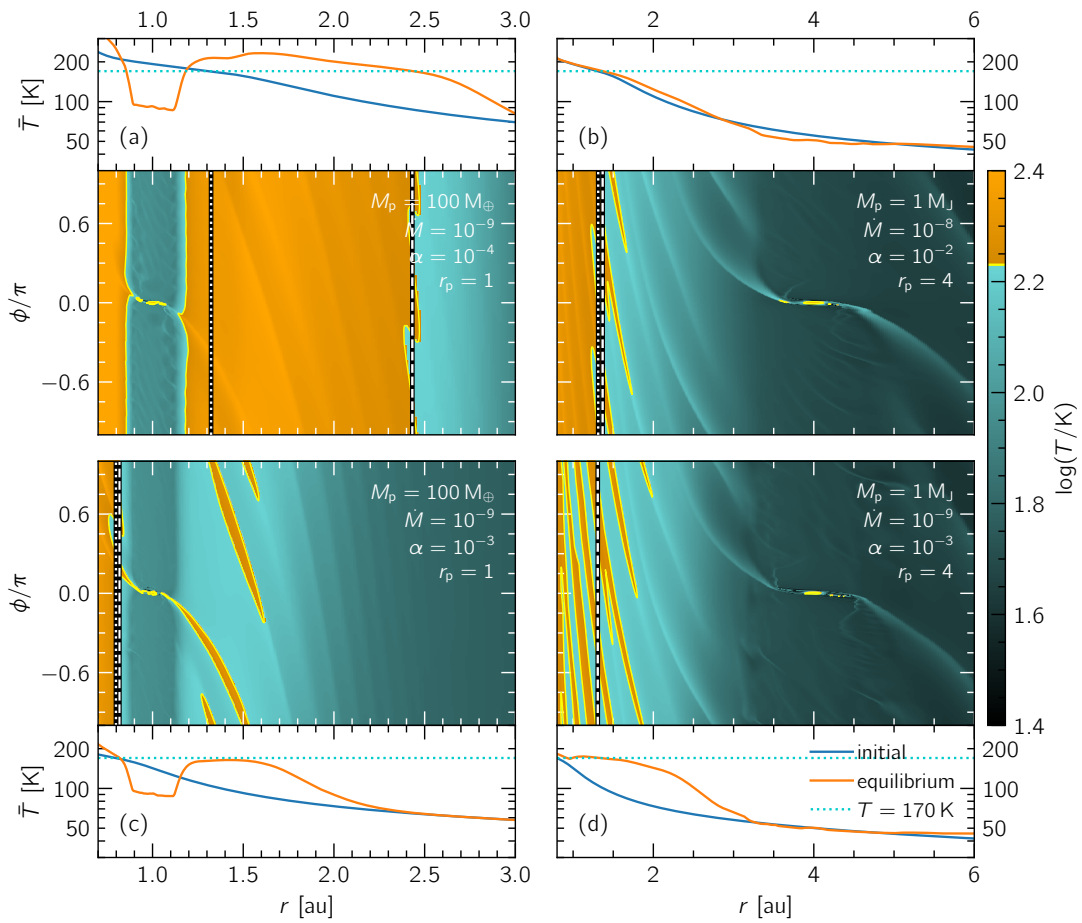

Fig. 10. Temperature maps for four models, showcasing azimuthal structure of ice line. Two colormaps denote the blue "cold" ($T < T_{ice}$) and orange "hot" ($T > T_{ice}$) disk, respectively, with a yellow line separating the two regions ($T = T_{ice}$). The initial location of the ice line (before a planet is embedded) is marked with a dotted white line, while its azimuthally-averaged location in equilibrium is shown with a dashed white line.

6.5. Simplifications and assumptions

Throughout our study, we have made several assumptions about the various physical processes at play. Therefore, we find that it is important for us to draw attention to the potential impact they can have on our results.

First and foremost, there is our two-dimensional approximation in simulating global, adiabatic disks. Lyra et al. (2016) point out that the additional degree of freedom in the vertical expansion of an adiabatic shock results in overall weaker shocks, suggesting that our results overestimate shock heating. This can be amended by "scaling" our results to refer to more massive planets.

Secondly, regarding the smoothing length chosen for the planet's gravitational potential, we chose to evaluate the scale height locally ($H(r, \phi)$) instead of using that at the planet's location (H_p). The reason behind this choice is that it corrects for the disk's finite thickness, as shown by Müller et al. (2012). However, this assumption might be problematic in radiative simulations. For example, the planet's accretion luminosity can result in a "hot bubble" around the planet (Klahr & Kley 2006), where the scale height can increase sharply with respect to its initial value. Nevertheless, this smoothing length becomes important at a scale far smaller than the planet's Hill radius and, therefore, it should have a minuscule effect on the latter's gravitational potential.

Additionally, our model of stellar irradiation contains a simplification in that disk self-shadowing is ignored. Specifically, we assume that the star illuminates a disk where the scale height does not significantly change (such that $\frac{\mathrm{d}\log H}{\mathrm{d}\log r}$ is constant and refers to a power-law profile for H) but then we point out that shock heating can, in fact, strongly affect said disk property. While this assumption leads to a very straightforward and stable numerical implementation of stellar irradiation, it occasionally results in a disparity between our assumption of $9/7$ for $\frac{\mathrm{d}\log H}{\mathrm{d}\log r}$ and the actual value. This disparity is greater for optically thick disks and diminishes with the increasing distance from the star. We can, therefore, justify our choice by remembering that stellar irradiation is indeed a dominant heat source at large radii, where the approximation holds best, and gives way to viscous and shock heating near the star, rendering it insignificant regardless of how well the approximation holds.

7. Conclusions

In our study, we examined the thermodynamical impact of planets on the ambient protoplanetary disk in which they are

embedded. In order to do so, we first calculated an estimate for the amount of heat a planet can deliver into the disk through spiral shocks and we showed that such heating can be significant. We noted that this process is strongest in the immediate vicinity of the planet but it also has the potential to influence a larger area depending on disk optical depth. We then carried out a grid of 2D numerical hydrodynamics simulations with included radiative effects in order to find out if and how much this planet-generated heating can influence the disk and displace or deform the otherwise axisymmetric water ice line, defined as the radius $r_{\rm ice}$, where $T(r_{\rm ice}) = T_{\rm ice} = 170\,{\rm K}$.

We find that spiral shock heating is most important in optically thick, viscosity-dominated disks. Both of these requirements suggest a long cooling timescale, and they are fulfilled in the inner few au of a protoplanetary disk. On the other hand, the irradiation-dominated outer disk suppresses shock heating by raising the aspect ratio. However, even when a planet is embedded in the outer disk, its inner spirals can heat up the disk as they propagate inwards.

We also show that both a high viscosity or aspect ratio inhibit the gap opening process, a result which is consistent with previous studies. On top of that, treating radiative effects allows us to probe the gas temperature inside the gap region. We find that in the inner disk, where viscous heating determines the gas temperature, a cold gap can be seen around massive planets. This is not visible in the outer disk, where the temperature both inside and out of the gap region is determined by the irradiation temperature.

By tracing the planet's spiral arms we find that planet-induced spiral shocks are scaled in strength with planet mass to the extent that shock heating is strongest for massive planets. This leads to a noticeable difference between spiral and background temperatures, with clear implications on the pitch angles of said spirals. We also show that due to the fact that the aspect ratio can increase dramatically by high-mass planets, fitting pitch angles with a standard flaring-aspect-ratio prescription would not, in principle, yield accurate results when shock heating is important.

Subsequently, we investigated the planet's ability to displace the water ice line. We found that shock heating by the planet can increase temperatures high enough to push the ice line away from the star. This outward displacement of the ice line can happen to various degrees depending on the optical depth of the disk. Optically thicker disks are unable to efficiently radiate away excess heat and are prone to larger ice line displacements. Shock heating can then lead to either a uniform outward movement or a non-axisymmetric deformation of the ice line.

In the inner few au of our disks, planets that are massive enough to carve a gap can create a cold ring around their semi-major axis. This gap cooling effect can easily overpower shock heating in the immediate vicinity of the planet, pulling the ice line inwards to the inner gap edge if it was initially near the soon-to-be-opened gap region or creating a "hot ring" outside of the planet's location if the ice line maintains a radius greater than that of the planet's semimajor axis.

However, it is also possible that the ice line becomes deformed due to the temperature contrast between spirals and the disk background to the extent that it bends to follow spiral trajectories. This deformation is clearest for strong shocks in optically thin disks, where the ice line can trace the inner or outer spirals depending on the initial disk temperature profile. Such spirals will then be water-poor, with possible implications on dust growth in their vicinity. These effects can impact the accretion rate and composition of accreted particles on a planet.

As far as the scale of this displacement and deformation is concerned, we find that planet mass plays a leading role in determining both shock strength as well as gap width to the extent that any effect related to the location or shape of the ice line is amplified for higher planet masses.

Finally, we report that accounting for radiative diffusion in the disk midplane leads to no significant differences in temperature profiles or ice line deformation. In addition, it leads to barely any observable differences in azimuthally-averaged ice line locations and spiral arm opening angles. As a result, radiative diffusion can be safely ignored in the context of this study.

Acknowledgements. The authors thank the referee for the constructive comments and suggestions for improving the quality of the manuscript. S.A. acknowledges support from the Swiss NCCR PlanetS. C.P.D. and W.K. acknowledge funding from the DFG research group FOR 2634 "Planet Formation Witnesses and Probes: Transition Disks" under grant DU 414/22-1 and KL 650/30-1. We acknowledge the support of the DFG priority program SPP 1992 "Exploring the Diversity of Extrasolar Planets" under grant KL 650/27-1. The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden–Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG. This research was supported by the Munich Institute for Astro- and Particle Physics (MIAPP) of the DFG cluster of excellence "Origin and Structure of the Universe". All plots in this paper were made with the Python library `matplotlib` (Hunter 2007).

References

Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, ApJ, 869, L41
Ataiee, S., Baruteau, C., Alibert, Y., & Benz, W. 2018, A&A, 615, A110
Bell, K. R., & Lin, D. N. C. 1994, ApJ, 427, 987
Bitsch, B., Raymond, S. N., & Izidoro, A. 2019, A&A, 624, A109
Crida, A., Morbidelli, A., & Masset, F. 2006, Icarus, 181, 587
de Val-Borro, M., Edgar, R. G., Artymowicz, P., et al. 2006, MNRAS, 370, 529
Dong, R., Rafikov, R. R., & Stone, J. M. 2011, ApJ, 741, 57
Drążkowska, J., & Alibert, Y. 2017, A&A, 608, A92
Goodman, J., & Rafikov, R. R. 2001, ApJ, 552, 793
Hallam, P. D., & Paardekooper, S. J. 2018, MNRAS, 481, 1667
Hubeny, I. 1990, ApJ, 351, 632
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Jang-Condell, H., & Turner, N. J. 2012, ApJ, 749, 153
Keppler, M., Benisty, M., Müller, A., et al. 2018, A&A, 617, A44
Klahr, H., & Kley, W. 2006, A&A, 445, 747
Kley, W. 1989, A&A, 208, 98
Kley, W., & Crida, A. 2008, A&A, 487, L9
Kley, W., & Nelson, R. P. 2012, ARA&A, 50, 211
Lin, D. N. C., & Papaloizou, J. 1985, in Protostars and Planets II, eds. D. C. Black, & M. S. Matthews (Tucson: University of Arizona Press), 981
Lodato, G. 2007, Rivista del Nuovo Cimento, 30, 293
Lyra, W., Richert, A. J. W., Boley, A., et al. 2016, ApJ, 817, 102
Masset, F. 2000, A&AS, 141, 165
Menou, K., & Goodman, J. 2004, ApJ, 606, 520
Mignone, A., Bodo, G., Massaglia, S., et al. 2007, ApJS, 170, 228
Mignone, A., Flock, M., Stute, M., Kolb, S. M., & Muscianisi, G. 2012, A&A, 545, A152
Müller, T. W. A. 2014, PhD Thesis, University of Tuebingen, Tuebingen, Germany
Müller, T. W. A., & Kley, W. 2012, A&A, 539, A18
Müller, T. W. A., Kley, W., & Meru, F. 2012, A&A, 541, A123
Muto, T., Grady, C. A., Hashimoto, J., et al. 2012, ApJ, 748, L22
Ogilvie, G. I., & Lubow, S. H. 2002, MNRAS, 330, 950
Paardekooper, S. J., Baruteau, C., Crida, A., & Kley, W. 2010, MNRAS, 401, 1950
Rafikov, R. R. 2016, ApJ, 831, 122
Schoonenberg, D., Ormel, C. W., & Krijt, S. 2018, A&A, 620, A134
Semenov, D., Henning, T., Helling, C., Ilgner, M., & Sedlmayr, E. 2003, A&A, 410, 611
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 500, 33
Toomre, A. 1964, ApJ, 139, 1217
Zhang, K., Blake, G. A., & Bergin, E. A. 2015, ApJ, 806, L7
Zhang, S., Zhu, Z., Huang, J., et al. 2018, ApJ, 869, L47
Zhu, Z., Dong, R., Stone, J. M., & Rafikov, R. R. 2015, ApJ, 813, 88

A&A 633, A29 (2020)

Appendix A: Grid and code validation and physics justification

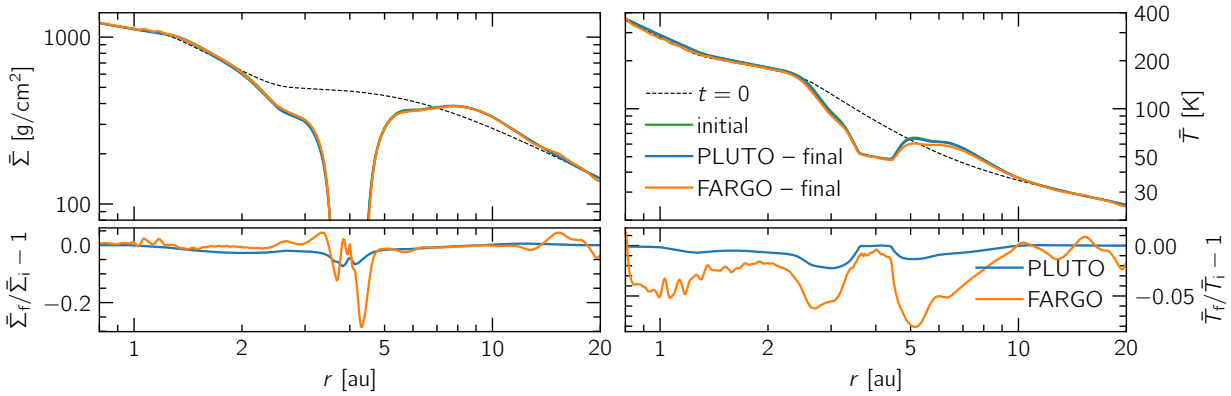

Fig. A.1. Comparison between PLUTO and FARGO, after restarting from identical disk state in quasi-equilibrium (t_i = 6250 orbits) and running independently until $t_f = t_i + 650$ orbits. Both surface density and temperature profiles are in good agreement across codes in the outer disk, while the different treatment of shock heating between the codes becomes evident only near to the planet.

In order to verify our setup, we ran a comparison test against the numerical hydrodynamics code FARGO (Masset 2000). For this test, we simulated the first 6250 orbits (50 kyr) for a fiducial model ($M_p = 100\,M_\oplus$, $r_p = 4$ au, $\dot{M} = 10^{-8}\,M_\odot\,\mathrm{yr}^{-1}$, $\alpha = 10^{-3}$) using PLUTO, then transferred the current, quasi-equilibrium disk state to FARGO and ran it for an additional 650 orbits using both codes and using exactly the same physics. The final disk state for surface density and temperature is plotted in Fig. A.1. We find that the two codes produce similar results in the inner and outer disk, differing only around the planet. We rationalize this by pointing out the fundamentally different treatment of shocks between the two codes: PLUTO utilizes a Godunov-type scheme (a conservative, finite-volume approximation combined with a Riemann solver) that captures and resolves shocks with very good accuracy, in contrast to FARGO's treatment of shocks.

Overall, the level of agreement between the two codes provides good grounds to assume that our setup is working as intended and that we can proceed with simulating our suite of models using PLUTO.

Next, we verified our grid size. We used enough cells in the radial direction such that the pressure scale height H is resolved by six or more cells at the planet's location. We used an appropriate grid size in the azimuthal direction to maintain square cells (roughly twice the number of radial cells). To check whether this grid size was large enough, we reran our fiducial model with double the resolution on both the r and ϕ axes (using PLUTO) and compared the azimuthally-averaged surface density and temperature profiles (see Fig. A.2). The results turned out to be quite similar (to roughly 90%) so for our qualitative study this resolution of six cells per scale height is justified. The grid size used for our simulations is shown in Table 2.

Finally, we investigated the influence of radiative diffusion on the phenomena we intended to study, namely the shock strength of planetary spirals and the influence of the planet's shock heating on the water ice line. We find that accounting for radiative diffusion within the disk midplane barely affected the outcome of the two simulations. It was enabled in the fiducial model for $M_p = 10$ and $100\,M_\oplus$ so that a case where no gap opens can also be studied. We report on the effect of radiative diffusion for these two models in more detail in Appendix C.

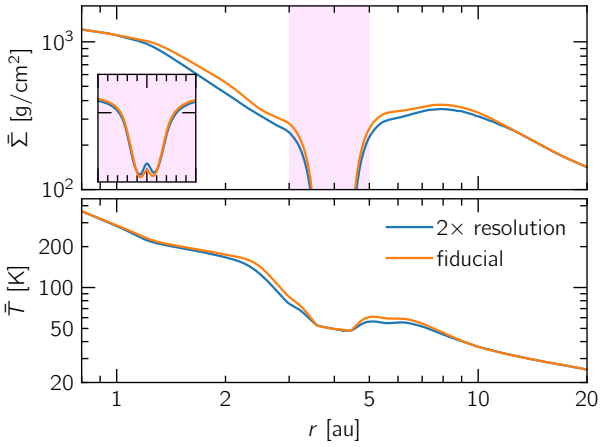

Fig. A.2. Two runs for our fiducial model, one of which (blue curves) has double resolution on both r and ϕ axes (for total of four times as many cells). The snapshots are taken at t = 100 kyr (12 500 orbits) where equilibrium has more or less been reached. The inset zooms in on the pink-tinted region, showcasing the match between the two gap profiles.

Appendix B: Grid structure

As described in Sect. 3.2, our first step was to generate 1D models for various combinations of $\dot{M}$ and α. These models were calculated for $r \in [0.2, 100]$ au and then an appropriate region was selected depending on planet semimajor axis r_p by constructing a grid that mimics the PLUTO grid structure and fitting our initial profiles onto it through linear interpolation. That grid typically extends from $r_p/5$–$5r_p$ except for simulations with $10\,M_\oplus$ planets, which were carried out earlier with a domain always between 0.5–20 au regardless of planet location. A verification test was carried out to make sure that that setup did not affect the quality of the simulations and produced results identical to those using the former setup, therefore these simulations did not need to be rerun.

As far as grid size is concerned, we measured the pressure scale height H_p at the planet's location and constructed a logarithmically-spaced array of N_r cells in the r-direction that satisfies:

$$\frac{H_\mathrm{p}}{\Delta r_\mathrm{p}} \geq 6, \quad H_\mathrm{p} = \frac{c_\mathrm{siso}}{\Omega_\mathrm{K}}\bigg|_{\mathrm{p,t=0}} = \sqrt{\frac{RT_\mathrm{p} r_\mathrm{p}^3}{\mu G M}}, \quad M = M_* + M_\mathrm{p}, \tag{B.1}$$

while the number of cells in the azimuthal direction was chosen so that cells are square, or:

$$N_\phi = \left\| \pi \frac{a+1}{a-1} \right\|, \quad a = \left(\frac{r_\mathrm{out}}{r_\mathrm{in}} \right)^{1/N_r} \tag{B.2}$$

This typically results in grids of around 600×1200 cells (e.g., our fiducial model with $M_\mathrm{p} = 100\,M_\oplus$, $\dot{M} = 10^{-8}\,M_\odot\,\mathrm{yr}^{-1}$, $\alpha = 10^{-3}$, $r_\mathrm{p} = 4$ au contains 531×1037 cells).

Appendix C: Effect of radiative diffusion

We repeated two simulations for our fiducial model ($r_\mathrm{p} = 4$ au, $\dot{M} = 10^{-8}\,M_\odot\,\mathrm{yr}^{-1}$, $\alpha = 10^{-3}$) with flux-limited diffusion (FLD) enabled in the disk midplane. This additional term should smear out peaks in the temperature structure of the disk and could become important along the trajectories of spirals. We choose $M_\mathrm{p} \in \{10, 100\}\,M_\oplus$ to examine its overall impact on the disk whether a gap is carved or not. However, we do not see a significant difference in the low- nor in the high-mass simulations except for the lower peak temperature of spirals in the FLD models and a slight inward movement of the ice line. This implies that vertical cooling happens at a much faster rate than the planar diffusion timescale. By comparing their timescales, we indeed find that thermal cooling readjusts disk temperatures roughly 100 times faster than radiative diffusion does (except for inside the gap region) to the extent that its effect on temperature profiles and the ice line is negligible.

Disk profiles and spiral arms

A comparison is plotted in Fig. C.1 for the high-mass case. Spiral arms show slightly lower temperature maxima and the azimuthally-averaged temperature profile is smoother overall, with lower highs and higher lows. This effect is strongest around parts of the disk that might contain steep temperature gradients, such as the region between 2.5–3.5 au for these models, but it still barely makes a difference of more than 3% with respect to the model where we did not account for radiative diffusion, leading to identical aspect ratios in the two models and, therefore, practically indistinguishable pitch angles along spirals. We note that for the 10-Earth-mass case, differences between the two models are much smaller.

Location and shape of the ice line

Temperature gradients are slightly different when accounting for radiative diffusion, and especially so around the gap edge. Since the ice line is relatively close to said gap edge in the two models where the module is enabled, we are more or less looking at the effects of radiative diffusion at its maximum potential. However, in the previous paragraph, we report that its effect barely changes

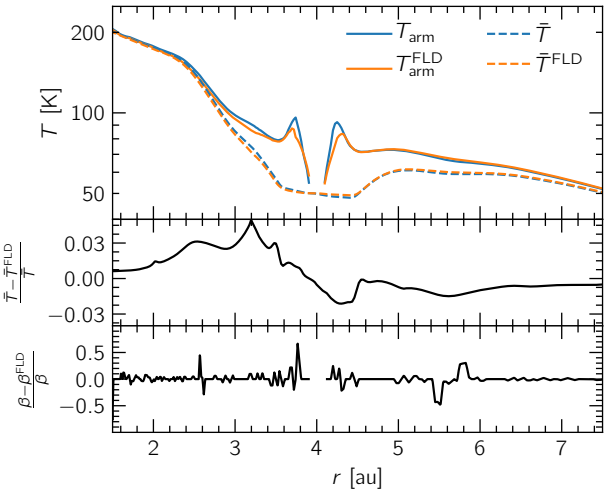

Fig. C.1. Comparison between two variants of our fiducial model ($M_\mathrm{p} = 100\,M_\oplus$), with and without radiative diffusion enabled. The effect of diffusion is visible near steep gradients along spiral crests or the gap edges but negligible in general. Spiral arms in both models overlap exactly, as shown in the *bottom panel*.

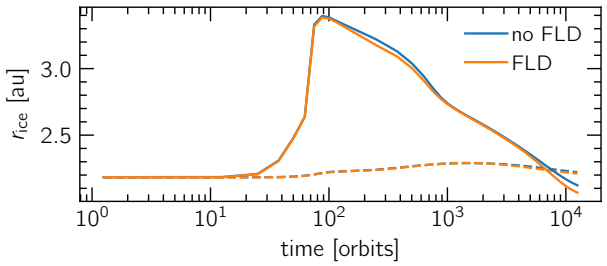

Fig. C.2. Evolution of ice line location for two pairs of simulations with and without treatment of radiative diffusion. Solid and dashed lines refer to models where $M_\mathrm{p} = 10$ and $100\,M_\oplus$, respectively.

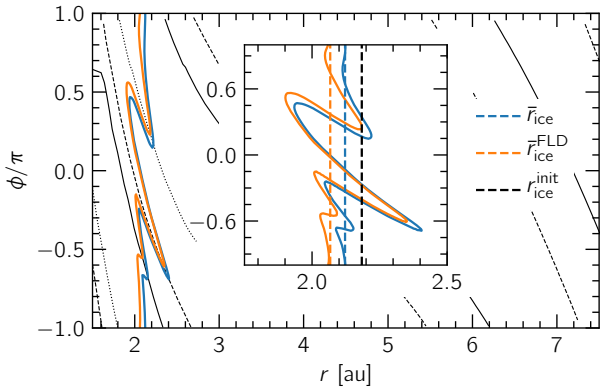

Fig. C.3. Shape of water ice line at equilibrium for two models with and without radiative diffusion ($M_\mathrm{p} = 100\,M_\oplus$). Aside from the slight inward shift of the ice line, its shape is generally unaffected. Here we plot an extended range of our simulation for context. Solid, dashed, and dotted black curves mark the location of primary, secondary, and tertiary spirals, respectively.

the picture with regard to shock strength, gap width, or spiral location. Because of these three points, the ice line's location over time is expected to be slightly, but not significantly, different when compared to that found in our standard simulations.

In Fig. C.2, we compare the location of $\bar{r}_{\rm ice}$ over time between our standard models and their respective FLD-enabled models. Indeed, the softer temperature profile in the inner disk allows the ice line to recede slightly more inwards when radiative diffusion is enabled. Nevertheless, the effect is still minuscule for the high-mass case and negligible for the low-mass case.

A 2D analysis of the ice line's shape offers similar results. As shown in Fig. C.3, accounting for radiative diffusion does not change the shape of the ice line with respect to the standard model, but, instead, it shifts it slightly inwards, as shown in Fig. C.2. As with the 1D analysis, this difference is practically nonexistent for the low-mass case as the temperature profile is softer overall: shocks by the 10-Earth-mass planet are significantly weaker and a gap does not open.

Appendix D: Implementation of radiative diffusion

To examine the effect of radiative diffusion along the disk midplane, we implemented an external module that is coupled with PLUTO and solves the following equation after every timestep:

$$\frac{\partial c_{\rm v}\Sigma T}{\partial t} = -\nabla\cdot(2H\boldsymbol{F}_{\rm rad}), \qquad \boldsymbol{F}_{\rm rad} = -\frac{4\sigma_{\rm SB}}{\lambda\rho\kappa}\nabla T^4, \tag{D.1}$$

where $\boldsymbol{F}$ denotes the radiation flux across the disk midplane and is defined as:

$$\boldsymbol{F}_{\rm rad} = -\frac{4\sigma_{\rm SB}}{\lambda\kappa\rho_{\rm mid}}\nabla T^4, \tag{D.2}$$

where λ is a flux limiter, following Kley (1989).

By defining a diffusion coefficient K as

$$K = 2H\frac{4\sigma_{\rm SB}}{\lambda\rho\kappa}(4T^3) = \frac{32\sigma_{\rm SB}\sqrt{2\pi}}{\lambda\Sigma\kappa}H^2T^3, \tag{D.3}$$

we discretize Eq. (D.1) following Appendix A.1 in Müller (2014)

$$c_{\rm v}\Sigma\frac{\partial T}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rK\frac{\partial T}{\partial r}\right) + \frac{1}{r^2}\frac{\partial}{\partial\phi}\left(K\frac{\partial T}{\partial\phi}\right), \tag{D.4}$$

on a grid where i and j denote cell indices along the r- and ϕ-direction respectively, and obtain:

$$\begin{aligned} c_{\rm v}\Sigma_{ij}&\frac{T^{n+1}_{i,j}-T^n_{i,j}}{\Delta t} \\ &= \frac{1}{r_i\Delta r_i}\left((rK)_{i+\frac{1}{2},j}\frac{T^{n+1}_{i+1,j}-T^{n+1}_{i,j}}{r_{i+1}-r_i} - (rK)_{i-\frac{1}{2},j}\frac{T^{n+1}_{i,j}-T^{n+1}_{i-1,j}}{r_i-r_{i-1}}\right) \\ &+\frac{1}{r_i^2\Delta\phi^2}\left(K_{i,j+\frac{1}{2}}\left(T^{n+1}_{i,j+1}-T^{n+1}_{i,j}\right) - K_{i,j-\frac{1}{2}}\left(T^{n+1}_{i,j}-T^{n+1}_{i,j-1}\right)\right), \end{aligned} \tag{D.5}$$

where n and $n+1$ denote the states at time t and $t+\Delta t$, respectively.

We now have to solve for T^{n+1}. We can group up the right hand side to form a linear system:

$$\begin{aligned} T^n_{i,j} &= A_{ij}T^{n+1}_{i-1,j} + C_{ij}T^{n+1}_{i+1,j} + D_{ij}T^{n+1}_{i,j-1} + E_{ij}T^{n+1}_{i,j+1} + B_{ij}T^{n+1}_{i,j} \\ &\Rightarrow \overline{\mathbf{M}}\cdot T^{n+1} = T^n, \end{aligned} \tag{D.6}$$

where

$$\begin{aligned} A_{ij} &= -\frac{\Delta t}{c_{\rm v}\Sigma_{ij}}\frac{1}{r_i\Delta r_i}\frac{(rK)_{i-\frac{1}{2},j}}{r_i-r_{i-1}}, \qquad C_{ij} = -\frac{\Delta t}{c_{\rm v}\Sigma_{ij}}\frac{1}{r_i\Delta r_i}\frac{(rK)_{i+\frac{1}{2},j}}{r_{i+1}-r_i}, \\ D_{ij} &= -\frac{\Delta t}{c_{\rm v}\Sigma_{ij}}\frac{1}{r_i^2\Delta\phi^2}K_{i,j-\frac{1}{2}}, \qquad E_{ij} = -\frac{\Delta t}{c_{\rm v}\Sigma_{ij}}\frac{1}{r_i^2\Delta\phi^2}K_{i,j+\frac{1}{2}}, \\ B_{ij} &= 1 - A_{ij} - C_{ij} - D_{ij} - E_{ij}, \end{aligned} \tag{D.7}$$

are elements of the matrix $\overline{\mathbf{M}}$. We solve this system using successive overrelaxation (SOR). Therefore, we calculate and fix $\overline{\mathbf{M}}$ before iterating over:

$$\begin{aligned} \tilde{T}^{k+1}_{i,j} &= (1-\omega)\tilde{T}^k_{i,j} \\ &- \frac{\omega}{B_{ij}}\left[A_{ij}\tilde{T}^k_{i-1,j} + C_{ij}\tilde{T}^k_{i+1,j} + D_{ij}\tilde{T}^k_{i,j-1} + E_{ij}\tilde{T}^k_{i,j+1} - T^{k=0}_{i,j}\right], \end{aligned} \tag{D.8}$$

with $\omega = 1.5$ until $\tilde{T}_{i,j}$ converges. Boundary conditions during this iterative process are closed in order to conserve total thermal energy through the radiative diffusion substep.

C.2 Ziampras et al. (2020b)

A&A 637, A50 (2020)
https://doi.org/10.1051/0004-6361/201937048

Astronomy & Astrophysics

Importance of radiative effects in gap opening by planets in protoplanetary disks

Alexandros Ziampras[1], Wilhelm Kley[1], and Cornelis P. Dullemond[2]

[1] Institut für Astronomie und Astrophysik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany
e-mail: alexandros.ziampras@uni-tuebingen.de
[2] Institute for Theoretical Astrophysics, Heidelberg University, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany

Received 4 November 2019 / Accepted 2 March 2020

ABSTRACT

Recent ALMA observations revealed concentric annular structures in several young class-II objects. In an attempt to produce the rings and gaps in some of these systems, they have been modeled numerically with a single embedded planet assuming a locally isothermal equation of state. This is often justified by observations targeting the irradiation-dominated outer regions of disks (approximately 100 au). We test this assumption by conducting hydrodynamics simulations of embedded planets in thin locally isothermal and radiative disks that mimic the systems HD 163296 and AS 209 in order to examine the effect of including the energy equation in a seemingly locally isothermal environment as far as planet–disk interaction is concerned. We find that modeling such disks with an ideal equation of state makes a difference in terms of the number of produced rings and the spiral arm contrast in the disk. Locally isothermal disks produce sharper annular or azimuthal features and overestimate a single planet's gap-opening capabilities by producing multiple gaps. In contrast, planets in radiative disks carve a single gap for typical disk parameters. Consequently, for accurate modeling of planets with semimajor axes up to about 100 au, radiative effects should be taken into account even in seemingly locally isothermal disks. In addition, for the case of AS 209, we find that the primary gap is significantly different between locally isothermal and radiative models. Our results suggest that multiple planets are required to explain the ring-rich structures in such systems.

Key words. protoplanetary disks – planet–disk interactions – methods: numerical

1. Introduction

Planets are born in protoplanetary disks. While they do not always make their presence clear like in the case of PDS 70b and c (Keppler et al. 2018; Haffert et al. 2019), the ALMA observations of the DSHARP survey (Andrews et al. 2018) have provided theorists with high-fidelity datasets for testing the planet–disk interaction theory. It is fascinating how well a single planet can reproduce the annular substructures in some of the observed systems, for instance, Zhang et al. (2018).

Most of the aforementioned systems show resolved features at the 100 au scale, suggesting that the disk temperature profile is set by a balance between heating by stellar irradiation and thermal cooling. This makes modeling these systems with a locally isothermal equation of state quite attractive because the cooling timescale in the optically thin irradiation-dominated outer disk is commonly sufficiently short (shorter than a hundredth of an orbit) to render radiative effects negligible. At the same time, the very low level of effective viscosity that is inferred for these disks (Zhang et al. 2018) indicates minuscule contributions by viscous heating to the thermal budget of the disk, such that the locally isothermal assumption seems further justified.

However, an embedded planet also interacts with the disk gravitationally, forming spiral arms (Ogilvie & Lubow 2002; Rafikov 2002) that permeate the disk and can steepen into shocks (Zhu et al. 2015). Several studies have shown that heating by these spiral shocks is another significant heat source at distances of several au for solar-type stars (Richert et al. 2015; Lyra et al. 2016; Rafikov 2016; Ziampras et al. 2020). While the planet's contribution to the energy budget decreases with increasing distances from the star, there is a gray area at a few tens of au where shock heating could operate to some degree, even when the cooling timescale is a small fraction of the local orbital period.

In addition, Miranda & Rafikov (2019) showed that even in seemingly locally isothermal scenarios, assuming an adiabatic equation of state results in fundamentally different physics concerning the angular momentum flux in protoplanetary disks, and it therefore leads to noticeable changes in dust continuum profiles. They quoted distances of about 80 au as the lower limit beyond which a locally isothermal equation of state can indeed be justified, and showed that deviations of about 10^{-3} from $\gamma = 1$ can lead to noticeable differences (of about 10%) in simulated outcomes and therefore observables such as continuum emission intensity.

This subject is quickly gaining clarity as more effort is made to understand planet-induced gap opening in order to interpret observations and constrain the properties of protoplanetary disks. During the reviewing process for this paper, two publications by Miranda & Rafikov (2020) and Zhang & Zhu (2020) were made available, which discuss the effect of a finite cooling timescale on the location and number of gaps that can be produced by a single planet. Our study does not reiterate these new results, but rather enriches them by handling radiative effects differently. It serves as an additional point of view in arriving at a similar conclusion: radiative cooling is of key importance in determining the radial structure of a disk with an embedded planet.

A&A 637, A50 (2020)

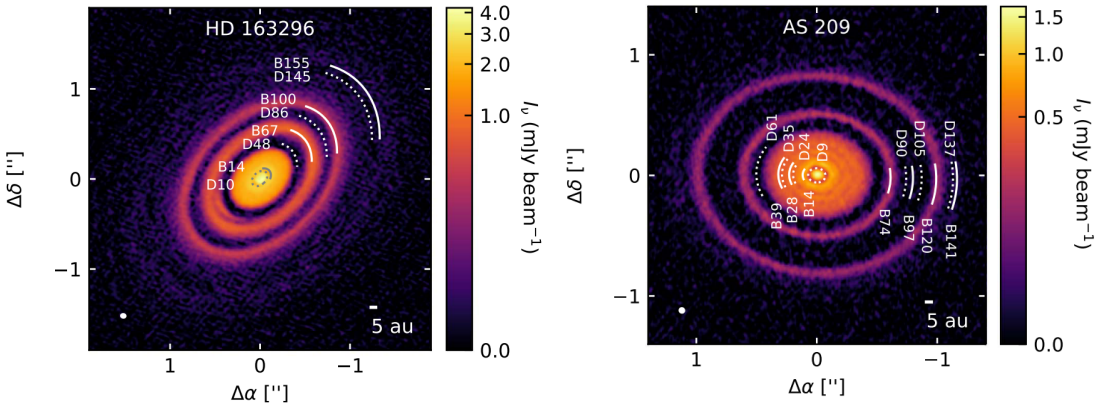

Fig. 1. ALMA continuum emission observations of the two systems we used as testbeds for our numerical models, as shown in Huang et al. (2018). Solid and dotted white arcs mark the location of bright rings and dark gaps, respectively, along with their distance from the host star in au. It is likely that a planet is responsible for one or more of the gaps observed in either system. *Left*: HD 163296, a 12.6 Myr old likely isolated system that features at least two clearly visible gaps at 48 and 86 au. Our models focus on the 20–60 au range, with a planet carving a gap at 48 au. *Right*: AS 209, a 1 Myr old system in the Ophiuchus region that is rich in annular structures, boasting five rings between 20 and 130 au. This source has been modeled by Zhang et al. (2018), who found that a fit with a single planet at 99 au could reconstruct all five of these rings.

We use numerical simulations to address the importance of proper treatment of radiative effects for two systems that mimic the properties of HD 163296 and AS 209 (Andrews et al. 2018; Zhang et al. 2018) and show that an adiabatic equation of state can produce different results with respect to a locally isothermal model when the gap-opening capabilities of a planet are modeled.

In Sect. 2 we present our physical and numerical setup in modeling these two systems. We present our results in terms of disk structures in Sect. 3, and discuss their implications in Sect. 4. Finally, we summarize our results and conclude this study in Sect. 5.

2. Sources, physics, and numerics

In this section we describe the physical and numerical modeling of HD 163296 and AS 209. We first present the two systems that we use as our reference. We then describe the source terms in the vertically integrated hydrodynamics equations, our physical assumptions in terms of star, planet, and disk properties, and the initial and boundary conditions for our simulations.

2.1. Sources: HD 163296 and AS 209

Rather than carrying out a study on an arbitrary toy model of a protoplanetary disk, we decided to model two of the sources targeted by the DSHARP survey (Andrews et al. 2018), namely HD 163296 and AS 209 (see Fig. 1). Both of these sources feature annular structures in the form of bright rings and dark gaps in dust continuum emission images, with the former also showing a crescent at roughly 50 au and the latter standing out with its ring-rich emission profile. These systems have been modeled in the past and their various features have been studied (Huang et al. 2018; Dullemond et al. 2018; Fedele et al. 2018; Zhang et al. 2018; Zhang & Zhu 2020), and using them as testbeds for our study can provide a deeper insight on any planets they might harbor.

2.2. Hydrodynamics, heating, and cooling

The two-dimensional (2D) vertically integrated Navier–Stokes equations in cylindrical $\{r, \phi, z\}$ coordinates for a perfect gas with surface density Σ, velocity $\boldsymbol{v}$ and specific midplane internal energy e read

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = -\Sigma\nabla\cdot\boldsymbol{v}, \tag{1a}$$

$$\Sigma\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\nabla p + \Sigma\boldsymbol{g} + \nabla\cdot\boldsymbol{\Sigma}, \tag{1b}$$

$$\frac{\mathrm{d}\Sigma e}{\mathrm{d}t} = -\gamma\Sigma e\nabla\cdot\boldsymbol{v} + Q_{\rm visc} + Q_{\rm irr} - Q_{\rm cool}, \tag{1c}$$

where γ is the adiabatic index and $p = (\gamma-1)\Sigma e$ is the vertically integrated pressure. External source terms (in our case, gravitational forces) are contained in $\boldsymbol{g}$, and $\boldsymbol{\Sigma}$ denotes the viscous stress tensor (Tassoul 1978).

For a thin disk of gas rotating on a Keplerian orbit with orbital frequency $\Omega_{\rm K} = \sqrt{GM/r^3}$, at distance r around a star of mass M, the pressure scale height of the gas is $H = c_{\rm siso}/\Omega_{\rm K}$. Here, $c_{\rm siso} = \sqrt{p/\Sigma}$ is the isothermal sound speed and relates to the adiabatic sound speed $c_{\rm s}$ as $c_{\rm s} = \sqrt{\gamma}c_{\rm siso}$. The aspect ratio of the disk is then $h = H/r$. The gravitational constant and the mean molecular weight of the gas are denoted by G and μ, respectively. We adopted a typical solar-composition disk of molecular H–He gas, therefore $\gamma = 7/5$ and $\mu = 2.35$.

In a locally isothermal framework, the energy equation in Eq. (1) is not solved. Instead, a fixed sound speed profile is adopted. This in turn defines a (fixed) radial temperature profile T, as $c_{\rm siso} = \sqrt{R_{\rm gas}T/\mu}$, with $R_{\rm gas}$ being the gas constant. In the locally isothermal scenario this temperature profile is constant on cylinders and only dependent on radius, but we focus on the midplane temperature in this 2D approximation.

More generally, however, the full system of equations is solved. In this case, we include viscous heating, stellar irradiation, and thermal cooling in the energy equation as follows:

$$Q_{\rm visc} = \frac{1}{2\nu\Sigma}\mathrm{Tr}(\boldsymbol{\Sigma}^2) = \frac{1}{2\nu\Sigma}\left(\sigma_{rr}^2 + 2\sigma_{r\phi}^2 + \sigma_{\phi\phi}^2 + \sigma_{zz}^2\right) \tag{2a}$$

$$Q_{\rm irr} = 2\frac{L_*}{4\pi r^2}(1-\epsilon)\left(\frac{\mathrm{d}\log H}{\mathrm{d}\log r} - 1\right)h\frac{1}{\tau_{\rm eff}}, \tag{2b}$$

$$Q_{\rm cool} = 2\sigma_{\rm SB}\frac{T^4}{\tau_{\rm eff}}, \tag{2c}$$

Table 1. Physical and numerical parameters used in our modeling of the sources HD 163296 and AS 209.

System	HD 163296						AS 209
Parameter	(1) Base	(2) $\gamma \approx 1$	(3) Low κ	(4) κ by LP85	(5) Low α	(6) Shallow Σ	(7) Base
M_* [$M_\odot$]	2.089	–	–	–	–	–	0.83
L_* [$L_\odot$]	16.98	–	–	–	–	–	1.41
M_p [M_J]	0.5	–	–	–	–	–	0.083
M_p [$10^{-3}\,M_*$]	0.29	–	–	–	–	–	0.1
M_disk [% $M_\odot$]	5.58	–	–	–	–	7.37	68.1
M_disk [% M_*]	2.67	–	–	–	–	3.53	82.0
r_p [au]	48	–	–	–	–	–	99
$\Sigma_{t=0}(r)$	$\propto r^{-3/2}$	–	–	–	–	$\propto r^{-1}$	$\propto r^{-1}$
Σ_p [g cm^{-2}]	9.61	–	–	–	–	–	10.0
h_p [%]	5.601	–	–	–	–	–	8.175
T_p [K]	34.5	–	–	–	–	–	14.2
γ	1.4	1.01	–	–	–	–	1.4
κ [cm^2 g^{-1}]	0.45	–	0.045	LP85	–	–	LP85
$\log\alpha$	−4	–	–	–	−5	–	−5
r_min–r_max [r_p]	0.2–5	–	–	–	–	–	0.1–10
r_min–r_max [au]	9.6–240	–	–	–	–	–	9.9–990
$N_r \times N_\phi$	573 × 1118	–	–	–	–	–	673 × 1312
$N_\mathrm{cells}/H_\mathrm{p}$	10	–	–	–	–	–	12

Notes. The opacity model by Lin & Papaloizou (1985; here referred to as LP85) functionally translates to $\kappa = 5\times10^{-4}\,T^2$ within our simulation domain ($T < 170$ K). Only the two models labeled "base" (Cols. 1 and 7) are discussed in the results section; the remaining models (Cols. 2–7) were used to verify our results and are discussed in Appendix A. Dashes imply that a parameter is inherited from the base model (Col. 1 for HD 163296). In our radiative models, the energy equation allows temperature-related quantities (i.e., h_p, T_p, κ if defined by LP85) to evolve throughout the simulation. These are otherwise kept fixed in the corresponding locally isothermal runs.

where ν is the kinematic viscosity. For this we used the α-ansatz (Shakura & Sunyaev 1973), which for thin disks reads $\nu = \alpha c_\mathrm{s} H$. The Stefan-Boltzmann constant is denoted by σ_SB, and τ_eff is an effective optical depth following Hubeny (1990)

$$\tau_\mathrm{eff} = \frac{3\tau}{8} + \frac{\sqrt{3}}{4} + \frac{1}{4\tau}, \tag{3}$$

where τ is the vertical optical depth measured from the midplane to the disk surface,

$$\tau = \int_0^\infty \kappa\rho \mathrm{d}z \approx c_1 \kappa_\mathrm{mid}\rho_\mathrm{mid} H. \tag{4}$$

Here, ρ is the volume density and $\kappa_\mathrm{mid}(\rho, T)$ is the Rosseland mean opacity at the disk midplane, which we refer to as simply κ and evaluate in Sect. 2.3 below, either using a constant value or adopting the temperature-dependent opacity law. To match the opacity drop with height, we included a correction factor $c_1 = 1/2$ (Müller & Kley 2012). The midplane gas density ρ_mid is related to the surface density such that $\Sigma = \sqrt{2\pi}\rho_\mathrm{mid} H$ (i.e., corresponding to a Gaussian vertical stratification at hydrostatic equilibrium).

Our cooling and irradiation prescription adapts the model by Menou & Goodman (2004) for a disk with an albedo of $\epsilon = 1/2$ around a star with luminosity L_*. The factor $\frac{\mathrm{d}\log H}{\mathrm{d}\log r}$ was assumed to be constant and equal to $9/7$ (i.e., disk self-shadowing was not considered). For more details, see the physical setup by Ziampras et al. (2020).

2.3. Numerics

We used the PLUTO code (Mignone et al. 2007) along with the FARGO method (Masset 2000; Mignone et al. 2012) for our simulations. Models with an embedded planet were run on a polar $\{r, \phi\}$ grid, logarithmically spaced in the radial direction.

In order to set our initial conditions, we constructed two different disks that mimic the structure of HD 163296 and AS 209. We therefore set M_* and L_* according to Andrews et al. (2018), and decided to embed a single planet with mass $M_\mathrm{p} = 0.5\,M_J$ at $r_\mathrm{p} = 48$ au for HD 163296 and $M_\mathrm{p} = 0.083\,M_J$ at $r_\mathrm{p} = 99$ au for AS 209. We adopted a viscous α following Zhang et al. (2018) (either 10^{-4} or 10^{-5}, constant throughout the disk). For the Rosseland mean opacity we chose either a constant value $\kappa = 0.45\,\mathrm{cm^2\,g^{-1}}$ (adapted from Birnstiel et al. 2018) or a temperature-dependent opacity law by Lin & Papaloizou (1985). The latter dictates that $\kappa = 5\times10^{-4}T^2$ for temperatures under 170 K, which is true for the full extent of the simulated disk. A list of the physical and numerical parameters used in our models is given in Table 1.

The initial surface density needs to be prescribed, and the temperature is simply given by $Q_\mathrm{visc} + Q_\mathrm{irr} = Q_\mathrm{cool}$ (see Eq. (1c)) in the absence of a planet (we can ignore compression heating in this scenario because its contribution is negligible). We adopted values inferred in Table 3 by Zhang et al. (2018) for HD 163296 such that the surface density at $r_\mathrm{gap} = \{10, 48, 86\}$ au is $\Sigma(r_\mathrm{gap}) = \{100, 10, 3\}\,\mathrm{g\,cm^{-2}}$, respectively. We then fit a power law to these three points and find that a fit of

$$\Sigma(r)^\mathrm{HD} \approx 10\left(\frac{r}{r_\mathrm{p}}\right)^{-3/2}\frac{\mathrm{g}}{\mathrm{cm^2}} \approx 3200\left(\frac{r}{\mathrm{au}}\right)^{-3/2}\frac{\mathrm{g}}{\mathrm{cm^2}} \tag{5}$$

matches very well. For AS 209, we simply adopt a setup similar to that by Zhang et al. (2018), such that

$$\Sigma(r)^\mathrm{AS} \approx 10\left(\frac{r}{r_\mathrm{p}}\right)^{-1}\frac{\mathrm{g}}{\mathrm{cm^2}}. \tag{6}$$

As far as temperature is concerned, the effect of viscous heating $Q_{\rm visc}$ is functionally negligible within our simulation range for our α value. In the absence of a planet, the temperature profile is therefore set by a balance between irradiation heating and thermal cooling. In Eqs. (2b) and (2c), the optical depths for absorption and emission in $Q_{\rm irr}$ and $Q_{\rm cool}$, respectively, can in principle be different, but as a first approximation we assumed that they are identical and can be factored out when we set $Q_{\rm irr} = Q_{\rm cool}$. The temperature profile then only depends on stellar properties so that

$$T(r) = T_{\rm p}\left(\frac{r}{r_{\rm p}}\right)^{-3/7} \text{K} \quad \propto \frac{h^2}{r} \quad \Rightarrow \quad h(r) = h_{\rm p}\left(\frac{r}{r_{\rm p}}\right)^{2/7} \tag{7}$$

for our irradiation prescription. By replacing the geometrical factor $\left(\frac{{\rm d}\log H}{{\rm d}\log r} - 1\right)h$ with a constant irradiation angle ϕ (e.g., Dullemond et al. 2018), we recover the familiar formula where $T \propto r^{-1/2}$ and $h \propto r^{1/4}$ instead (not used here).

After constructing our initial surface density and temperature profiles, we first ran a one-dimensional radiative simulation without a planet and verified that the profiles indeed correspond to an equilibrium state. We then embedded the planet, and using the power-law profiles given in Eqs. (5)–(7) as initial conditions, executed various radiative as well as locally isothermal simulations.

In all setups, Σ is damped to the initial profiles at the boundaries according to de Val-Borro et al. (2006) over a timescale of 0.3 boundary orbital periods. The planet does not accrete material or migrate through the disk, and traces circular orbits around the star. The target simulation time was 1000 orbits for HD 163296 – which at 48 au translates into roughly 230 kyr – and 5000 orbits for AS 209 (5.4 Myr at 99 au) because these models have a very low viscosity of 10^{-5}. We used a resolution of 573×1118 and 673×1312 cells for HD 163296 and AS 209, respectively, which results in square cells with 10–12 cells per scale height in the radial direction at the planet's location.

Similar to Ziampras et al. (2020), the planet's presence is described by an additional gravitational force. We accounted for the shift of the system barycenter that is caused by the planet, but neglected backreaction of the disk onto the star and planet.

To prevent singularities near the planet, we introduced a softening length $\epsilon = 0.6H$ following Müller et al. (2012), using the local pressure scale height and a correction factor that accounts for the disk thickness.

3. Results

In this section we present the results of our numerical simulations. As stated above, we tested several different models. We focus in this section on the locally isothermal and radiative models with parameters described in the previous section. The remaining models are used for comparison purposes and are discussed in Appendix A.

First, we present our results on HD 163296. We start with focusing on overall disk profiles and differences between the locally isothermal and radiative models in terms of the radial distribution of gas in the disk. We then compare and contrast the structure of spiral arms between the two models in terms of surface density and temperature contrast with respect to the disk background. This is then followed up by results on AS 209 regarding the first point only because this is the most crucial to the discussion.

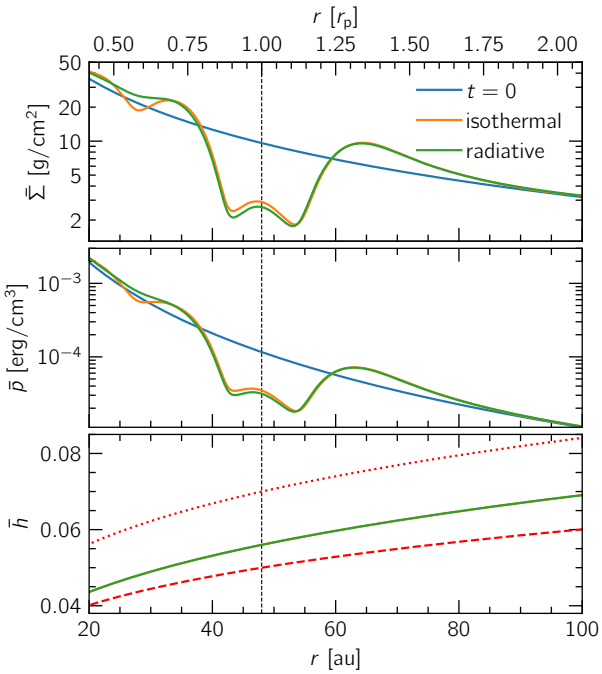

Fig. 2. Azimuthally averaged surface density, midplane pressure, and aspect ratio (used as a proxy for temperature, see Eq. (7)) of HD 163296 for a locally isothermal (in orange) and a radiative (in green) disk after 1000 orbits with an embedded planet with mass 0.5 M_J at 48 au. Vertical black lines mark the location of the planet. The dashed and dotted red lines in the *bottom panel* denote the aspect ratio profiles used in the simulations by Zhang et al. (2018), where $h \propto r^{1/4}$ and $h_{\rm p} = 0.05$ and 0.07, respectively. The fact that our aspect ratio profile is bounded by these two curves allows us to validate our results by comparing them to that study (see Sect. 4).

3.1. HD 163296

The first system we studied is HD 163296. This system features several rings, as shown in Fig. 1.

3.1.1. Disk profiles and secondary gaps

Our main results are plotted in Fig. 2. By comparing our radiative and isothermal runs and computing the aspect ratio as a proxy for temperature (because $h \propto \sqrt{rT}$), we find that the azimuthally averaged temperature profiles of the two are practically identical with each other (bottom panel in Fig. 2), and therefore with the profile that an irradiation-dominated disk should have. However, the azimuthally averaged surface density profiles of the two models differ in the inner disk and around the planet within 1000 orbits. The difference lies in the additional depression of the gas surface density at around 28 au for the locally isothermal case, in contrast to the smoother profile in the radiative run. This secondary gap, while shallow, can warrant the existence of a pressure bump in the inner disk, creating a dust trap and therefore a second and potentially visible dust ring. In contrast to the isothermal run, this pressure minimum disappears in the radiative simulation (middle panel in Fig. 2).

A rather small (compared to that at 28 au) but visible difference can also be seen within the gap region. This may be related to nonlinear effects caused by the planet's gap opening, and discuss this in Sect. 4. Finally, the azimuthally averaged surface

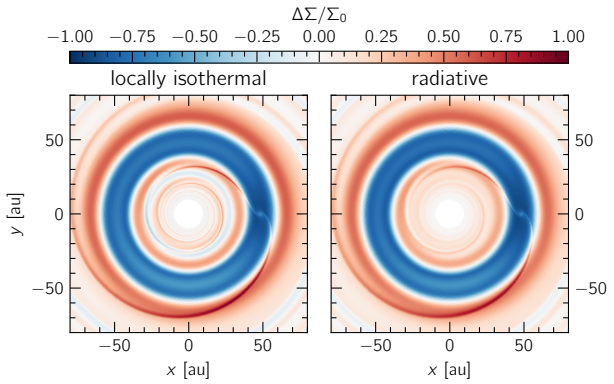

Fig. 3. Surface density perturbation with respect to the initial profiles for the two models of HD 163296. The locally isothermal run features a secondary gap at roughly 28 au and slightly more prominent inner spirals.

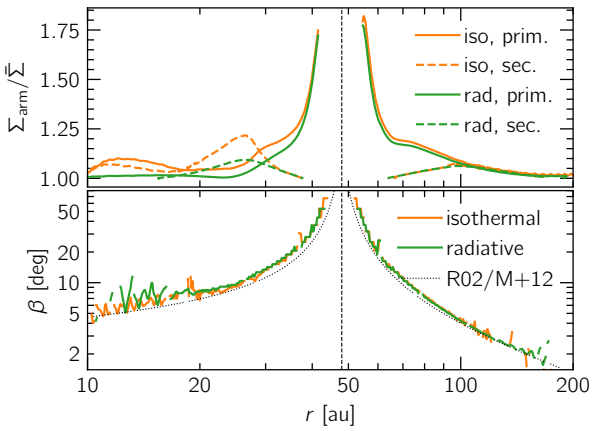

Fig. 4. Comparison of spiral arm properties between our two models for HD 163296. *Top*: primary and secondary spiral arm contrast with respect to the disk background. We find that the locally isothermal model shows consistently higher contrast. *Bottom*: pitch angle of spirals as a function of radius. Because the two models have identical aspect ratios, the overlap of the two curves is to be expected. The vertical black line marks the planet location. The dotted black line corresponds to the analytical formulae in Rafikov (2002), Eq. (44), and Muto et al. (2012), Eq. (1).

density in the outer disk is indistinguishable between the two models.

In Fig. 3 we compare the surface density in the disk to its initial profile. A gap and ring are clearly visible at 28 and 34 au, respectively, in the locally isothermal model. Additionally, spiral arms in the inner disk have slightly higher contrast with the background for that model compared to its radiative counterpart. We carry out a more detailed comparison in the next section.

3.1.2. Spiral arm contrast

As mentioned in the work of Miranda & Rafikov (2019), the locally isothermal equation of state overestimates the contrast of structures in protoplanetary disks. We already saw this behavior in Fig. 2, where a secondary gap is visible at 28 au. However, nonaxisymmetric features in the disk such as spiral arms are lost when averaging along annuli. With this in mind, it is useful to track the spiral arms that are excited by the planet and compare the surface density along their crests with its azimuthally averaged values. In doing so, we can estimate their contrast with respect to the disk background.

This comparison was carried out by tracking the trajectory of each individual spiral, logging the surface density $\Sigma_{\rm arm}$ along their peaks and then plotting the ratio $\Sigma_{\rm arm}/\bar{\Sigma}$, where $\bar{\Sigma}$ is the azimuthally averaged surface density (see Ziampras et al. 2020 for details). Our results are summarized in Fig. 4 and show that spirals in the locally isothermal model are indeed consistently "stronger" (i.e., have a higher contrast with respect to the disk), with the effect being more prominent in the inner disk (top panel).

The pitch (or opening) angle β of these spirals can be defined (e.g., Zhu et al. 2015) through

$$\tan\beta \equiv \frac{{\rm d}r}{r{\rm d}\phi} \approx \frac{c_{\rm s}}{r|\Omega(r)-\Omega_{\rm p}|} = \frac{h}{|1-(r/r_{\rm p})^{3/2}|}. \tag{8}$$

Equation (8) suggests that the pitch angle of spirals should be a function of the aspect ratio (again, a proxy for the temperature) for a given distance r. With this in mind, and given that the two models share an identical radial temperature profile, we expect and observe that the pitch-angle profiles of both primary and secondary spirals match very well for the locally isothermal and radiative model (lower panel).

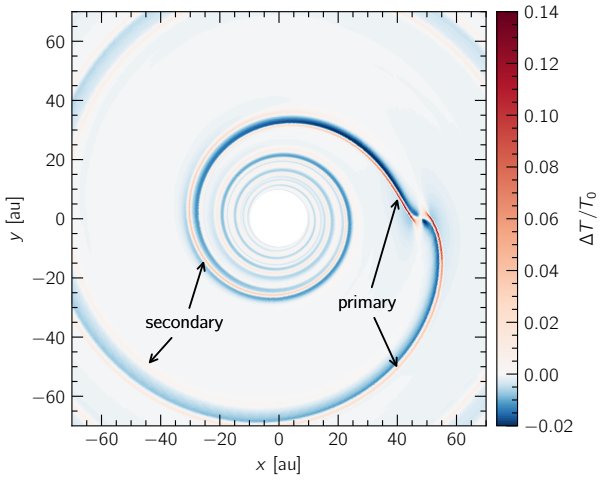

Fig. 5. Temperature contrast of spirals with respect to the disk background for the radiative run of HD 163296. The effect of spiral heating on the global disk is negligible, and the temperature along spiral crests can increase by up to 15%.

We also compared the temperature along spiral crests with the initial axisymmetric profile for the radiative model. We found that differences are about 10–15% or smaller, and that spiral heating on the entire disk is negligible (see Fig. 5). This is to be expected in the optically thin irradiation-dominated region of the disk at $r \gtrsim 20$ au (e.g., Rafikov 2016).

3.2. AS 209

The goal in modeling this additional source was to see how this disparity between the locally isothermal and adiabatic equations of state would affect the gas surface density profile of a system with more gaps observable in continuum emission: five in AS 209 (at $r \in \{24, 35, 61, 90, 105\}$ au) compared to only one in

A&A 637, A50 (2020)

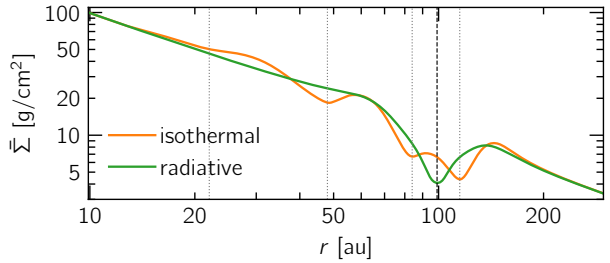

Fig. 6. Surface density profile for our numerical model of AS 209 for a locally isothermal and a radiative equation of state after 5000 orbits. The differences in the inner disk as well as the gap region are clear between the two models. Vertical dotted lines mark location candidates for dust gaps based on the formula determined by Zhang et al. (2018), and the dashed line marks the planet location.

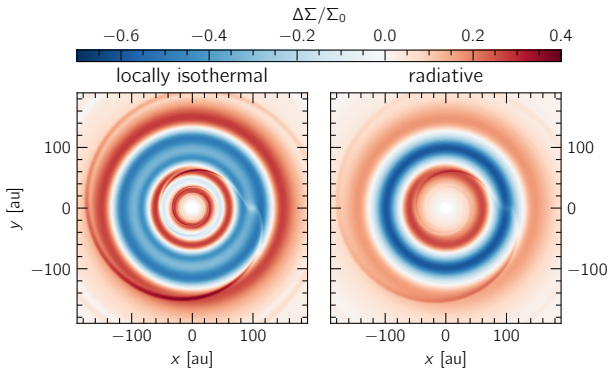

Fig. 7. Similar to Fig. 3, but for AS 209. The sharper annular and angular structures in the locally isothermal model are clearly visible.

HD 163296 (at $r = 48$ au) for our simulation domain (Huang et al. 2018). While reproducing all five gaps might be difficult with an aspect ratio $h_p \approx 0.08$, we see substantial differences between the two setups not only in the number of producible gaps, but also in the shape of the primary gap around the planetary orbit.

Our results are plotted in Figs. 6 and 7 and match our findings for HD 163296 in that the radiative model does not agree with the locally isothermal one. A single planet at 99 au reproduces at least three gaps at 47, 84 and 115 au in the locally isothermal model (with a possible fourth gap at 22 au, which is however subject to boundary effects), whereas the radiative model produces a very smooth monotonic profile with no gaps except for one exactly at the planet location (Fig. 6). In addition, the primary gap structure is entirely different between the two models, with the locally isothermal model showing a wider gap region that contains more material at the orbital radius of the planet, however. This can be seen in both figures, and is discussed in detail in Sect. 4.

4. Discussion

In Sect. 3 we showed that modeling a protoplanetary disk with a locally isothermal equation of state can under certain circumstances prove incorrect because this assumption might affect the planet–disk interaction process even when the temperature profile is largely unaffected by the choice of equation of state. By comparing against a radiative simulation where viscous and

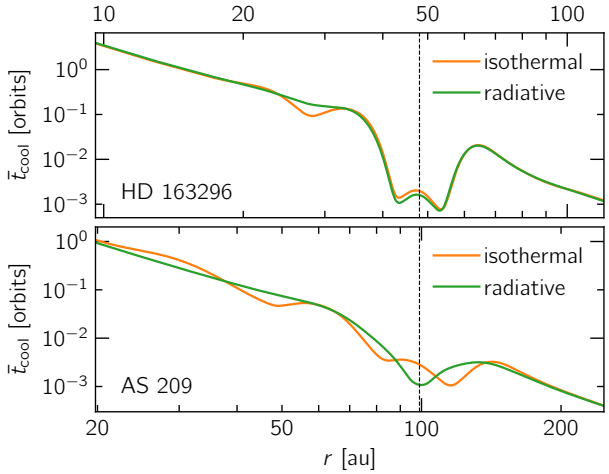

Fig. 8. Azimuthally averaged cooling timescale in units of the local orbital period of the disk after 1000 and 5000 planetary orbits for HD 163296 and AS 209, respectively. Differences between locally isothermal and radiative models become clear in the inner disk, where the cooling timescale corresponds to a non-negligible fraction of the orbital period (roughly 1–10% near the planet, and comparable to the orbital period farther in). In the outer disk, cooling is sufficiently fast so that the two models overlap as far as azimuthally averaged profiles are concerned.

irradiation heating and thermal cooling are included, we showed that radiative effects play an important role in the disk evolution even at large distances from the source star.

Our testbeds aimed at reproducing the inner 50 au structure of the system HD 163296 and the gap-rich structure of AS 209. At distances of 48 au for HD 163296 and 99 au for AS 209 (which is where planets are suspected to be), the disk is strongly dominated by irradiation, whereas viscous and spiral heating effects are negligible. Even so, the cooling timescale at the range of 20–100 au in these systems is still significant enough to invalidate the locally isothermal assumption.

We can estimate the cooling timescale as

$$\frac{\partial \Sigma e}{\partial t} \sim \frac{\Sigma e}{t_{\rm cool}} \sim Q_{\rm cool} \Rightarrow t_{\rm cool} \propto \frac{\tau_{\rm eff}\Sigma}{T^3}. \tag{9}$$

We then compare $t_{\rm cool}$ to the local orbital period $P_{\rm orb} = 2\pi/\Omega_{\rm K}$ in Fig. 8. We find that $t_{\rm cool}$ corresponds to 2–20% of $P_{\rm orb}$ in the 20–50 au range in the absence of a planet for HD 163296, and 1–10% in the 40–100 au range for AS 209. These fractions are clearly significant in this context, implying that the angular momentum flux driven by the spiral waves of a planet is described by an adiabatic framework (Miranda & Rafikov 2019).

At the same time, we see that the cooling timescale drops below 0.1% of the orbital period at distances of ∼100 au for HD 163296, or ∼200 au for AS 209 (see Fig. 8). At these distances we find that the locally isothermal model approaches the radiative one both in terms of radial disk structure and spiral arm contrast, suggesting that the locally isothermal assumption can still be justified when the cooling timescale is short enough. This is consistent with the latest results by Miranda & Rafikov (2020), who suggested that a cooling timescale of about 10^{-3}–10^{-2} orbital periods is sufficiently short to match locally isothermal disks for low- and high-mass planets, respectively. Regarding this last statement, the different gap structure between equations of state around the planetary orbit in AS 209 can be

explained by comparing the planet mass to the thermal mass of the disk (Zhu et al. 2015)

$$M_{\rm th} \equiv \frac{c_{\rm s}^3}{G\Omega_{\rm p}} = h_{\rm p}^3 M_* \approx 1\, M_J \left(\frac{h_{\rm p}}{0.1}\right)^3 \left(\frac{M_*}{M_\odot}\right). \tag{10}$$

Using this formula, we find that a $0.5\,M_J$ planet in HD 163296 corresponds to $1.3\,M_{\rm th}$, and a $0.083\,M_J$ planet in AS 209 to $0.18\,M_{\rm th}$. This means that the former system is prone to nonlinear effects due to gap opening, and therefore the surface density profiles around the primary gap are largely the same between the two different equations of state. On the other hand, our numerical model of AS 209 hosts a planet that falls in the linear regime, and therefore a cooling time of about 1% of the orbital period is not short enough to allow the radiative and locally isothermal models to agree with each other. Instead, the two profiles show a substantially different structure in the region that corotates with the planet, and they only overlap at the outskirts of the disk (where planet-driven effects are weak and/or damped anyway).

To verify that the locally isothermal profile can indeed be recovered with a more efficient cooling prescription, we executed several additional radiative simulations of HD 163296 where we limited adiabatic effects ($\gamma = 1.01$) or artificially reduced the cooling timescale (using $\kappa = 0.045$ cm^2 g^{-1}), thereby limiting radiative effects. These results are shown in Appendix A.

While this exercise shows that the inner disk is prone to radiative effects in our simulations, this result should be taken with a grain of salt when comparing to observations. Both the optical depth and surface density (which depend on the opacity and initial conditions) are uncertain and not set in stone. Additionally, if a substantial part of the dust grains has grown into larger grains, the opacity (and in turn, the optical depth) could be much lower, reducing the cooling timescale.

Our results on HD 163296 can also be compared to the 2D locally isothermal simulations by Zhang et al. (2018). In Fig. 2 of their study, they show that a planet with $M_{\rm p} = 0.3\,M_J$ ($q = 2.9 \times 10^{-4}$, in our case $q = 2.3 \times 10^{-4}$) can open a secondary gap at roughly $0.6\,r_{\rm p}$ (28 au for our single-planet model of HD 163296) when the aspect ratio $h_{\rm p}$ is between 0.05 and 0.07 and scales with $r^{1/4}$. Because our computed aspect ratio lies within the profiles generated using those values (see Fig. 2), we expect exactly (or at least) one secondary gap in our locally isothermal simulation. The fact that we indeed observe it suggests that our results agree with Zhang et al. (2018) in the locally isothermal limit. In addition, we observe the secondary gap at the location inferred by the fitting formula suggested in that study. However, we used a steeper surface density profile than they did ($\Sigma_0 \propto r^{-3/2}$ as opposed to r^{-1}). We carried out two more simulations with a shallower surface density profile and found similar results (see Appendix A).

By taking into account the above points, we see that the secondary-gap-opening capabilities of a single planet are exaggerated in a locally isothermal disk, such that a secondary gap in gas surface density can form within that framework but not always when radiative effects are properly treated. In the case of HD 163296 this could constrain the time of planet formation because a ring at 34 au is indeed not visible (Huang et al. 2018). This assumes that our estimates of α among other model parameters are viable.

Our simulations do not include a dust component, meaning that we cannot verify whether this secondary gap generated in the locally isothermal simulation is visible. Nevertheless, we expect that large grains can be trapped in the pressure bump

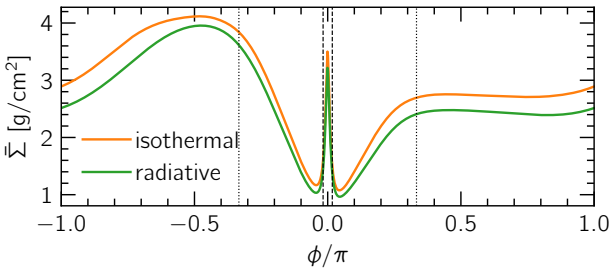

Fig. 9. Azimuthal cut at $r = r_{\rm p} = 48$ au. The surface density trailing behind the planet (left side, $\phi < 0$) is slightly higher than that at the leading Lagrange point ($\phi > 0$). The dashed black lines mark the planet's hill radius and contain 20 cells. The dotted lines mark the location of the L4 (left) and L5 (right) Lagrange points.

formed at 34 au, forming a ring and therefore increasing the contrast between the inner disk and the secondary gap region. A similar argument can be used for the crescent structure within the gap that can be seen in the ALMA observations of HD 163296. We find that the perturbed gas surface density at the trailing Lagrange point is higher than that at the leading one (see Fig. 9), and we suggest that dust–gas interaction could collect grains at the L4 point. This feature, however, should be transient as the gap region continues to empty as the disk evolves.

The system HD 163296 shows more structure between 50 and 200 au, namely a clear second gap in continuum emission at 86 au and a third, slightly less visible gap at 145 au. Assuming the existence of planets at all three locations, we can justify our results as far as annular structures in the 10–50 au range are concerned because the planet at 48 au will shield the inner disk by opening a gap and therefore halt the propagation of spiral waves by the outer planet(s).

Nevertheless, it would be interesting to include more planets and model the 10–200 au range of this system, assuming that each gap in dust continuum stems from a single planet opening a corresponding gap in gas density. This is further motivated by the kinematic detection of Jupiter-sized planets at 83 and 137 au by Teague et al. (2018). We therefore carried out a simulation with three planets at 48, 86, and 145 au, respectively, and find similar results in the 10–50 au range. These results are shown and discussed in Appendix B.

There are several other sources for which ALMA has provided high-fidelity observational datasets. As far as ring structures are concerned, one system stands out: AS 209 (Andrews et al. 2018) shows at least five rings with as many gaps in dust continuum emission (Huang et al. 2018). Numerical models strongly suggest that one or more growing planets are responsible for their formation (e.g., Fedele et al. 2018).

It has been suggested that a single planet at 99 au might be able to carve most of these gaps (Zhang et al. 2018). We therefore found it useful to examine the importance of radiative effects on a system with such a rich annular structure, and carried out a comparison similar to that for HD 163296 against a locally isothermal simulation, shown in Sect. 3.2. Strikingly, while our locally isothermal model produces several rings throughout the disk, we find that this is not the case in the radiative model, which shows no ring structure except for the unavoidable pressure bump in the outer disk that is caused by the clearing of the planet's corotating region. Our results suggest that we would need multiple planets to explain the ring–gap structure in such systems.

5. Conclusions

Our goal was to examine the importance of radiative effects in disk evolution with regard to the gap-opening capabilities of a single planet. To this end, we carried out 2D numerical simulations of planets embedded in disks that resemble two systems that have recently been imaged in high angular resolution: HD 163296 and AS 209. We then compared locally isothermal simulations, where a fixed radial sound speed profile is prescribed, to setups where radiative effects are self-consistently treated.

We found that locally isothermal models exaggerate the contrast of planet-generated features with respect to radiative models and therefore overestimate the planet's ability to carve a secondary gap in its disk. This is consistent with previous results and implies that a single planet cannot always explain the existence of multiple gaps.

We also found that the contrast of spiral arms launched by a planet is artificially sharpened within a locally isothermal framework. While this phenomenon is weak or negligible in the optically thinner outer disk, it becomes more significant for inner spirals. Regardless, spiral shock heating is negligible, with spirals having a low temperature contrast with the background and the disk being sufficiently optically thin.

Finally, by running a simulation of HD 163296 over an extended range that contained three planets, we found that our results in the limited range of 10–60 au (i.e., inside and around the innermost planet) remain unchanged. We also showed that an interplay between planet mass and the cooling timescale can lead to slight differences between a locally isothermal and an adiabatic equation of state even at 145 au, although viscous evolution might render such differences negligible.

In conclusion, the locally isothermal assumption proves to be dangerous even at the range of tens of au regarding planet–disk interaction and should therefore be in general avoided in favor of an adiabatic equation of state with a prescription for radiative cooling in the disk. By estimating the cooling timescale $t_{\rm cool}$, the usage of such an assumption in a regime where cooling occurs very rapidly compared to the orbital period $P_{\rm orb}$ might be justified. This corresponds to $t_{\rm cool}$ at least shorter than 1% of $P_{\rm orb}$ for massive planets, or 0.1% for low-mass planets, in good agreement with Miranda & Rafikov (2020).

Acknowledgements. C.P.D. and W.K. acknowledge funding from the DFG research group FOR 2634 "Planet Formation Witnesses and Probes: Transition Disks" under grant DU 414/22-1 and KL 650/29-1, 650/30-1. The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG. All plots in this paper were made with the Python library `matplotlib` (Hunter 2007).

References

Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, ApJ, 869, L41
Birnstiel, T., Dullemond, C. P., Zhu, Z., et al. 2018, ApJ, 869, L45
Crida, A., Morbidelli, A., & Masset, F. 2006, Icarus, 181, 587
de Val-Borro, M., Edgar, R. G., Artymowicz, P., et al. 2006, MNRAS, 370, 529
Dullemond, C. P., Birnstiel, T., Huang, J., et al. 2018, ApJ, 869, L46
Fedele, D., Tazzari, M., Booth, R., et al. 2018, A&A, 610, A24
Haffert, S. Y., Bohn, A. J., de Boer, J., et al. 2019, Nat. Astron., 3, 749
Huang, J., Andrews, S. M., Dullemond, C. P., et al. 2018, ApJ, 869, L42
Hubeny, I. 1990, ApJ, 351, 632
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Keppler, M., Benisty, M., Müller, A., et al. 2018, A&A, 617, A44
Lin, D. N. C., & Papaloizou, J. 1985, Protostars and Planets II, eds. D. C. Black & M. S. Matthews (Tucson, AZ: University of Arizona Press), 981s
Lyra, W., Richert, A. J. W., Boley, A., et al. 2016, ApJ, 817, 102
Masset, F. 2000, A&AS, 141, 165
Menou, K., & Goodman, J. 2004, ApJ, 606, 520
Mignone, A., Bodo, G., Massaglia, S., et al. 2007, ApJS, 170, 228
Mignone, A., Flock, M., Stute, M., Kolb, S. M., & Muscianisi, G. 2012, A&A, 545, A152
Miranda, R., & Rafikov, R. R. 2019, ApJ, 878, L9
Miranda, R., & Rafikov, R. R. 2020, AJ, 892, 1
Müller, T. W. A., & Kley, W. 2012, A&A, 539, A18
Müller, T. W. A., Kley, W., & Meru, F. 2012, A&A, 541, A123
Muto, T., Grady, C. A., Hashimoto, J., et al. 2012, ApJ, 748, L22
Ogilvie, G. I., & Lubow, S. H. 2002, MNRAS, 330, 950
Rafikov, R. R. 2002, ApJ, 569, 997
Rafikov, R. R. 2016, ApJ, 831, 122
Richert, A. J. W., Lyra, W., Boley, A., Mac Low, M.-M., & Turner, N. 2015, ApJ, 804, 95
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Tassoul, J.-L. 1978, Theory of Rotating Stars (Princeton: Princeton University Press)
Teague, R., Bae, J., Bergin, E. A., Birnstiel, T., & Foreman-Mackey, D. 2018, ApJ, 860, L12
Thun, D., & Kley, W. 2018, A&A, 616, A47
Zhang, S., & Zhu, Z. 2020, MNRAS, 493, 2287
Zhang, S., Zhu, Z., Huang, J., et al. 2018, ApJ, 869, L47
Zhu, Z., Dong, R., Stone, J. M., & Rafikov, R. R. 2015, ApJ, 813, 88
Ziampras, A., Ataiee, S., Kley, W., Dullemond, C. P., & Baruteau, C. 2020, A&A, 633, A29

Appendix A: Approaching the locally isothermal limit in HD 163296

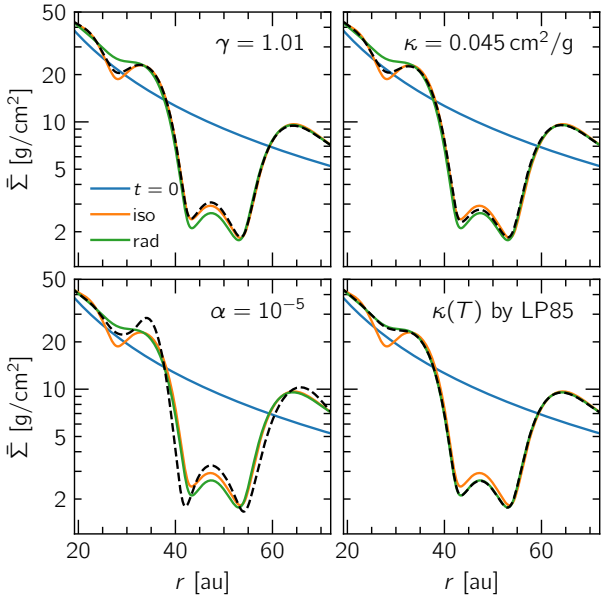

Fig. A.1. Azimuthally averaged surface density profiles after 1000 orbits for several comparison tests, attempting to recover the locally isothermal limit by constraining radiative effects. The orange and green curves correspond to our fiducial locally isothermal and radiative models, and the change being tested in each panel is plotted with a black dashed line. An adiabatic equation of state with $\gamma = 1.01$ or $\kappa = 0.045\,\mathrm{cm^2/g}$ (instead of 1.4 and $0.45\,\mathrm{cm^2/g}$, respectively) roughly reconstructs the gap structure of the locally isothermal model and could allow the formation of a shallower secondary gap. A model with lower viscosity ($\alpha = 10^{-5}$ instead of 10^{-4}) also shows a secondary gap, but differences in the width of the primary gap and in the outer disk are also visible. The opacity model by Lin & Papaloizou (1985) yields no observable difference.

To verify whether the locally isothermal results for HD 163296 concerning the primary and secondary gaps can be recovered by appropriately tweaking radiative effects, we tested several cases where we either restrained changes in internal energy by setting $\gamma = 1.01$, or amplified the contribution of irradiation and cooling by lowering the opacity to $\kappa = 0.045\,\mathrm{cm^2\,g^{-1}}$. A detailed description of the model parameters used here is given in Table 1, Cols. 2 and 3.

Our results, plotted in Fig. A.1, show that locally isothermal conditions can be emulated to an extent by manipulating the contribution of the energy equation altogether through the adiabatic index, or controlling the cooling timescale through the opacity (Eq. (9) suggests that $t_{\rm cool} \propto \tau \propto \kappa$). It might be possible to completely recover the locally isothermal limit by further constraining $\gamma \to 1$ or artificially lowering the opacity even further.

In addition, to test the planet's secondary-gap opening capabilities under conditions that could potentially lead to the reemergence of the secondary gap, as well as to verify our opacity model of choice, we executed a few more pairs of locally isothermal and radiative simulations. We set $\alpha = 10^{-5}$ in the first pair to facilitate the gap opening process. We find that this does lead to the surface density contrast (and the pressure bump) reappearing in the inner disk, but at the same time, it also results in a wider gap and a different outer disk profile (see Fig. A.1). This value of α is extremely low and gap opening is known to be easier when the viscosity is low (e.g., Crida et al. 2006), therefore we argue that this adds nothing new to our results.

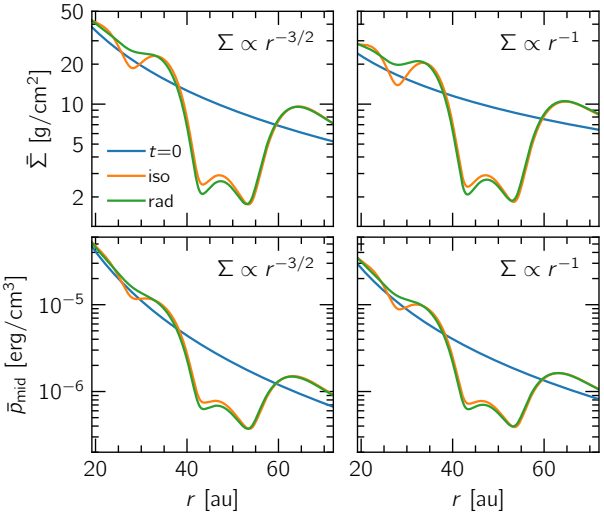

Fig. A.2. Azimuthally averaged surface density and midplane pressure profiles after 1000 orbits for two comparison tests where disks with different surface density profiles are simulated. *Right panels*: a shallower surface density profile ($\Sigma_0 \propto r^{-1}$ instead of $\propto r^{-3/2}$) is used.

In the second pair, we compared our constant-opacity models and the analytical opacity model by Lin & Papaloizou (1985), which dictates a relation $\kappa \propto T^2$ for $T < 170\,\mathrm{K}$ (i.e., within our simulation domain). We find that the choice of opacity model has little to no effect as far as the secondary-gap opening capabilities of the planet are concerned (Fig. A.1). This makes sense because irradiation heating and thermal cooling dominate the energy equation, and equating these two together factors the opacity out given our prescription in Eq. (2) while still yielding a similar value for the cooling timescale at the 20–40 au region.

Finally, in the third pair we prescribed a shallower initial surface density profile $\Sigma(r) \approx 10\,(r/r_{\rm p})^{-1}$ in an attempt to reduce the cooling timescale in the inner disk while preserving the conditions in the planet's vicinity. In doing so, we can also compare against the results from locally isothermal simulations by Zhang et al. (2018). The results are summarized in Fig. A.2 and paint a picture similar to that for the steeper, $\Sigma(r) \propto r^{-3/2}$ profile. However, because the cooling timescale is shorter for the shallower Σ profile (by a factor of 30 and 40% at the location of the secondary pressure bump and gap, respectively), the disk acts more "locally isothermally" and the secondary density bump is still visible after 1000 orbits but a pressure minimum does not form in the midplane. At the same time, pressure torques near the secondary gap edge are weaker for the shallower profile; this effect assists the gap-opening process. A detailed description of the model parameters used in these models is given in Table 1, Cols. 4–6.

A&A 637, A50 (2020)

Appendix B: Modeling HD 163296 with three planets

Table B.1. Parameters used for the model of HD 163296 containing three planets, compared to the base model of the same system.

Parameter	Base	Three planets
M_* [$M_\odot$]	2.089	–
L_* [$L_\odot$]	16.98	–
M_p [M_J]	0.5	0.5/0.5/0.5
M_p [10^{-3}_*]	0.29	0.29/0.29/0.29
M_disk [% $M_\odot$]	5.58	9.4
M_disk [% M_*]	2.67	4.5
r_p [au]	48	48/86/145
$\Sigma_{t=0}(r)$	$\propto r^{-3/2}$	–
Σ_p [g cm^{-2}]	9.61	9.61/4.01/1.83
h_p [%]	5.601	5.601/6.617/7.684
T_p [K]	34.5	34.5/26.9/21.5
γ	1.4	–
κ [cm^2 g^{-1}]	0.45	–
$\log\alpha$	–4	–
r_min–r_max [au]	9.6–240	9.6–576
$N_r \times N_\phi$	573 × 1118	800 × 1561
$N_\mathrm{cells}/H_\mathrm{p}$	10	11/13/15

Notes. Dashes imply that a parameter is inherited from the base model.

As mentioned in Sect. 4, we conducted an additional simulation of HD 163296 with three planets at 48, 86, and 145 au to examine the structure of the system on a larger scale, more easily comparable to the DSHARP observation (see Fig. 1, left panel). To do so, we employed a fourth-order Runge–Kutta N-body integrator (Thun & Kley 2018) that allows the planets to interact gravitationally and accounts for the noninertial term that arises due to centering our system around $r = 0$ instead of the barycenter of the system.

In our model, all three planets have the same mass of 0.5 M_J. Because the thermal mass of the disk scales purely with the aspect ratio at a planet's location, we expect a damping of nonlinear effects as we slowly transition to the linear regime upon moving farther out in the disk. More specifically, the ratio $M_\mathrm{p}/M_\mathrm{th}$ is 1.3, 0.8, and 0.5 for the three planets in ascending distance from the star. Because the cooling timescale at 86 and 145 au is 0.33 and 0.06% of the local orbital period, respectively, it is possible that we can still observe deviations between the locally isothermal and radiative models around the planets' corotating regions because we now study the low-mass regime.

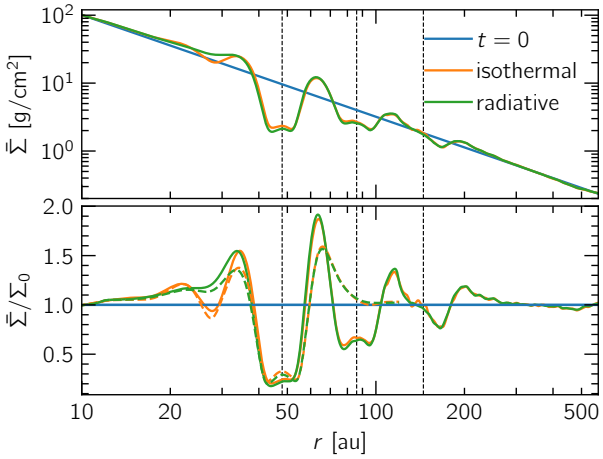

Fig. B.1. Azimuthally averaged surface density profile after 1000 orbits at 48 au (417 at 86 au, 190 at 145 au) for the three-planet model (see Table B.1). The colored dashed lines refer to the base model with a single planet at 48 au. Vertical black lines show the location of planets in the disk.

The resulting surface density profile is plotted in Fig. B.1. As expected, the two models show better overall overlap with increasing distances from the star. Nevertheless, small deviations are still visible within the corotating regions of the outer planets. We rationalize this outcome both through the cooling timescale argument and by considering how the interaction between a planet and the disturbances caused by other planets (e.g., spirals) could affect their corotating regions. It should be noted, however, that the viscous timescale at 145 au is roughly three times longer than that at 48 au, so it is possible that differences between equations of state can emerge on much longer timescales, which might not be reasonable for such a young system.

Finally, by comparing the 10–60 au range of the three-planet simulation against that of our base model with a single planet (lower panel of Fig. B.1) we find that the multi-planet system shows stronger perturbations in surface density around the secondary gap at 28 au, but still fails to form a pressure maximum at 34 au in the radiative model. This helps support the relevance of our base model, which targeted a limited range of HD 163296.

C.3 Rometsch et al. (2021)

Astronomy & Astrophysics manuscript no. paper
September 7, 2021

Survival of planet-induced vortices in 2D disks

Thomas Rometsch, Alexandros Ziampras, Wilhelm Kley, and William Béthune

Institut für Astronomie und Astrophysik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany
e-mail: {thomas.rometsch, alexandros.ziampras}@uni-tuebingen.de

September 7, 2021

ABSTRACT

Context. Several observations of protoplanetary disks display non-axisymmetric features, often interpreted as vortices. Numerical modeling has repeatedly shown that gap-opening planets are capable of producing large and long-lasting vortices at their outer gap edge, making massive planets popular candidates as the source of such features.
Aims. We explore the lifetime of vortices generated by Jupiter-sized planets as a function of the thermal relaxation timescale, the level of turbulence, and the effect of disk self-gravity.
Methods. We conduct 2D numerical simulations using the hydrodynamics codes PLUTO and FARGO, scanning through several physical and numerical parameters. Vortex properties are automatically extracted from thousands of simulation snapshots.
Results. We find that vortices that spawn at the outer gap edge can survive for about 100–3000 planetary orbits for typical disk parameters, where the shortest lifetimes occur for a moderate efficiency of dissipation and cooling. However, we also observe a different regime of long-lasting vortices with lifetimes of at least 15 000 orbits for very low viscosity and very short thermal relaxation timescales. Disk self-gravity significantly shortens the lifetime of regular vortices but still allows long-lived ones to survive.
Conclusions. Our results suggest that the cooling timescale plays an important role in vortex formation and lifetime and that planet-generated vortices should be observable at large distances from the star for typical thermal relaxation timescales and turbulence levels.

Key words. protoplanetary disks – planet–disk interaction – hydrodynamics – methods: numerical

1. Introduction

Planets are born and grow in accretion disks around young stars. This is supported by observations of protoplanets embedded in a disk of gas and dust captured during their growth phase (e.g. Keppler et al. 2018). A protoplanet interacts with the disk around it in every stage of its growth (Kley & Nelson 2012) for example via exchange of angular momentum. This results in the launching of spiral arms (Ogilvie & Lubow 2002) and, if the planet is massive enough, the opening of a gap and, in some cases, the formation of multiple rings around the planet's orbit (Rafikov 2002). The number of spirals, gaps, and rings as well as their contrast scales with the planet's mass, such that Jupiter-sized planets can have a strong impact on their environment in the right conditions, possibly resulting in multiple ring-like and non-axisymmetric observable features (Zhang & Zhu 2020; Miranda & Rafikov 2020a). This makes the planet–disk interaction scenario a popular interpretation for the numerous high-fidelity ALMA observations of such features.

One promising scenario to explain observational asymmetries is the existence of vortices because they naturally accumulate dust at the pressure maxima in their center (see for example Marel et al. 2013; Bae et al. 2016; Pérez et al. 2018; Hammer et al. 2019; Barge & Sommeria 1996).

Among the various ways to form vortices, the Rossby-wave instability (RWI, Lovelace et al. 1999) is particularly relevant in the vicinity of gaps. The RWI readily happens in 2D disks at the outer and inner edge of planet-opened gaps (Li et al. 2005; Val-Borro et al. 2007). Additional mechanisms that could be relevant in this context are the subcritical baroclinic instability (SBI, Klahr & Bodenheimer 2003; Lesur & Papaloizou 2010) and the zombie-vortex instability (ZVI, Marcus et al. 2015, 2016). Vortices are then susceptible to viscous spreading as well as secondary instabilities such as the elliptical instability (Lesur & Papaloizou 2009), which cause vortex decay. The lifetime of vortices is therefore determined by a competition between vortex-forming and -decaying mechanisms.

Aside from possibly causing observable features in the disks, vortices can also affect planet migration in a stochastic fashion (McNally et al. 2019) and even cause temporary outward migration (Lega et al. 2021) for otherwise inwardly-migrating planets. Understanding their formation pathways and lifetimes is therefore critical to the modeling of planet migration using global, low-viscosity simulations.

In previous numerical studies, vortex properties have been found to depend on various physical processes such as turbulent viscosity and disk self-gravity. Lower viscosity allows vortices to live longer (Val-Borro et al. 2007) whereas the inclusion of self-gravity tends to weaken vortices, shortening their lifespan (Lin & Papaloizou 2011; Zhu & Baruteau 2016; Regály & Vorobyov 2017).

In recent numerical studies, radiative effects have been discovered to have a significant impact on the gap-opening capabilities of planets and therefore the structure of said gaps (Ziampras et al. 2020b; Miranda & Rafikov 2020b), possibly affecting the development of the RWI and by extension vortices around their edge. The present study aims to investigate the role of radiative effects for properties of vortices created by planets. More precisely, we explore how the thermal relaxation timescale of the gas affects the lifetime of vortices created during the growth of Jupiter-sized planets.

Recently, Fung & Ono (2021) ran 2D shearing box simulations of RWI-induced vortices. As their simulations did not include a planet, the RWI was triggered by an artificial density bump. They described a baroclinic effect that spins down vortices where the decay is fastest for thermal relaxation times of the order of a tenth of the vortex turnover time.

We ran a suite of global two-dimensional hydrodynamics simulations with an embedded Jupiter-sized planet, which naturally creates vortices in the disk, for different choices for the turbulent viscosity and the thermal relaxation timescale, among other physical parameters. The results of these simulations are then post-processed with our newly-developed pipeline for the detection and characterization of vortices.

In Sect. 2 we describe our physical model and numerical setup. We present a typical life track of a vortex in our models in Sect. 3, report the dependence of vortex properties on physical parameters in Sect. 4, and present the case of long-lived vortices in Sect. 5. We discuss and comment on our findings in Sect. 6. Finally, Sect. 7 contains a summary of our main results and our conclusions.

2. Physics and numerics

In this section, we describe the physical and numerical framework that we used in our simulations. We justify the approximations in our model, explain in detail the initialization process, and list technical parameters such as our grid setup and parameter space.

2.1. Hydrodynamics

We consider a thin disk of neutral, ideal gas with adiabatic index $\gamma = 7/5$ and mean molecular weight $\mu = 2.353$ that is orbiting around a star with one solar mass $M_\star = \mathrm{M}_\odot$. The two-dimensional, vertically integrated Navier-Stokes equations in a polar coordinate system $\{r, \phi\}$ read

$$\frac{\partial \Sigma}{\partial t} + \boldsymbol{u} \cdot \nabla \Sigma = -\Sigma \nabla \cdot \boldsymbol{u} \tag{1a}$$

$$\Sigma \frac{\partial \boldsymbol{u}}{\partial t} + \Sigma(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p + \Sigma \boldsymbol{g} + \nabla \cdot \boldsymbol{\sigma} \tag{1b}$$

$$\frac{\partial (\Sigma \varepsilon)}{\partial t} + \boldsymbol{u} \cdot \nabla (\Sigma \varepsilon) = -\gamma \Sigma \varepsilon \nabla \cdot \boldsymbol{u} + Q_\mathrm{visc} + Q_\mathrm{relax} \tag{1c}$$

where $\boldsymbol{u} = (u_r, u_\phi)$ and ε are the velocity and specific internal energy of the gas evaluated at the midplane, and Σ is the surface density. The vertically integrated pressure p is defined through the ideal gas law $p = (\gamma - 1)\Sigma\varepsilon = \mathrm{R_g}\Sigma T/\mu$, with $\mathrm{R_g}$ being the gas constant and T the gas temperature. The isothermal sound speed of the gas is then given by $c_\mathrm{s,iso} = \sqrt{p/\Sigma} = \sqrt{\mathrm{R_g} T/\mu}$ and relates to the adiabatic sound speed c_s as $c_\mathrm{s,iso} = c_\mathrm{s}/\sqrt{\gamma}$. For a disk in Keplerian motion and vertical hydrostatic equilibrium, we can also write $c_\mathrm{s,iso} = H\Omega_\mathrm{K}$, where $\Omega_\mathrm{K} = \sqrt{\mathrm{G}M_\star/r^3}$ is the Keplerian orbital frequency at radius r and H is the pressure scale height of the gas.

The viscous stress tensor $\boldsymbol{\sigma}$ (following Tassoul (1978)) appears in both the momentum equation (1b) as well as the dissipation function:

$$Q_\mathrm{visc} = \frac{1}{2\nu\Sigma}\mathrm{Tr}(\boldsymbol{\sigma}^2) = \frac{1}{2\nu\Sigma}\left(\sigma_{rr}^2 + 2\sigma_{r\phi}^2 + \sigma_{\phi\phi}^2 + \sigma_{zz}^2\right), \tag{2}$$

where $\nu = \alpha c_\mathrm{s} H$ is the kinematic viscosity parametrized according to the α-viscosity model of Shakura & Sunyaev (1973). Here, α is a parameter that captures both radial angular momentum transport that leads to accretion onto the star and heating of the disk due to viscous friction. Numerical simulations of (magneto)hydrodynamical instabilities such as the vertical shear instability (VSI, Nelson et al. 2013) or the magneto-rotational instability (MRI, Balbus & Hawley 1991) have provided numerical estimates of α, while observations of young stellar objects surrounded by disks have constrained these estimates (Dullemond et al. 2018). To probe a wide range of diffusion regimes from practically inviscid to moderately viscous, we choose $\alpha \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$ for our models.

Viscous dissipation leads to the heating of the disk. An embedded planet can also deposit significant amounts of thermal energy via the dissipation of spiral shocks (Rafikov 2016; Ziampras et al. 2020a). As a cooling solution, we allow the disk to relax to a prescribed temperature profile T_0 (see Eq. (5)) over a relaxation timescale $\tau_\mathrm{relax} = \beta/\Omega_\mathrm{K}$.

The thermal relaxation term appears as an additional source term to the energy equation

$$Q_\mathrm{relax} = -\Sigma c_\mathrm{v} \frac{T - T_0}{\beta}\Omega_\mathrm{K} \Rightarrow \frac{\partial T}{\partial t} = -\frac{T - T_0}{\tau_\mathrm{relax}} \tag{3}$$

where $c_\mathrm{v} = \frac{\mathrm{R_g}}{\mu(\gamma-1)}$ is the heat capacity of the gas at constant volume. The parameter β controls the relaxation timescale, as well as the overall planet–disk interaction process (Miranda & Rafikov 2020b), we choose the values $\beta \in \{0.01, 1, 100\}$ which correspond to very fast, moderate, and very slow relaxation.

The gravity of the star and planet are included as a source term in $\boldsymbol{g}$. We work in a star-centered coordinate system and embed a planet with mass M_p at a position $\boldsymbol{r}_\mathrm{p}$. Thus, the source term reads

$$\begin{aligned}\boldsymbol{g} &= \boldsymbol{g}_\star + \boldsymbol{g}_\mathrm{p} + \boldsymbol{g}_\mathrm{ind} \\ &= -\frac{\mathrm{G}M_\star}{r^3}\boldsymbol{r} - \frac{\mathrm{G}M_\mathrm{p}}{(d^2 + \epsilon^2)^{3/2}}\boldsymbol{d} - \frac{\mathrm{G}M_\mathrm{p}}{r_\mathrm{p}^3}\boldsymbol{r}_\mathrm{p}, \quad \boldsymbol{d} = \boldsymbol{r} - \boldsymbol{r}_\mathrm{p}.\end{aligned} \tag{4}$$

The terms $\boldsymbol{g}_\star$, $\boldsymbol{g}_\mathrm{p}$, and $\boldsymbol{g}_\mathrm{ind}$ denote the acceleration due to the star, the planet, and the indirect term which is a correction needed because the star-centered frame is not an inertial frame. Disk feedback on the star and planet is neglected. The planet's gravitational pull (2nd term in the RHS of Eq. (4)) is smoothed using a Plummer potential with a smoothing length $\epsilon = 0.6H(\boldsymbol{r})$ that captures the effect of the vertical structure of a more realistic 3D disk (Müller et al. 2012) and prevents singularities near the planet's location.

For simplicity, we do not allow the planet to migrate. We chose to limit the degrees of freedom in our model to focus on the dynamics of the vortex and avoid complex and potentially chaotic interplay of the vortex with the planet (Lega et al. 2021). For the same reason of simplicity, we neglect planetary accretion in our models.

2.2. Numerics

We use two different codes for our numerical models: `PLUTO 4.2` (Mignone et al. 2007), a finite-volume, energy-conserving, shock-capturing code that treats transport by solving the Riemann problem across the interfaces of adjacent cells in both directions (r, ϕ) in an unsplit fashion, and our custom `FARGO` (Masset 2000) version, `FargoCPT` (Rometsch et al. 2020), which

uses a finite-difference, dimensionally-split, second-order upwind method for gas advection. Both codes utilize the FARGO method (implemented into PLUTO by Mignone et al. 2012), in which orbital advection is essentially performed via the Keplerian rotation on top of which the code solves for the residual velocity deviations, significantly relaxing timestep limitations and reducing numerical dissipation in the process (Masset 2000).

The inherent differences between the two numerical schemes make it worthwhile to carry out our simulations using both codes, to verify the robustness of our results and test for numerical convergence. Namely, the strictly energy-conserving nature of PLUTO and the necessity for artificial viscosity to stabilize FargoCPT are discussed in more detail in Sect. 6.7, among others.

2.2.1. Grid setup

Our computational domain spans the full azimuthal extent and a radial range of $r \in [0.2, 5.0]\, r_\mathrm{p} = [1.04, 26.0]$ au, with square cells logarithmically spaced so that the cell aspect ratio is preserved. After carrying out a thorough investigation on the effects of our numerical resolution of the recovery of both radial and azimuthal features caused by the planet, we decide to execute our simulations using a resolution of 8 and 16 cells per scale height (hereafter "cps") in both directions (r, ϕ). At this resolution, the two codes reach good convergence in terms of the presence and contrast of features shaped by the planet and results agree between the codes. This translates to a fiducial resolution of $(N_r, N_\phi) = (528, 1024)$ cells for 8 cps, or $(1056, 2048)$ cells for 16 cps. In addition, using the same resolution in both directions in terms of cps ensures that the effects of numerical viscosity are isotropic (see Appendix A).

2.2.2. Initial and boundary conditions

Our disk is initially axisymmetric and in equilibrium in the radial direction, such that the initial radial velocity profile results a constant accretion rate through the disk. The azimuthal velocity is close to the Keplerian profile, with the correction due to the radial pressure gradient. The initial surface density and temperature profiles are simple power laws such that

$$\Sigma_0(r) = 222\,\mathrm{g/cm^2} \left(\frac{r}{r_\mathrm{p}}\right)^{-1/2}, \quad T_0(r) = 120.7\,\mathrm{K} \left(\frac{r}{r_\mathrm{p}}\right)^{-1}, \tag{5}$$

with $r_\mathrm{p} = 5.2$ au. This temperature profile translates to a disk with a constant aspect ratio $h(r) = H/r = 0.05$. While the general consensus is that protoplanetary disks are flared (i.e., the aspect ratio increases with distance, see for example Dullemond (2000)), we choose to use a constant aspect ratio since the behavior and lifetime of vortices depends on this quantity (Hammer et al. 2021). Thus, we can isolate the dependence of vortices on the physical and numerical parameters in our suite of simulations.

The radial and azimuthal velocity components at $t = 0$ are then

$$u_r(r) = -\frac{3}{2}\frac{\nu}{r}, \quad u_\phi(r) = r\Omega_\mathrm{K}\sqrt{1 - 1.5\,h^2} \tag{6}$$

Near the boundaries, within the radial extent $r \in [0.2, 0.25] \cup [4.2, 5.0]\, r_\mathrm{p}$, the surface density and velocity are both damped to their initial profiles (see Eqs. (5), (6)) using the method of De Val-Borro et al. (2006), over a damping timescale of 0.3 periods at the respective boundary. While the radial boundary edges are closed, this minimizes the reflection of spiral waves back into the computational domain. The boundaries are periodic in the azimuthal direction.

We then embed a Jupiter-sized planet ($M_\mathrm{p} = 1\mathrm{M_J} = 10^{-3} M_\star$) in most models, with some simulations instead containing a less massive planet of $M_\mathrm{p} = 0.5\,\mathrm{M_J}$. To smoothly introduce the planet into the disk, we typically allow the planet to grow over 100 orbits at r_p using the formula by De Val-Borro et al. (2006). The importance of the growth timescale and planet mass will be discussed in Sect. 4.

2.3. Vortex Detection

We use the gas vortensity

$$\varpi = \frac{(\nabla \times \boldsymbol{u}) \cdot \hat{z}}{\Sigma} \tag{7}$$

as a proxy to detect and track the evolution of vortices over hundreds of snapshots for every model. Since these vortices consist of anticyclonic motion, the center of a vortex corresponds to a local minimum in vorticity, $\omega = (\nabla \times \boldsymbol{u}) \cdot \hat{z}$. Because vortices tend to accumulate mass towards their center and Σ is enhanced inside the vortex, the transition from the background flow to the vortex region is stronger and the vortex is more easily identified in a map of ϖ than in the case of ω alone.

More precisely, we use the gas vortensity normalized by the background vortensity from the initial conditions, $\varpi_0 = (\nabla \times \boldsymbol{u}_\mathrm{K}) \cdot \hat{z} \,/\, \Sigma_0$. This eliminates the radial dependence of the Keplerian velocity and the disk's surface density and ensures that our vortex proxy quantity, ϖ/ϖ_0, is of order unity everywhere in the disk except for the gap region, due to its very low surface density. The quantity ϖ/ϖ_0 usually varies between -1 for strongly counter-rotating vortices to 1 for the background flow.

We use our new Python module, called Vortector, that extracts iso-vortensity contours using the computer vision library OpenCV (Bradski 2000) to detect vortex candidates and then fits a 2D Gaussian to the vortensity and surface density data. The FWHM (or 2.355σ) of this Gaussian is used to define the radial and azimuthal extent of a vortex. Using this method, we also extract information about the shape of the vortex, including its radial and azimuthal extent and the mass it encloses. A more detailed description can be found in Appendix B.

This automated process has the drawback of sometimes producing detection artifacts, as can be seen for example on Fig. 6 below (top panel, dashed orange line), such that the vortex size (and thus its mass) is overestimated near the end of its lifetime as it blends into the disk background. While this effect is partly counteracted by using a median filter in time, we do not manually edit the output of the Vortector on a model-by-model basis.

In the following three sections, we present the results of our simulations. First, we present a typical example of vortex formation and evolution (Sect. 3). We then go on to describe the dependence of vortices on physical parameters for the group of vortices with short and intermediate lifetime (Sect. 4). Finally, long-lived and migrating vortices are presented (Sect. 5).

3. Typical life track of a vortex

The Jupiter-sized embedded planet opens a deep gap in all of our simulations. Fig. 1 shows maps of Σ (left) and ϖ (right) normal-

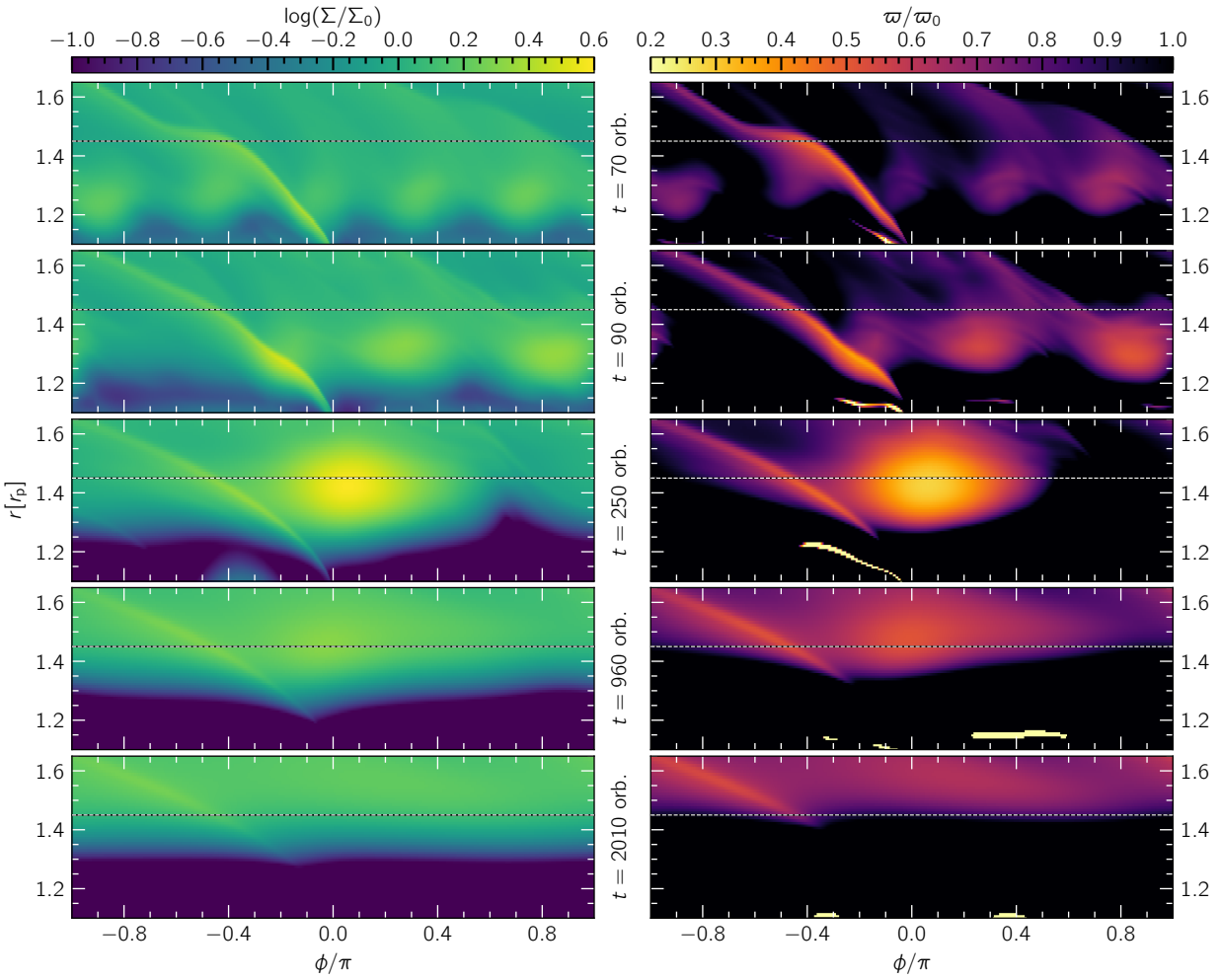

Fig. 1. Multiple snapshots of the $\alpha = 10^{-5}$, $\beta = 1$, 8 cps model showcasing the vortex merging process during the early stage of gap opening, the resulting vortex's fully grown size and its subsequent decay. The surface density and vortensity contrast compared to their initial profiles is shown on the left and right panels, respectively. Time is quoted in units of planetary orbits. The horizontal line at $r = 1.45\,r_p$ serves to highlight the outward radial movement of those structures as the gap around the planet grows wider. The planet is located at $r = 1\,r_p$ and $\phi = 0$.

ized by their initial values at five timestamps during the vortex lifetime for a model with $\alpha = 10^{-5}$, $\beta = 1$, and a resolution of 8 cps performed with the `FargoCPT` code. Horizontal dotted lines at $r = 1.45\,r_p$ are superimposed as a reference marking the final location of the vortex center. Here, four small-scale vortices (top row) first merge into two slightly larger vortices (second row) and then finally into one massive vortex (middle) that will last for a little over 1100 orbits. The vortex slowly decays over time, maintaining a large size (fourth row). In the later stages, the vortex is no longer present anymore (bottom row). The non-axisymmetric structure still visible exists due to the planet's spiral arm and is corotating with the planet.

During the early gap opening process, the outer gap edge grows Rossby-wave unstable (Lovelace et al. 1999) and several small-scale vortices form around it (top two rows). Fig. 2 shows radial profiles of the Lovelace parameter, $\mathcal{L}$, (top) and Σ (bottom) at different timestamps during the vortex formation up until $t =$ 180 orbits. The Lovelace parameter is defined as

$$\mathcal{L} = \frac{S^{2/\gamma}}{\varpi} \tag{8}$$

with the entropy $S = P/\Sigma^\gamma$. The development of a maximum in $\mathcal{L}$, which is one condition for the onset of the RWI, is visible. Vertical lines at the center of the $\mathcal{L}$ maxima (as determined by eye) are added to both panels to guide the eye for a comparison of the location of the maxima in $\mathcal{L}$ and Σ at each timestamp. The $\mathcal{L}$ maxima are located on the slope of the gap edge slightly inward of the Σ maxima and coincide with the location of the small vortex centers. The maximum in $\mathcal{L}$ moves outward following the maximum in Σ as the gap opens. This illustrates that the vortices form due to the RWI at the slope of the outer gap edge.

In the absence of self-gravity, these small vortices then quickly merge together (within ∼100 planet orbits) into a single large vortex that slowly moves outwards following the gap edge as the gap deepens and widens (third and fourth row in Fig. 1). The surviving vortex then typically decays over ∼200–2000 orbits. The evolution of three vortex properties is illustrated

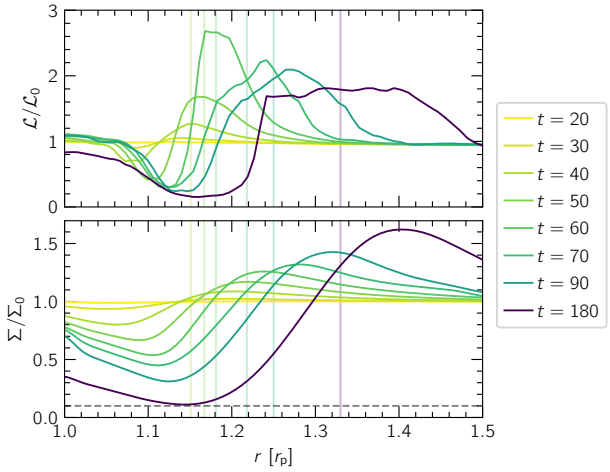

Fig. 2. Evolution of radial Lovelace parameter (see Eq. (8)) and Σ profiles during vortex formation over the first 200 orbits of the sample case from Sect. 3. The vertical lines indicate the center of the maximum in $\mathcal{L}$ (estimated by eye) to guide the eye to the corresponding location of the Σ profile. $\mathcal{L}$ is calculated as the azimuthal average at each radius. The dotted horizontal line in the bottom panel marks 10% of Σ_0 which we define as the location of the gap edge.

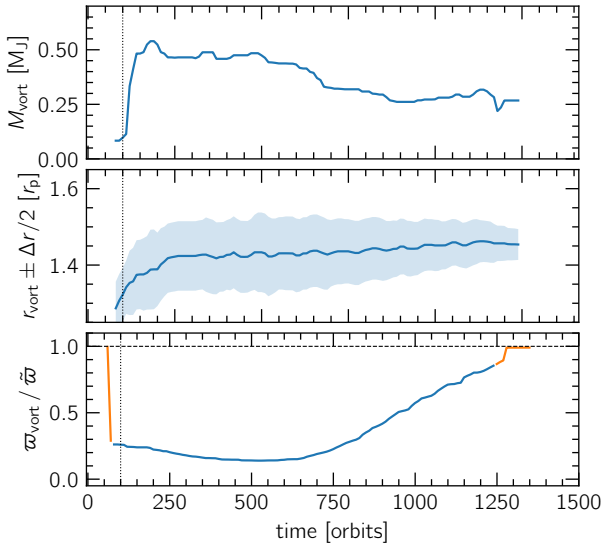

Fig. 3. Evolution of vortex properties for the showcase simulations presented in Sect. 3. The panels show, from top to bottom, the mass enclosed in the FWHM ellipse of the vortex fit, M_{vort} in Jupiter masses, the radial location of the vortex, r_{vort} and its FWHM, Δr, indicated by the shaded area, and the ratio between minimum vortensity inside the vortex and the azimuthal median of vortensity at the radial location of the vortensity minimum. A dotted vertical line indicates the time when the planet has reached its final mass. The curves are smoothed with a median filter which spans over the next and last 5 datapoints (±50 orbits at r_p). The orange parts of the line in the bottom panel show the evolution of the vortensity prior to the "birth" and after the "death" of the vortex.

in Fig. 3. It shows, from top to bottom, the mass M_{vort} enclosed within the FWHM ellipse of the 2D Gaussian fit to Σ, the location of the center of the vortex r_{vort} and the radial FWHM width Δr as the shaded area, and the vortensity at the vortex center normalized by the azimuthal median. The vertical dotted lines indicate the time when the planet reached its full mass (typically 100 orbits). A short phase of vortex formation is followed by a slow and steady decay process, as can be seen in the decrease of mass and radial size. Because the vortensity contribution of the anticyclonic vortex is negative, an increase in vortensity indicates a decay as well. The line in the bottom panel of Fig. 3 is continued (in orange) for another 100 orbits after the vortex decayed according to our criterion presented below in Sect. 4.1, to illustrate the return of the curve to 1, which corresponds to an azimuthally symmetric state.

During its lifetime, the vortex can become as large as $\Delta r = 0.4\, r_p$ (2 au for $r_p = 5.2$ au) with a typical vortex aspect ratio ($r\Delta\phi/\Delta r$) of 6-10. Its mass, M_{vort}, is typically some tenths of M_J but can be as large as one M_J, with a surface density enhanced by a factor of up to 7 compared to the initial value.

The vortices form around the location where the radial Σ profile reaches 10% of its initial value (see bottom panel of Fig. 2), which we define as the gap edge similar to Crida et al. (2006). During their lifetime, most vortices tend to stick to this gap edge in the sense that their inner boundary, $r_{vort} - \Delta r/2$, roughly coincides with the gap edge location. For some models, we observe that the vortex detaches from the outer gap edge after several hundred orbits and starts migrating outward. These models will be discussed later in Sect. 5.

4. Dependence of vortex properties on physical parameters

Having described a typical lifetrack of a vortex in our simulations, we now present the effects of different physics and numerics on vortex lifetime, location, and impact on the overall disk structure. The model parameters are listed with the main results in Table C.1.

4.1. Vortex lifetime

We define the vortex lifetime as the time difference between its "birth" and "death" by analyzing the ratio of ϖ to the azimuthal median value, $\tilde{\varpi}$ as a function of time. The normalization with $\tilde{\varpi}$ instead of ϖ_0 is done to eliminate the ϖ evolution of the background disk due to changes in Σ and radial pressure gradients, which affect $(\nabla \times \boldsymbol{u}) \cdot \hat{z}$ by changing the azimuthal velocity.

The "birth" is identified as the time when $\varpi/\tilde{\varpi}$ drops from its initial value of 1 (for an axisymmetric disk) down to lower values (see bottom panel of Fig. 3). Because $\varpi/\tilde{\varpi}$ drops already for small vortices, the lifetime also includes the stage where there are multiple small vortices (see Sect. 3).

The "death" of the vortex, however, is less obvious to identify. At this stage, $\varpi/\tilde{\varpi}$ usually slowly rises back to the background disk's value. Usually, there is a "knee" visible in $\varpi/\tilde{\varpi}$ at or slightly after the point in time where the vortex dies and where ϖ approaches the background flow (see the orange part of the line in the bottom panel of Fig. 3 where $\varpi/\tilde{\varpi}$ is continued for another 100 orbits after the vortex disappeared at $t = 1250$ orbits). For some models, this "knee" is not visible, and we manually inspect the 2D contour plots of $\varpi/\tilde{\varpi}$ and identify when no closed iso-value lines (with spacing in $\varpi/\tilde{\varpi}$ of 0.05) are present anymore. As an additional measure for less obvious cases, we analyze the gas streamlines at different timestamps.

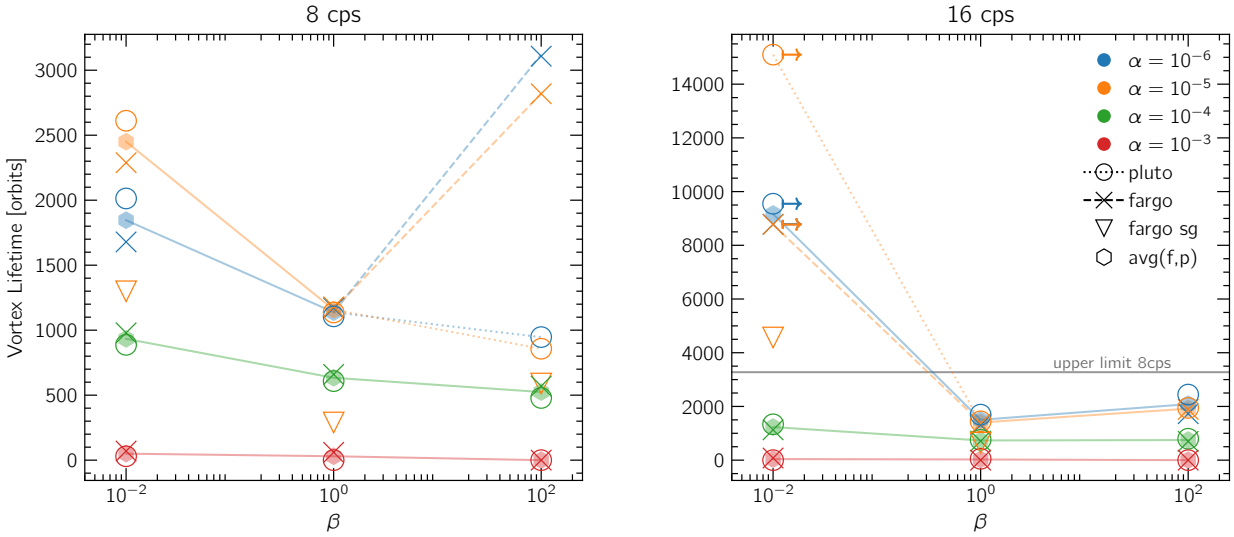

Fig. 4. Lifetime of vortices as a function of β for 8 cps (left) and 16 cps resolution (right). Colors encode α, and the different symbols denote the code and the inclusion of self-gravity. The solid lines help guide the eye and connect the lifetime averages between the two codes (without self-gravity) for each value of α, where the two codes agree sufficiently. For parameters where there is a difference between the codes, dashed and dotted lines connect to the datapoints of the `FargoCPT` and `PLUTO` runs, respectively. A "↦" next to a symbol marks models that were terminated due to runtime constraints but still contain an active vortex. The horizontal gray line in the right panel indicates the top of the y-axis of the left panel. A list of all vortex lifetimes shown here is provided in Table C.1.

In our models, the drop in $\varpi/\tilde{\varpi}$ happens in a matter of tens of orbits. A conservative estimate for the uncertainty of this "birth" time measurement is 50 planetary orbits. From applying the manual method to models where the "knee" exists in the $\varpi/\tilde{\varpi}$ curve (implying the "death" of the vortex), we estimate a conservative uncertainty to be 100 planetary orbits. This leaves a total uncertainty of 150 planetary orbits for the lifetime of our vortices.

The lifetime of vortices in our grid of simulations is shown as an overview in Fig. 4. The left and right panels show vortex lifetimes as a function of β for 8 and 16 cps, respectively. The viscous α is encoded in color, and the symbol indicates the simulation code and the inclusion of self-gravity. For each value of α and β, we calculated the average ("avg(f,p)") between the two codes (not including the self-gravity models) when the results are close together. The solid-colored lines connect the averages to help visualize the trends. For parameters for which the two codes showed different vortex lifetimes, we added separate lines connecting the average to the `FargoCPT` and `PLUTO` results to highlight the differences.

Lifetimes range from some hundred to 2000 orbits for the shorter-lived vortex group up to at least 15 000 orbits for the long-lived vortices discussed later in Sect. 5. The most prominent features of the distribution are the trend of decreasing lifetime with increasing α and the minimum of vortex lifetime at $\beta = 1$ for low α and high resolution.

In the following sections, we address the influence of our model parameters on vortex lifetime.

4.2. Influence of the thermal relaxation timescale

The dimensionless thermal relaxation timescale β has a strong effect on vortex lifetime. For $\alpha = 10^{-4}$, lifetimes are of the order of several hundred to 1000 orbits with a downward trend as β increases. Vortex lifetimes are shortest for $\beta = 1$ (around 1250 orbits) and increase towards both sides to around 2000 orbits for $\beta = 100$ and to values of the order of 10 000 orbits for $\beta = 0.01$, high-resolution runs. Exceptions to this trend are the 8-cps `PLUTO` models for $\alpha = 10^{-5}$–10^{-6} and $\beta = 100$. We could not identify the reason why the two codes did not agree for these parameters, but we note that the two codes match well once again for 16 cps in the same configurations. Models with very long vortex lifetimes will be analyzed later in Sect. 5.

Fung & Ono (2021) reported a similar trend in vortex lifetime in two-dimensional shearing-box simulations without planets, in which the vortex was introduced by initializing the simulation with a radial density bump. They found that vortex decay is fastest for intermediate β in the range 1–10, but their disk model assumes a constant background disk, without gradients in T and Σ which change baroclinic effects. Our results indicate that a similar mechanism might be at play in the presence of an embedded planet with strong spiral arm shocks. However, the strong enhancement of vortex lifetime for $\beta = 0.01$ hints at the presence of an additional mechanism which keeps the vortices alive. We discuss these hypotheses further in Sect. 6.1.

For a comparison of vortex evolution at different β see Fig. 5, where the evolution of vortex properties (analogous to Fig. 3) of three `FargoCPT` simulations at 8 cps resolution with $\beta = 0.01, 1, 100$ is shown. The absolute radial location of vortices varies with β, as well. This is due to the tendency of the vortices to form and subsequently stick to the outer planet gap edge and the gap opening process being strongly influenced by β. Miranda & Rafikov (2020b) showed that "extreme" values of β (i.e., $\beta \to 0$ or $\beta \to \infty$) result in narrower planet-opened gaps but additional gaps in the inner disk, whereas intermediate values of $\beta \sim 0.1$–10 lead to a single, wide gap around the planet's orbit. In our simulations, models with $\beta = 0.01$ show the widest gaps, narrower gaps are present for $\beta = 1$, and $\beta = 100$ models showed an even slightly narrower gap. This is reflected in the vortex locations which are further in for higher β (see the center panel of

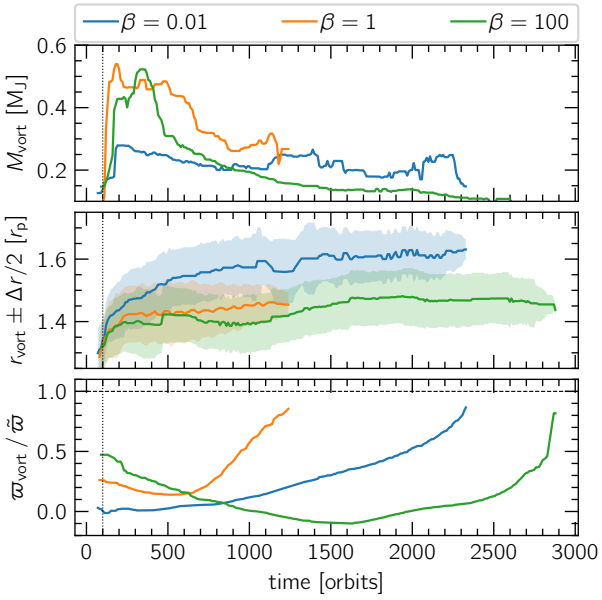

Fig. 5. Evolution of vortex properties for varying values of the thermal relaxation β parameter. The panels are as in Fig. 3. Shown are models run with the FargoCPT code with $\alpha = 10^{-5}$ and at 8 cps resolution (orange "f8" dots in Fig. 4).

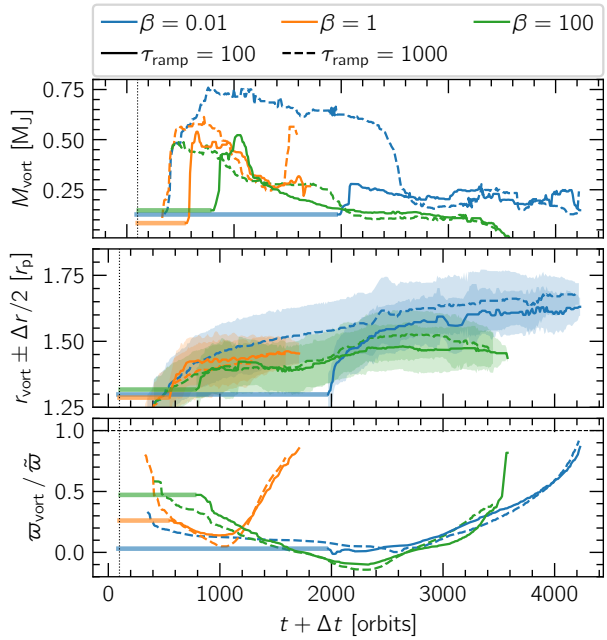

Fig. 6. Influence of the planet introduction time on the evolution of vortex properties. The panels are as in Fig. 3. Solid and dashed lines show models with a $\tau_{ramp} = 100$ orbits and 1000 orbits, respectively. The $\tau_{ramp} = 100$ orbits curves are shifted to the right (see the horizontal lines) to illustrate that the curves have the same shape in the decay phase, independent of τ_{ramp}. Note that the final evolution of the vortex, after it has reached its minimum in vortensity, is the same independent of planet introduction time.

Fig. 5). The difference to Miranda & Rafikov (2020b) might be due to the presence of the vortex.

4.3. Planet growth timescale

Hammer et al. (2017) observed that the lifetime of planet-induced vortices can depend on the timescale over which the planet mass is increased in order to introduce the planet into the simulation. They found that vortex lifetime decreased with a longer planet growth time. In our models, increasing the planet growth timescale from 100 to 1000 orbits caused vortices to live longer by 470 for $\beta = 1$ up to 1900 orbits for $\beta = 0.01$. Fig. 6 shows the evolution of vortex quantities comparing the FargoCPT runs with a $\tau_{ramp} = 100$ orbits, already presented in Fig. 5, with their respective counterparts with $\tau_{ramp} = 1000$ orbits. The curves of runs with $\tau_{ramp} = 100$ orbits are shifted to the right by a time Δt from 470 to 1900 orbits. This shift clearly illustrates that the decay of these vortices is almost the same for both values of τ_{ramp} in terms of their mass, location and vortensity curves. The only difference caused by the planet injection timescale is how long it takes for the vortex to reach the turnover point, after which it starts to decay.

4.4. Planet mass

From our $M_p = 0.5\,M_J$ models we can not draw any conclusions regarding the dependence of vortex lifetime on planet mass, because for the set of parameters, $\beta = 0.01$ and $\alpha = 10^{-5}$, the vortices are long-lived outliers like the ones discussed in Sect. 5. However, the location of the vortex is also influenced by the planet's mass. Lower-mass planets open narrower gaps and cause the location of the vortex, given that this is linked to the gap edge's location, to be further in compared to more massive planets. In our models, the vortices in the $M_p = 0.5\,M_J$ were located $\sim 0.15\,r_p$ closer to the star.

4.5. Viscosity

The observed vortex lifetime typically increases with lower values of α. Simulations with $\alpha = 10^{-3}$ show only small vortices forming. They disappear within 100 orbits, thus, they are already gone by the time the planet has grown to its full mass. For models with $\alpha = 10^{-4}$, we observe vortex lifetimes of up to around 1000 orbits.

Simulations with a lower viscosity ($\alpha = 10^{-6}$–10^{-5}) show even longer lifetimes, usually in the range between 1000 and 2000 orbits, excluding the outliers that we discuss later in Sect. 5. For this range of α, vortices usually have similar lifetimes for simulations sharing the same β value. For an example, see Fig. 7 which shows, from top to bottom, the evolution of the mass enclosed in the vortex's region (FWHM), the location and radial extent (in FWHM) of the vortex as determined by the surface density fit, and the ratio of normalized vortensity to the azimuthal median of the latter at the location of the vortex.

The vortex location is not influenced by viscosity. Although the gap opening time is $t_{gap} \approx 2700 \left(\frac{\alpha}{10^{-3}}\right)^{-3/2} T_{orb}$ according to the estimate in Kanagawa et al. (2017), the bulk of the gas in the planet's vicinity is cleared within the first few hundred orbits. During this time, Σ is lowered by two orders of magnitude within the gap region, and the radial gradient of Σ becomes steep enough to facilitate vortex formation.

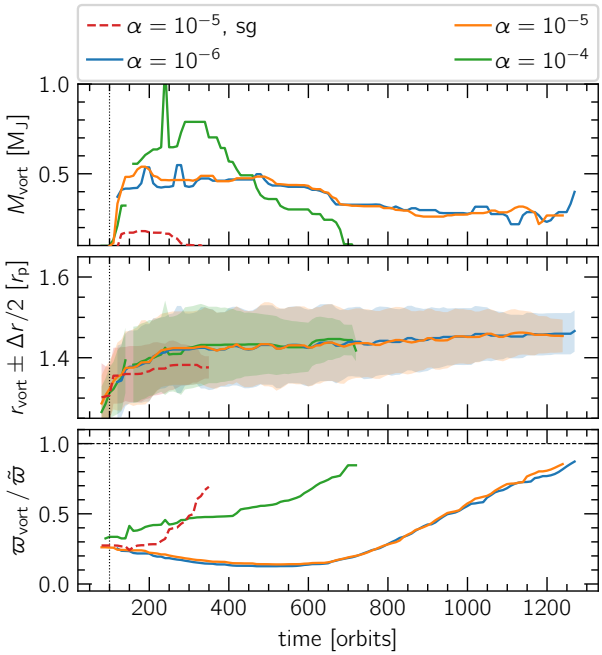

Fig. 7. Evolution of vortex properties for varying values of α. Panels are shown as in Fig. 3. Shown are models run with the `FargoCPT` code with $\beta = 1$ and at 8 cps resolution. The $\alpha = 10^{-3}$ run is excluded because no vortex forms. In addition, a run with disk self-gravity enabled is added for the $\alpha = 10^{-5}$ case. The similarity between simulations with $\alpha = 10^{-5}$ and 10^{-6} is apparent.

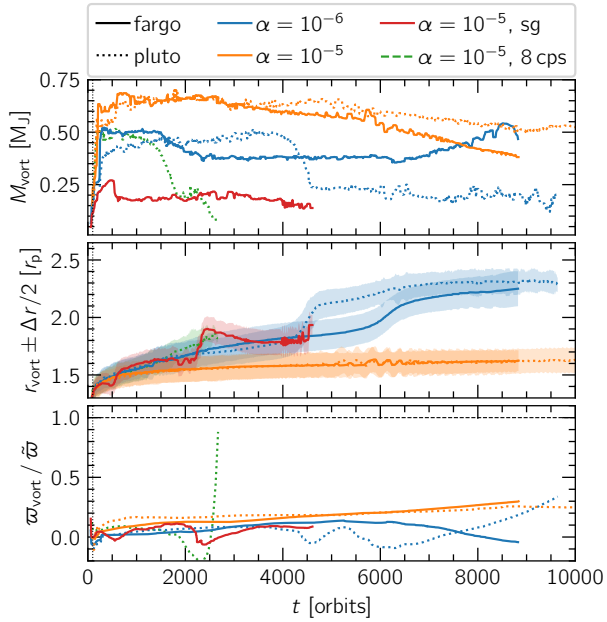

Fig. 8. Selection of models with long-living and migrating vortices at 16 cps resolution for the two different codes. Both codes agree remarkably well for the blue and orange cases. The panels are as in Fig. 3. In models shown here, $\beta = 0.01$. The values of α, resolution and code used are indicated in the legend.

4.6. Self-gravity

Several studies showed that vortices in weakly or strongly self-gravitating disks might not grow as large because small vortices do not merge into one large vortex (Lin & Papaloizou 2011) and dissipate more rapidly due to stretching in the azimuthal direction (Lovelace & Hohlfeld 2013; Regály & Vorobyov 2017; Zhu & Baruteau 2016). This can be the case even for low-mass disks as long as the Toomre stability parameter Q is lower than 50 or $hQ \leqslant \frac{\pi}{2}$. For the choice of parameters in our models, the Toomre parameter is $Q \approx 25\,(r/r_\mathrm{p})^{-3/2}$ ($hQ \approx 1.25\,(r/r_\mathrm{p})^{-3/2}$), dropping under 5 at roughly $r/r_\mathrm{p} = 2.8$. To check the effect that disk self-gravity has played in our models, we ran additional simulations with `FargoCPT` with self-gravity activated for all three values of $\beta = 0.01, 1, 100$ and for $\alpha = 10^{-5}$.

The lifetimes of vortices in these simulations are shown in Fig. 4 as the rightmost datapoint in each column (models "f8sg"). An example evolution of their properties is shown in Fig. 7.

Self-gravity inhibits the merging of the small initially formed vortices in to one large vortex. Instead, two smaller vortices usually remain until they decay. This leads to a significantly shorter lifetime compared to the analogous simulations without self-gravity, consistent with the above studies. This, however, does not apply to the long-lived, migrating vortices which are discussed in Sect. 5.

Fig. 7 shows that the center of the vortex in a model with self-gravity and $\beta = 1$ is further in compared to its non-self-gravitating counterpart. This is due to a smaller radial extent of the vortex in the run with self-gravity and the tendency of the inner edge of each vortex to coincide with the gap edge.

Because self-gravity does not noticeably change the radial disk profile for the mass regime of our models, the inner edge of the vortices is at the same location, independent of whether self-gravity is included or not. The same effect is also observed for $\beta = 0.01$ and $\beta = 100$.

5. Long-lived and migrating vortices

In some of the cases, a much longer-lived vortex is observed. In these models, vortices stay close to their peak mass for several thousand orbits and, in some cases, migrate outwards after having stayed at the planet gap edge. This happens only for very low viscosities ($\alpha \leq 10^{-5}$) and $\beta = 0.01$ or locally isothermal simulations ($\beta \to 0$). For our standard $M_\mathrm{p} = 1\,\mathrm{M_J}$ planets, the long-lived outliers appear only at the highest resolution of 16 cps but not at 8 cps. For the corresponding $M_\mathrm{p} = 0.5\,\mathrm{M_J}$ model, the long-lived vortex also appeared at 8 cps. Spiral arms launched by the vortex are clearly visible for these long-lived large vortices (see Fig. B.2). They are more pronounced for lower values of α.

Figure 8 shows the evolution of vortex properties for a selection of models to highlight the observed behavior. The most prominent example is the model with $\alpha = 10^{-5}$, $\beta = 0.01$ and a 16 cps resolution. The vortex in those runs lived for 15 100 orbits before we terminated the two simulations due to their long runtime. Both codes, `PLUTO` and `FargoCPT`, agree well for the long-lived cases. Specifically, they are in exceptionally close agreement for $\alpha = 10^{-5}$ and only differ at later stages for $\alpha = 10^{-6}$ (see orange and blue lines in Fig. 5).

We do not currently fully understand the mechanism that allows these long-lived vortices to sustain themselves for such

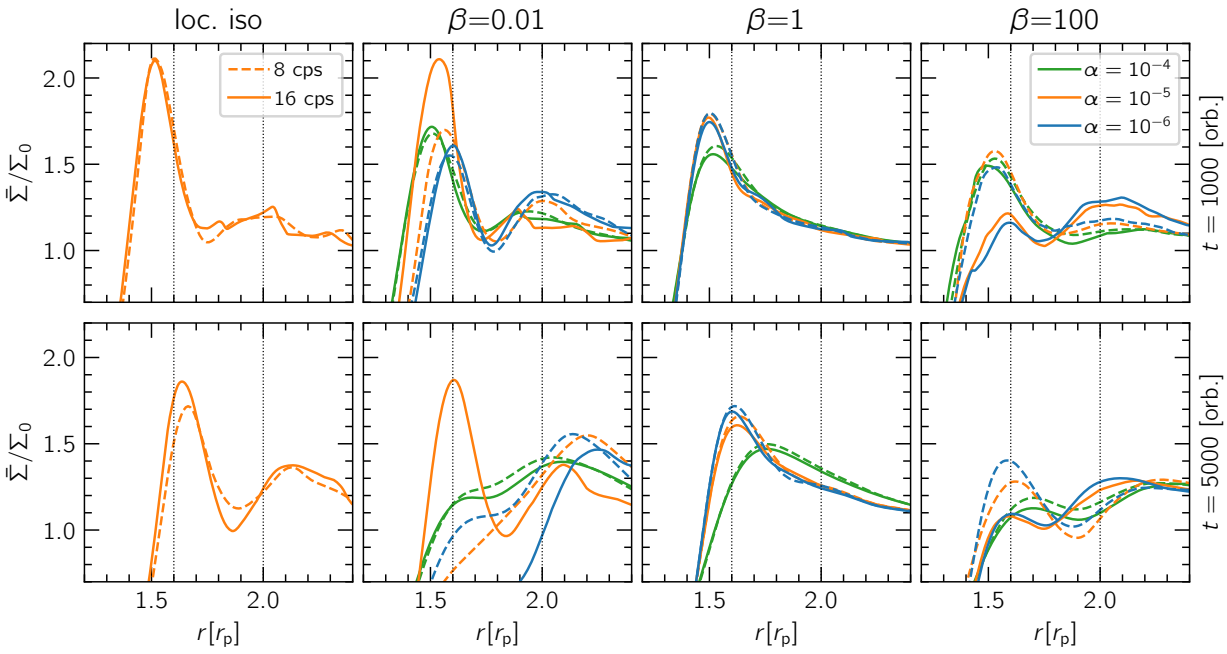

Fig. 9. Azimuthally averaged surface density profiles as a function of different physical (α, β) and numerical (cps) parameters at two different timestamps. The peak around r/r_p = 1.5–1.6 corresponds to the pressure bump formed by the planet as the latter pushes material away, forming a gap around its orbit. The smaller, secondary peak at around r/r_p = 2.1 is caused by the vortex that forms near the "primary", planet-generated bump. Top: radial profiles at t = 1000 orbits. At this stage, all models pictured feature a vortex near the primary bump. We note the absence of a secondary bump for the models with β = 1. Bottom: the same profiles at t = 5000 orbits. Here, the primary bump has moved radially outwards as the planet's gap gets deeper and wider. We highlight the depletion of gas near the "primary" pressure bump for the second panel from the left (β = 0.01). This is caused by the combination of a vortex migrating outwards to the secondary bump, and the inability of the planet to resupply that zone with material from its now-depleted gap region. Also note the difference between resolutions of 8 and 16 cps (dashed and solid lines), especially for the β = 100 models and the β = 0.01, $\alpha = 10^{-5}$ model (the evolution of this model is shown as the orange line in Fig. 8).

long timescales. We attempt to provide a speculative explanation in Sect. 6.1.

5.1. Migration and secondary vortices

For $\beta \neq 1$, a secondary radial density and pressure bump is observed in the outer disk. This is the result of the vortex generating spiral arms which transport angular momentum. Radially outwards, this results in the accumulation of mass in a second bump (see panels for $\beta \neq 1$ in Fig. 9). This does not happen for $\beta = 1$ due to the less efficient angular momentum transport by spiral arms for this intermediate value of β (Miranda & Rafikov 2020b).

For $\beta = 0.01$, some models show vortices migrating radially outwards (e.g., the $\alpha = 10^{-6}$ models in Fig. 8). This is likely related to the formation of the secondary bump outside of the vortex location (see Fig. 9) and the fact that vortices typically migrate towards pressure bumps (Paardekooper et al. 2010).

For vortices that migrate far enough outside, which only happens for $\beta = 0.01$, a weaker secondary vortex appears between them and the planet's gap edge (see Fig. B.2). These secondary vortices then decay over a few hundreds of orbits, already having decayed by the time the "primary" vortex disappears. While they are treated as independent entities, they are not included in Fig. 4 or the discussion above.

Their occurrence is likely the result of a multistage process which begins with the secondary bump forming and the "primary" vortex migrating radially outwards towards it and meanwhile supplying mass towards the planet-generated gap edge. This then feeds the emerging "secondary" vortex.

6. Discussion

In this section, we address some ways in which our results could be interpreted and their relevance in explaining observations. We also underline some caveats of our models.

6.1. On the conditions to form and sustain a vortex

To form a vortex, one needs to create a local vortensity extremum. In the absence of non-conservative forces, the evolution equation for the vortensity in a two-dimensional flow reads

$$\frac{\partial \varpi}{\partial t} + \boldsymbol{u} \cdot \nabla \varpi = \frac{\nabla \Sigma \times \nabla P}{\Sigma^3} \cdot \hat{z} + \mathcal{V} = \mathcal{B} + \mathcal{V} \tag{9}$$

where $\mathcal{B} = \frac{\nabla\Sigma \times \nabla P}{\Sigma^3} \cdot \hat{z}$ is the baroclinic term and $\mathcal{V}$ describes viscous diffusion of vortensity which can lead to vortex decay.

As outlined in the introduction, several instabilities have been discovered that provide a mechanism to form or destroy large-scale vortices, but they all fundamentally rely on Eqn. 9 to change the vortensity of the flow. The mechanism responsible for the formation of the vortex in our simulations is most likely the RWI which is triggered during the gap opening process, as we demonstrated in Sect. 3 and Fig. 2.

To check whether vortices can only form during the gap opening process and not in the quasi-steady state after the bulk of the gas has been pushed out of the gap region, we removed the long-lived vortex from the $\alpha = 10^{-5}, \beta = 0.01$, 16 cps model by replacing the velocities and Σ with their azimuthal median values for $r > r_p$ during the peak of its activity ($t = 1880$ orbits). The fact that there is no vortex forming again is an indication that the formation of vortices in our simulations depends on the gap opening process to produce conditions that can trigger the RWI. This is also backed by the observation that the peak in $\mathcal{L}$ is strongest for an intermediate time, $t = 70$ orbits, during the gap opening process, after which the maximum disappears and a plateau in $\mathcal{L}$ forms.

Vortex decay happens due to at least two mechanisms. Viscous spreading attacks the vortices for high $\alpha = 10^{-4}$–10^{-3}, as illustrated by the trend of lower vortex lifetime for higher α, and vortex stretching due to self-gravity effects additionally limits vortex lifetime if it is considered (Lin & Papaloizou 2011; Zhu & Baruteau 2016; Regály & Vorobyov 2017).

For sufficiently low α, another process that depends on β starts to be dominant. We do not fully understand the mechanism but we observed some similarities to the recent work by Fung & Ono (2021). They found that, in their simulations, vortices decay the fastest for $\beta = 1$–10 and decay slower for both smaller and larger β. Vortex lifetime in their simulations changed by up to an order of magnitude depending on β. We also find a minimum in vortex lifetime for $\beta = 1$ with lifetimes increasing as $\beta \neq 1$.

Fung & Ono (2021) explained the decay mechanism by asymmetries in the structure of $\mathcal{B}$ around the vortex center, which they found to be quadrupolar (see their Fig. 6) and to change with β. We also find asymmetries in the structure of β, however our simulations differ from theirs in some fundamental aspects. Our simulations are global with radially varying Σ and T profiles and include a planet that continually perturbs the disk, whereas their simulations consider a local shearing sheet with a constant background Σ and T, with only an initial perturbation in the form of a density bump. As a consequence of the radially varying T in our simulations, the structure of $\mathcal{B}$ around the vortex center is dipolar in the azimuthal direction, as can be expected for a Gaussian-like density maximum. Additionally, the planetary spiral arms strongly influence $\mathcal{B}$. Figure 10 shows a 2D map of $\mathcal{B}$ for two simulations with $\alpha = 10^{-5}$ and 16 cps resolution. The left panels show a short-lived vortex with $\beta = 1$ at $t = 1000$ orbits and the right panels show the long-lived vortex model which exhibits the "secondary" vortex (see Sect. 5.1 for a description and Fig. B.2 for ϖ and Σ maps at the same time). The actual shape of the perturbation of $\mathcal{B}$ inside and around the vortex is time-dependent and depends on the phase w.r.t. to the spiral arm. It is not clear to us, at the moment, how the changes in structure of $\mathcal{B}$ lead to the change in vortex decay and how this proposed mechanism depends on the various parameters in our system.

The long-lived group of vortices for low β (see Sect. 5) indicates that there might another vortex formation mechanism at play. Given that the RWI already caused finite perturbations in the disk and our disks exhibit a radial entropy gradient, the SBI (Klahr & Bodenheimer 2003; Lesur & Papaloizou 2010) seems to be a natural candidate. However, we verified that the SBI is not active in our disks by analyzing the Richardson number, the ratio of the buoyancy (also called Brunt-Väisälä) frequency to the shear rate, which needs to be negative in a radially extended region over the full azimuth of the disk for the SBI to operate. The Richardson number in our simulation is positive, except for narrow stripes following the spiral arms, which rules out that the SBI is active.

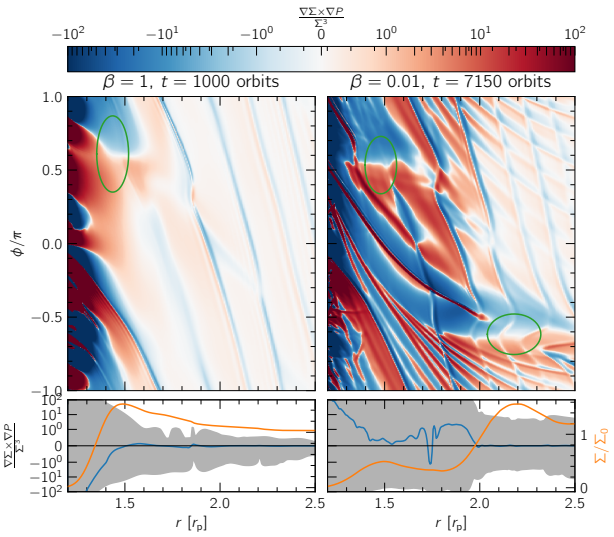

Fig. 10. Baroclinic term (RHS of Eqn. (10)) in the outer disk for a short-lived model ($\beta = 1$ at $t = 1000$ orbits) and a long-lived vortex ($\beta = 0.01$ at $t = 7150$ orbits, see also Fig. B.2) with $\alpha = 10^{-5}$ and 16 cps resolution on the left and right side, respectively. The top row shows maps of the baroclinic term with the detected vortices indicated with green ellipses as obtained from the Σ fit. The bottom row shows the radial Σ profile in orange, the azimuthally averaged baroclinic term in blue and the region between its minimum and maximum shaded in gray.

To rule out that the difference in lifetime is a result of the initial vortex formation during gap opening, we took the long-lived vortex out of the $\alpha = 10^{-5}, \beta = 0.01$, 16 cps model and inserted it into the $\alpha = 10^{-5}, \beta = 1$, 16 cps model. Although this artificial vortex has the same structure as in its original $\beta = 0.01$ model, it decays over nearly the same time as the standard $\beta = 1$ vortex. This is an additional indication that the difference in lifetime is cause by the dependence of the decay process on β or a possible additional vortex formation channel that sustains the vortex at low β.

This leaves us with the hypothesis that the interaction of the spiral arms with the vortices might play a major role in either slowing down vortex decay or providing an additional vortex formation channel. The hypothesis is motivated by the strong impact of the spiral arms on $\mathcal{B}$ and the dependence of spiral arm properties on β (Ziampras et al. 2020b; Miranda & Rafikov 2020b). Another contribution might be the vortensity jump across the spiral arm shock, which was recently illustrated to be important for the evolution of vortensity in the case of sub-thermal-mass planets (Cimerman & Rafikov 2021). Providing an analysis of both mechanisms in our context is, unfortunately, out of the scope of the present explorative study.

6.2. Effect of in-plane radiation transport

It has been shown that parametrizing radiative effects with β while omitting the effects of in-plane radiation transport can result in a potentially inaccurate radial surface density structure mainly in the inner disk and around the gap, due to how β affects the capability of a planet to open multiple "secondary" gaps at

$r < r_p$ (Miranda & Rafikov 2020b). Here, we are not interested in the annular structures of the inner disk, so we chose to ignore in-plane radiation transport. Nevertheless, to check for possible effects of in-plane radiation transport on the vortex dynamics, we repeated the $\alpha = 10^{-5}$, $\beta = 1$ model at 8 cps. This time we included a flux-limited diffusion (FLD) approach (Levermore & Pomraning 1981) similar to Ziampras et al. (2020a), but by parametrizing the diffusion coefficient $D_{\rm rad}$ following Eqs. (12)–(14) of Flock et al. (2017):

$$t_{\rm cool} \approx \frac{l_{\rm thin}^2}{3D_{\rm rad}} + \frac{H^2}{D_{\rm rad}} = \beta\Omega_{\rm K}^{-1} \Rightarrow D_{\rm rad} = \frac{\Omega_{\rm K}}{\beta}\left(\frac{l_{\rm thin}^2}{3} + H^2\right), \quad (10)$$

where $l_{\rm thin}$ is the photon mean free path. We found that including FLD slightly changes the radial surface density structure in the inner disk as predicted by Miranda & Rafikov (2020b) and reduces the vortex lifetime from 1200 to 900 orbits. Studying the effect of in-plane radiation transport in more detail requires further investigation.

6.3. About the assumption of a 2D disk

One of the main limitations of our models is the 2D assumption which was done due to runtime constraints in our rather wide exploration of the parameter space. It is entirely possible that various 3D effects can result in quantitative differences in vortex properties. 3D vortices can be susceptible to the elliptical instability (Lesur & Papaloizou 2009) which would lower their lifetime. On the other hand, the vertical modes of the SBI could provide an additional channel to sustain the vortices, and vertical gas circulation due to the VSI might interfere with vortex growth and decay (Flock et al. 2020).

To estimate the impact of including full-3D effects, we ran one 3D simulation with `FARGO3D` (Benítez-Llambay & Masset 2016) using a setup analogous to our 2D setup. We chose $\beta = 2\pi$, $\alpha = 0$ and a resolution of 8 cps in all three directions. Similar to our 2D models, a large vortex formed at the outer gap edge and lived for 7000 orbits. This illustrates that while there are differences, large vortices can survive in 3D disk simulations for a long time, even longer than in 2D for our example. We limited the 3D runs to this one test because its runtime at 8 cps resolution was close to 4 months with the simulation performed on 4 NVIDIA K80 GPUs.

6.4. On the observability of vortices at large radii

Sect. 4.2 illustrates that vortex lifetime is affected by the thermal cooling timescale β. The latter is expected to vary with radius in a disk, with values of 1–10 at 5 au, 0.1 at ∼10 au, and below 0.1–0.01 at ∼50 au (Ziampras et al. 2020b). Thus, we expect vortices to be in the short-lived regime close to the star and in the long-lived regime far from the star. From Fig. 4 we can estimate the lifetime of vortices in disks with $\alpha \leq 10^{-4}$ to be between 500 and 3000 orbits for $\beta \geqslant 1$ and between 1000 and 15 000 orbits for $\beta < 1$ for $\alpha \leq 10^{-4}$. Assuming a solar-mass star, this yields estimated lifetimes for a planet-induced vortex between 6–30 kyr at 5 au, 175–700 kyr at 50 au, and 1–15 Myr at 100 au. On the basis of a simple lifetime-centered argument, our results, therefore, suggest that planet-induced vortices are more likely to be observed at larger radii.

It should be noted, that planet growth timescales of 100 and 1000 planetary orbits are at the very low end of the spectrum of physically expected planet growth times. Hammer et al. (2017) provided estimates for more realistic planet growth-times of several thousand up to tens of thousands of orbits. It remains to be seen, whether the effects observed in this study still appear for longer, more realistic, planet growth timescales. However, simulating the disks at the required resolution of at least 16 cps for longer planet growth times along with the additional vortex evolution time is still computationally expensive.

6.5. On using the lifetime of vortices in simulations to explain observations

In the suite of simulations we carried out, the lifetime of vortices in models with identical physical parameters varies significantly with resolution. This was the case for low values of the viscous α parameter ($\alpha = 10^{-5}, 10^{-6}$). We argued that the numerical viscosity of our simulation codes is comparable to $\alpha_{\rm num} \lesssim 10^{-5}$. This suggests that simulations with a prescribed viscosity of the order of the numerical viscosity cannot be used as a controlled numerical experiment, at least as far as the occurrence and persistence of vortices is concerned. For prescribed viscosities well above the estimated numerical diffusion ($\alpha = 10^{-4}, 10^{-3}$ in our case), the consistency of vortex lifetimes between the two codes and numerical choices supports the idea that the numerical experiment is indeed a controlled one.

In light of some recent observations of molecular line broadening (e.g., Flaherty et al. 2018) and numerical studies of VSI turbulence (e.g., Flock et al. 2017) and planet–disk interaction (e.g., Zhang et al. 2018) hinting at low α values, this could pose a challenge for simulations of protoplanetary disks.

6.6. Resolution and numerical viscosity

Vortex evolution in "inviscid" disks is often studied using very high-resolution grids to minimize the effects of numerical viscosity (Li et al. 2005; Paardekooper et al. 2010; Lin & Papaloizou 2011; Zhu & Baruteau 2016; Hammer et al. 2017; McNally et al. 2019; Hammer et al. 2021; Fung & Ono 2021). While the resolution of 8 and 16 cells per scale height is likely enough to resolve planet-generated features such as the gap shape and spiral arms (see Appendix A), the numerical viscosity also needs to be low enough not to interfere with vortex decay.

An estimation of the numerical viscosity, valid for first-order schemes, is $\nu_{\rm num} \sim \frac{\Delta x^2}{\Delta t}$, with a representative cell size Δx and the timestep Δt. For our choices of parameters and assuming $\Delta t \approx \frac{\Delta x}{c_s}$ this corresponds to $\alpha_{\rm num} \sim 10^{-2}$–$10^{-1}$. Clearly, we see substantial changes in dynamics down to much lower values of the prescribed α. Because we employ a higher-order scheme, this simple estimate is not applicable. To our knowledge, there exists no formula to estimate the numerical viscosity for the higher-order schemes employed in this study, so we attempt to estimate it by comparing the results of our simulations at different values of α.

In general, we observe a similar behavior between models with $\alpha = 10^{-5}$ and 10^{-6}, both in terms of the behavior of vortices during their lifetime (size, mass, migration patterns) as well as the overall lifetime itself (see Fig. 4). This is also true across both codes that we used in this study, with the exception of the 8 cps models for $\beta = 100$. We attribute the similarity to the numerical diffusion inherent in the different advection schemes of the two codes and expect that this translates to an effective $\alpha_{\rm num}$ between 10^{-6} and 10^{-5} for our given choices of grid resolution. This implies that our experiments with $\alpha = 10^{-6}$ are most likely

not controlled ones, and for this reason, we typically group models with $\alpha \leq 10^{-5}$ together.

Nevertheless, we still observe a different behavior for some models with $\alpha = 10^{-6}$ when comparing them to those with $\alpha = 10^{-5}$, such as the migration of the long-lived models presented in Sect. 5 (see the different tracks of $r_{\rm vort}(t)$ in Fig. 8), most of which have a 16 cps resolution. This hints at a lower numerical diffusion for 16 cps of $\alpha_{\rm num} \sim 10^{-6}$.

Because the numerical viscosity in the 8 cps models might interferes with the prescribed $\alpha \leq 10^{-5}$, our 8 cps simulations might not be as trustworthy as our higher-resolution 16 cps, $\alpha \leq 10^{-5}$ runs.

6.7. On the different numerics of the two codes

We used two codes (PLUTO and FargoCPT) with fundamentally different numerical properties. The fact that the two codes agree in terms of results (see the orange lines in Fig. 8 for one striking example) is reassuring, but it is worth discussing their differences nonetheless.

FargoCPT requires an artificial viscosity prescription to stabilize the upwind method near regions of strong compression such as shocks. This provides additional dissipation which could affect the evolution of vortices whenever they interact with the spiral shocks induced by the planet. With the exception of the 8 cps models for $\beta = 100$, we found no significant differences in vortex lifetimes between the two codes. The one case for which the codes disagreed might be a result of insufficient resolution because the differences disappear for 16 cps.

On the other hand, PLUTO's strictly energy-conserving nature means that the evolved quantity in the energy equation is the sum of kinetic and thermal energy. Since kinetic energy dominates over thermal in typical Keplerian flows (for our setup, $E_{\rm kin}/E_{\rm th} \approx \frac{\gamma-1}{2h^2} = 80$), numerical errors in the calculation of total energy could affect the thermal energy budget of the disk due to subtractive cancellation error. In order to check this effect, we reran our fiducial model using the ENTROPY_SWITCH option of PLUTO, which ensures entropy conservation outside of the vicinity of shocks (which by definition do not conserve entropy, but are captured accurately by the Riemann solver). We found that this did not affect the life track of the generated vortex.

Finally, we also reran the fiducial model with PLUTO using a 3rd-order solver and parabolic reconstruction instead of the standard 2nd-order solver and linear reconstruction setup. We found no differences in vortex evolution or lifetime.

On the basis of our tests and the agreement of the codes for high resolution, we conclude that the vortex dynamics and effects we observed in our simulations are not numerical artifacts but that they are indeed physical.

7. Summary

We studied vortices created by planets in protoplanetary disks using two-dimensional viscous hydrodynamics simulations. The equation of state was assumed to follow an ideal gas, turbulence was included following the α parametrization, and thermal processes were considered by prescribing a thermal relaxation timescale using the β formalism. A focus was brought to vortices exterior to the gap opened by the planet. In order to verify our results, the simulations were carried out with both the FARGO and PLUTO codes which use different numerical schemes. The planet was treated as a non-accreting point mass with a smoothed gravitational potential and kept on a fixed circular orbit. Properties of vortices were automatically extracted using our newly-developed Vortector Python tool, which identifies and characterizes vortices. Vortex identification was performed by looking for elliptical shapes in iso-vortensity lines in the r-ϕ plane, and characterization was performed by fitting a 2D Gaussian to the vortensity and surface density.

Vortices formed during the gap opening process as the embedded Jupiter-mass planet was introduced into the simulation. At the outer gap edge, multiple small vortices formed that usually merged into a single large vortex that lived, depending on parameters, between 200 and several thousand orbits. These vortices had a full width at half maximum (as determined by the fitted 2D Gaussian) of up to $0.4\,r_{\rm p}$ (several au for a planet at $r_{\rm p} = 5.2$ au). The mass enclosed in this vortex area was up to one planetary mass (one Jupiter-mass in our models) for our choice of disk mass.

Vortex lifetime depended on the thermal relaxation timescale such that vortices lived shortest for intermediate cooling times ($\beta = 1$), a result also found by Fung & Ono (2021). We found two regimes for the vortices' lifetimes. A short-lived regime, with vortex lifetimes of up to 3000 orbits, was observed for slowly-cooling disks ($\beta \geq 1$), in which the vortices decayed faster than expected from viscous dissipation alone. In the long-lived regime, which was observed for fast cooling ($\beta \ll 1$) with the isothermal assumption as an extreme, vortices lived for a much longer time and did not decay rapidly. Vortex lifetimes were considerably longer in this regime, with a lower bound on the maximum lifetime being 15 000 orbits (the model was terminated while the vortex was still alive due to runtime constraints). From our analysis, we suspect that the long lifetime for small β is connected to the interaction of the vortex with the spiral arms, which are a source of vorticity. Details are left to future studies.

Additionally, including the disk's self-gravity in the models usually shortened the lifetime of vortices and stopped the small initial vortices from merging into one large vortex. Usually, two smaller vortices remained after the initial gap opening process, which then decayed faster compared to those in models where disk self-gravity was not accounted for. This finding, that self-gravity is detrimental to vortex survival, is in line with previous studies (Lovelace & Hohlfeld 2013; Regály & Vorobyov 2017; Zhu & Baruteau 2016).

Outward migration of the vortex was observed in some of the models with $\beta \ll 1$ and $\beta \gg 1$. In those cases, a second density (and thus pressure) bump formed outside of the vortex location, towards which the vortex then migrated (Paardekooper et al. 2010). In some $\beta = 0.01$ models, a small, short-lived, "secondary" vortex formed between the planet gap and the "primary" vortex.

Concerning the dependence of vortex lifetime on viscosity, we found the expected behavior that this lifetime was shorter for higher viscosity. For the highest viscosity of $\alpha = 10^{-3}$, practically no vortices were observed. For $\alpha = 10^{-5}$ and 10^{-6} we found nearly identical results, suggesting that the numerical viscosity in our models with a resolution of 8 and 16 cells per scale height was of the order of $\alpha_{\rm 8cps} \lesssim 10^{-5}$ and $\alpha_{\rm 16cps} \approx 10^{-6}$.

Allowing the planet to grow over a longer time, 1000 instead of 100 orbits, led to longer vortex lifetimes in all the cases we tested. This disagrees with the findings of Hammer et al. (2017), who found reduced vortex lifetimes for longer planet-growth times. In our models, vortices took longer to form in the case of the slower-growing planet. During their decay, however, their evolution was very similar, independent of planet introduction time (see Fig. 6), which in total increased their lifetime. The fact that vortex lifetime increased for longer planet-growth

timescales can be an indication that the effects presented in this study, including the long-lived vortex regime, are also applicable to longer, and arguably more realistic, planet-growth timescales of around 10 000 orbits.

Estimating vortex lifetime from our results, vortices are expected to live much longer at larger distances away from their host star. The increase in expected lifetime is firstly due to the longer orbital period at large radii, but also because the expected β values — the thermal relaxation timescale compared to the orbital timescale – are much lower and vortices then likely belong to the long-lived regime (see Sect. 5). From order-of-magnitude calculations, we find that large planet-induced vortices exterior to the planet at 50–100 au might live for several Myr. Considering the sensitivity of instruments like ALMA at these distances from the star, this suggests that these vortices should be observable more easily than planet-induced vortices at smaller radii.

Acknowledgements. TR and WK acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG) research group FOR 2634 "Planet Formation Witnesses and Probes: Transition Disks" under grant DU 414/22-1 and KL 650/29-1, 650/30-1. WB, WK, and AZ acknowledge support by the DFG-ANR supported GEPARD project (ANR-18-CE92-0044 DFG: KL 650/31-1). The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 37/935-1 FUGG. Plots in this paper were made with the Python library `matplotlib` (Hunter 2007).

References

Bae, J., Zhu, Z., & Hartmann, L. 2016, ApJ, 819, 134
Balbus, S. A. & Hawley, J. F. 1991, ApJ, 376, 214
Barge, P. & Sommeria, J. 1996, in NASA Conference Publication, Vol. 3343, NASA Conference Publication, 179–182
Benítez-Llambay, P. & Masset, F. S. 2016, ApJS, 223, 11
Bradski, G. 2000, Dr. Dobb's Journal of Software Tools
Cimerman, N. P. & Rafikov, R. R. 2021, arXiv e-prints, arXiv:2108.01423
Crida, A., Morbidelli, A., & Masset, F. 2006, Icarus, 181, 587
De Val-Borro, M., Edgar, R. G., Artymowicz, P., et al. 2006, MNRAS, 370, 529
Dullemond, C. P. 2000, A&A, 361, L17
Dullemond, C. P., Birnstiel, T., Huang, J., et al. 2018, ApJ, 869, L46
Flaherty, K. M., Hughes, A. M., Teague, R., et al. 2018, ApJ, 856, 117
Flock, M., Nelson, R. P., Turner, N. J., et al. 2017, ApJ, 850, 131
Flock, M., Turner, N. J., Nelson, R. P., et al. 2020, ApJ, 897, 155
Fung, J. & Ono, T. 2021, arXiv e-prints, arXiv:2108.06345
Hammer, M., Kratter, K. M., & Lin, M.-K. 2017, MNRAS, 466, 3533
Hammer, M., Lin, M.-K., Kratter, K. M., & Pinilla, P. 2021, MNRAS, 504, 3963
Hammer, M., Pinilla, P., Kratter, K. M., & Lin, M.-K. 2019, MNRAS, 482, 3609
Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90
Kanagawa, K. D., Tanaka, H., Muto, T., & Tanigawa, T. 2017, PASJ, 69
Keppler, M., Benisty, M., Müller, A., et al. 2018, A&A, 617, A44
Klahr, H. H. & Bodenheimer, P. 2003, ApJ, 582, 869
Kley, W. & Nelson, R. P. 2012, ARA&A, 50, 211
Lega, E., Nelson, R. P., Morbidelli, A., et al. 2021, A&A, 646, A166
Lesur, G. & Papaloizou, J. C. B. 2009, A&A, 498, 1
Lesur, G. & Papaloizou, J. C. B. 2010, A&A, 513, A60
Levermore, C. D. & Pomraning, G. C. 1981, ApJ, 248, 321
Li, H., Li, S., Koller, J., et al. 2005, ApJ, 624, 1003
Lin, M.-K. & Papaloizou, J. C. B. 2011, MNRAS, 415, 1426
Lovelace, R. V. E. & Hohlfeld, R. G. 2013, MNRAS, 429, 529
Lovelace, R. V. E., Li, H., Colgate, S. A., & Nelson, A. F. 1999, ApJ, 513, 805
Marcus, P. S., Pei, S., Jiang, C.-H., & Barranco, J. A. 2016, ApJ, 833, 148
Marcus, P. S., Pei, S., Jiang, C.-H., et al. 2015, ApJ, 808, 87
Marel, N. v. d., Dishoeck, E. F. v., Bruderer, S., et al. 2013, Science, 340, 1199
Masset, F. 2000, A&AS, 141, 165
McNally, C. P., Nelson, R. P., Paardekooper, S.-J., & Benítez-Llambay, P. 2019, MNRAS, 484, 728
Mignone, A., Bodo, G., Massaglia, S., et al. 2007, ApJS, 170, 228
Mignone, A., Flock, M., Stute, M., Kolb, S. M., & Muscianisi, G. 2012, A&A, 545, A152
Miranda, R. & Rafikov, R. R. 2020a, ApJ, 892, 65
Miranda, R. & Rafikov, R. R. 2020b, ApJ, 904, 121
Müller, T. W. A., Kley, W., & Meru, F. 2012, A&A, 541, A123
Nelson, R. P., Gressel, O., & Umurhan, O. M. 2013, MNRAS, 435, 2610
Ogilvie, G. I. & Lubow, S. H. 2002, MNRAS, 330, 950
Paardekooper, S.-J., Lesur, G., & Papaloizou, J. C. B. 2010, ApJ, 725, 146
Pérez, L. M., Benisty, M., Andrews, S. M., et al. 2018, ApJ, 869, L50
Rafikov, R. R. 2002, ApJ, 569, 997
Rafikov, R. R. 2016, ApJ, 831, 122
Regály, Z. & Vorobyov, E. 2017, MNRAS, 471, 2204
Rometsch, T., Rodenkirch, P. J., Kley, W., & Dullemond, C. P. 2020, A&A, 643, A87
Shakura, N. I. & Sunyaev, R. A. 1973, A&A, 500, 33
Tassoul, J.-L. 1978, Theory of rotating stars
Val-Borro, M. d., Artymowicz, P., D'Angelo, G., & Peplinski, A. 2007, A&A, 471, 1043
Zhang, S. & Zhu, Z. 2020, MNRAS, 493, 2287
Zhang, S., Zhu, Z., Huang, J., et al. 2018, ApJ, 869, L47
Zhu, Z. & Baruteau, C. 2016, MNRAS, 458, 3918
Ziampras, A., Ataiee, S., Kley, W., Dullemond, C. P., & Baruteau, C. 2020a, A&A, 633, A29
Ziampras, A., Kley, W., & Dullemond, C. P. 2020b, A&A, 637, A50

Appendix A: Grid resolution and numerical convergence

A grid resolution of 8 cells per scale height (cps) is often adopted in models of planet–disk interaction in literature. While it is widely agreed upon as sufficient, we test this statement by performing a series of test simulations with both `PLUTO` and `FargoCPT` using the same physical parameters as our locally isothermal models ($\alpha = 10^{-5}$), but using varying grid resolutions with 1, 2, 4, 8 and 16 cps in both directions (always maintaining square cells).

We found that we achieve numerical convergence on large-scale features such as the gap width and pressure bumps for a resolution of 4 cps. Convergence on more numerically-sensitive features such as gap depth and vortex formation is reached for a resolution of 8 cps, with 16 cps affecting the picture relatively weakly. This was observed across both codes, with the two showing very good agreement with each other both in terms of the resolution at which different features converge and the physical properties of said features across codes.

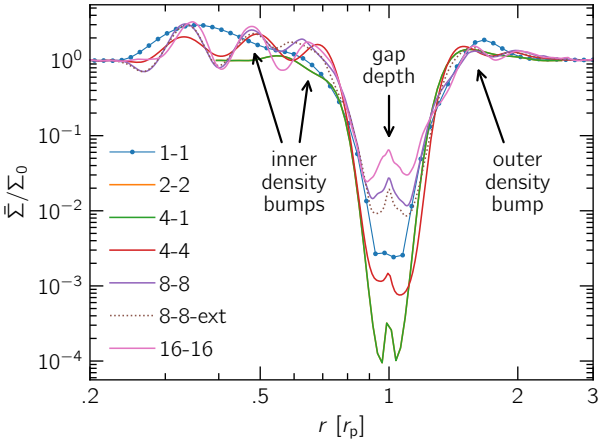

Fig. A.1. Results for our resolution study using `FargoCPT`. The overall shape of the gap is resolved with around 4 cps, while it takes 8 cps to properly resolve the gap depth and the contrast of most pressure bumps far from the gap. We are interested in the region between 0.5–2.0 r_p. Extending the outer boundary to $r = 10\,r_p$ in the "8-8-ext" model practically made no difference. It should be noted that the 16×16 cps model develops some small-scale vortices in the inner disk, which causes these differences around $0.7\,r_p$. Interestingly, a model that resolves the radial and azimuthal directions with 4 and 1 cps, respectively, captures these radial features almost as well as one with 4 cps in both directions.

Appendix B: The Vortector

A major task in this study was the identification and characterization of vortices in simulation data. For this purpose, we developed a Python package, the `Vortector`, that automates the process for relatively generic 2D hydrodynamics planet–disk simulations.

The `Vortector` package lets one visualize the vortex detection results (an example is shown in Fig. B.2), as well as includes information about the location, extent, and mass of a vortex along with various statistics related to the contour.

The package is publicly available on GitHub[1]. We hope to make the detection and characterization of vortices in simulation data easier for other members of the community and facilitate quantitative comparison of vortices between studies by providing a common detection pipeline.

To search for possible vortex candidates, a simple search for the location of minimum vortensity is sometimes enough to find the location of a vortex. Then, the value of the vorticity $\omega = (\nabla \times \boldsymbol{u}) \cdot \hat{z}$ can be used to learn how strongly the vortex rotates and the local surface density can be used as an indication for the mass enclosed in the vortex. This method, however, fails for many simulations, e.g., when the vortensity in a tiny region close to a spiral arm of the planet is lower than inside a vortex candidate, or when the gap region intrudes into the outer disk, which can induce strong anticyclonic motion at the outer gap edge.

To get around these issues, the `Vortector` uses the geometrical shape of vortices as they appear in a face-on image of the disk. Looking down on the surface of a disk, vortices appear as crescent-shaped objects. In the r-ϕ plane, which is more suitable for this task, large vortices appear as elliptical objects (see also Fig. 1 of Lesur & Papaloizou 2009). In fact, contour lines of the vortensity closely resemble ellipses in the r-ϕ plane. We can therefore identify vortices in a disk by finding closed contour lines that closely resemble ellipses. To solve this task programmatically, we can make use of the computer vision library `OpenCV` (Bradski 2000).

Our strategy to extract vortex candidates from a simulation snapshot can be then subdivided into three tasks:

1. Extract contour lines in the r-ϕ map of the vortensity,
2. identify nearly elliptical contours as vortex candidates, and
3. fit 2D Gaussians to ϖ and Σ for characterization.

The algorithm step by step

This section describes the vortex detection process using the model presented in Sect. 5.1 (`FargoCPT`, $\alpha = 10^{-5}$, $\beta = 0.01$, 16 cps) which shows the emergence of a secondary vortex. The data used for this analysis corresponds to a time $t = 7150$ orbits.

Before the analysis is performed on ϖ, the map is periodically extended in the ϕ direction in order to be able to identify vortices that intersect the periodic azimuthal boundary. The resulting image is shown in the left panel of Fig. B.1. There, the original domain is indicated by the thick solid rectangle which spans the azimuthal range from 0 to N, where N is the image size in pixels in the azimuthal direction. The top and bottom areas of the domain (orange and green) are repeated at the lower and upper boundaries, respectively.

Task 1: contour lines

Contour lines are extracted for a range of ϖ values ranging from 0 to 1 in increments of 0.05. For each value $\varpi_{\rm crit}$, a binary image is produced by setting each cell with $\varpi \geq \varpi_{\rm crit}$ to 1 and 0 otherwise. The binary image is then analyzed using `findContours` from `OpenCV`. Only closed contours are retained. This step usually results in up to a few thousand contours, depending on the dynamical state of the disk and the choice of increments in ϖ.

[1] https://github.com/rometsch/vortector

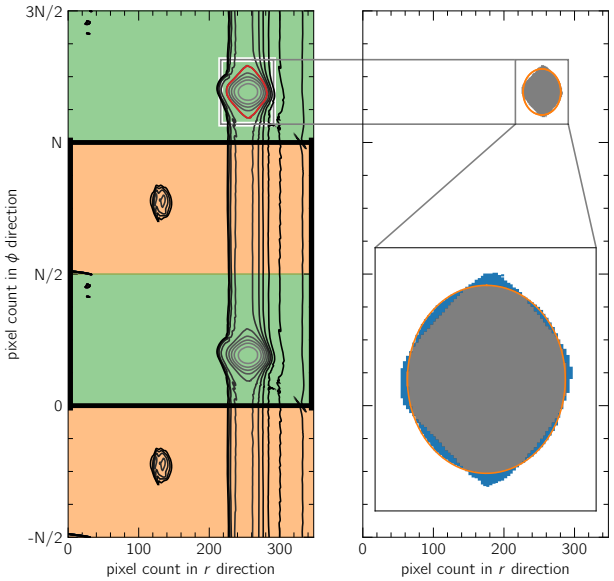

Fig. B.1. Periodically continued iso-vortensity line image (left) used for extracting contours for vortex candidates and an example contour illustrating the ellipse fit (right). The snapshot shown is at time t = 7150 orbits of the model with the "secondary" vortex that was discussed in Sect. 5.1. The left panel shows how the data array is mirrored in order to allow the detection of vortices that overlap with the periodic boundary. Areas with the same color are copies of one another. The red line indicates the outline of the grey area in the right panel. The original size is marked by the black rectangle ranging from 0 to N on the vertical axis. The area shaded in blue in the right panel illustrates the definition of the deviation from the ellipse that is used to select the vortex candidates from the closed contours. For the example shown, the ratio of difference area (blue) to the total contour area is 0.122, which is below the 0.15 threshold.

Task 2: Find closed contours resembling ellipses

Next, the `fitEllipse` function from `OpenCV` is used to fit an ellipse to each closed contour. One example of this is shown in the right panel of Fig. B.1, where the ellipse is visible as an orange line in the zoom-in.

The difference in area between the contour and fit is used as a measure of deviation. The deviation from an ellipse is then defined as the ratio of this difference and the area enclosed by the contour. We only keep contours for which the deviation is smaller than 0.15. The example contour in Fig. B.1 has a deviation of 0.122.

Finally, all the contours that are contained within the largest contour that satisfies this criterion are discarded, which leaves the example red contour in Fig. B.1 as the selected vortex candidate (see also the white contour line in Fig. B.2).

We only retain contours that enclose at least two other contours. With this restriction, we make sure that ϖ/ϖ_0 changes by a value of 0.1 from the outside to the inside of the vortex candidate. This has proven to be useful to filter out small fluctuations in the disk that otherwise appear as small transient vortices.

At this point, it becomes clear that the extent of the vortex and derived quantities such as the mass contained within are influenced by the choice of the levels used to produce the contour lines and the choice of the maximum relative ellipse deviation. The properties of the contour give an order-of-magnitude estimate nonetheless.

Task 3: Fit a 2D Gaussian

To remove the influence of the threshold parameters in the detection of the contour, a process that does not depend on our parameter choices but on the underlying data is needed.

Upon inspection of the curves of vortensity and surface density along a cut through the vortex, either radial or azimuthal, it becomes clear that these lines resemble Gaussian functions (see curves in Fig. B.2 around the 2D maps)

$$f(r,\phi) = c + a \exp\left(-\frac{(r-r_0)^2}{2\sigma_r^2}\right) \exp\left(-\frac{(\phi-\phi_0)^2}{2\sigma_\phi^2}\right). \tag{B.1}$$

Here, σ_r and σ_ϕ provide a measure for the vortex size and can even be used to give a definition of the vortex region that does not depend on additional parameters. In combination with the center coordinates r_0 and ϕ_0, σ_r and σ_ϕ can be used to define the vortex as the disk material contained within the ellipse given by

$$\left(\frac{r-r_0}{h_r}\right)^2 + \left(\frac{\phi-\phi_0}{h_\phi}\right)^2 = 1\,, \tag{B.2}$$

where $h_r = \sqrt{2\ln(2)}\,\sigma_r$ and $h_\phi = \sqrt{2\ln(2)}\,\sigma_\phi$ denote the half width at half maximum of the 2D Gaussian function defined in Eq. (B.1). We usually use the values obtained from the surface density fit because these are less time-sensitive compared to the vortensity fit and because the shape of Σ curves more closely resemble Gaussians (see Fig. B.2).

Appendix C: Data table

The lifetimes and parameters of all models mentioned in Sects. 4 and 5 are listed in Table C.1.

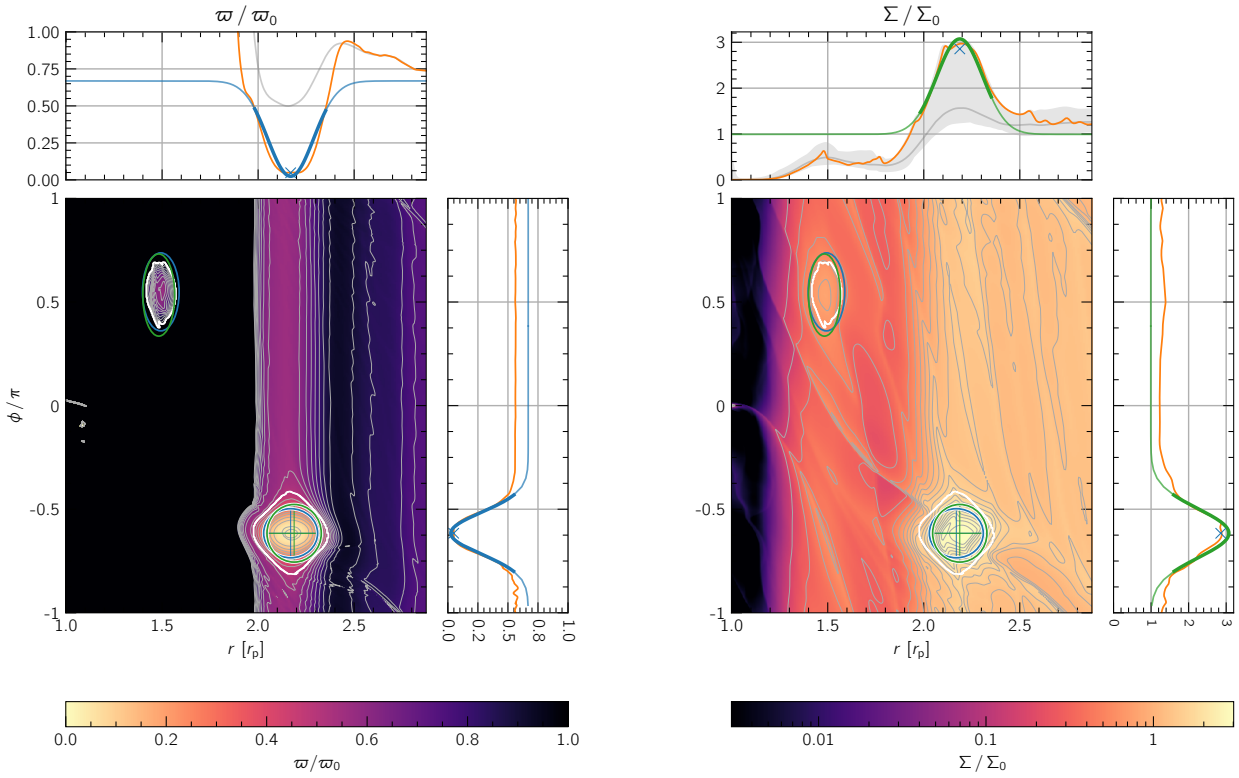

Fig. B.2. Overview of the results produced by the `Vortector` package for a model showing a secondary vortex discussed in Sect. 5.1 at t = 7150 orbits, with a 2D map of the vortensity on the left and surface density on the right. All detected vortex candidates are indicated in the 2D plots. The extracted contour (shown in Fig. B.1) is marked with a white line, the ellipse of the vortensity fit is shown in blue and the ellipse of the surface density fit is shown in green. Note that these ellipses are defined by σ_r and σ_ϕ from the fit of Eq. (B.2) and are different from the ellipse used to fit the contour. The ellipses of the most massive vortex include a crosshair indicating the center of the fit. Each 2D plot is accompanied by 1D plots of slices through the main vortex. They also show the values of the respective Gaussian fit in blue for vortensity and green for surface density. In this figure, the planet is located at $r = 1$ and $\phi = 0$.

Table C.1. Lifetimes of vortices in the simulations. Models with a resolution of 8, 16, and 32 cps have 528×1024, 1056×2048, and 2112×4096 cells, respectively.

code	β	α	cps	τ_{intro} [a]	special [b]	in Fig. 4	T_{vort} [c]
fargo	0.01	10^{-6}	16	100		x	>8784
pluto	0.01	10^{-6}	16	...		x	9548
fargo	0.01	10^{-6}	8			x	1679
pluto	0.01	10^{-6}	8			x	2013
fargo	0.01	10^{-5}	32				>739
fargo	0.01	10^{-5}	16			x	8774
fargo	0.01	10^{-5}	16		sg		>4567
pluto	0.01	10^{-5}	16			x	>15100
pluto	0.01	10^{-5}	16				5762
pluto	0.01	10^{-5}	16		vort rem		1843
fargo	0.01	10^{-5}	8	1000			3897
fargo	0.01	10^{-5}	8			x	2288
fargo	0.01	10^{-5}	8				7845
fargo	0.01	10^{-5}	8		sg	x	1299
fargo	0.01	10^{-5}	8	1000	sg		1269
pluto	0.01	10^{-5}	8			x	2611
pluto	0.01	10^{-5}	8	1000			2691
pluto	0.01	10^{-5}	8				10638
fargo	0.01	10^{-4}	16			x	1129
pluto	0.01	10^{-4}	16			x	1335
fargo	0.01	10^{-4}	8			x	979
pluto	0.01	10^{-4}	8			x	887
fargo	0.01	10^{-3}	16			x	79
pluto	0.01	10^{-3}	16			x	0
fargo	0.01	10^{-3}	8			x	69
pluto	0.01	10^{-3}	8			x	29
fargo	1	10^{-6}	16			x	1309
pluto	1	10^{-6}	16			x	1694
fargo	1	10^{-6}	8			x	1189
pluto	1	10^{-6}	8			x	1106
fargo	1	10^{-5}	16			x	1359
fargo	1	10^{-5}	16		sg		649
pluto	1	10^{-5}	16			x	1445
pluto	1	10^{-5}	16		art vort		916
fargo	1	10^{-5}	8	1000			1279
pluto	1	10^{-5}	8			x	1179
fargo	1	10^{-5}	8		sg	x	289
fargo	1	10^{-5}	8	1000	sg		509
pluto	1	10^{-5}	8			x	1136
pluto	1	10^{-5}	8	1000			956
fargo	1	10^{-4}	16			x	709
pluto	1	10^{-4}	16			x	757
fargo	1	10^{-4}	8			x	659
pluto	1	10^{-4}	8			x	607
fargo	1	10^{-3}	16			x	0
pluto	1	10^{-3}	16			x	49
fargo	1	10^{-3}	8			x	59
pluto	1	10^{-3}	8			x	0
fargo	100	10^{-6}	16			x	1729
pluto	100	10^{-6}	16			x	2441
fargo	100	10^{-6}	8			x	3108
pluto	100	10^{-6}	8			x	946
fargo	100	10^{-5}	16			x	1888
pluto	100	10^{-5}	16			x	1943
fargo	100	10^{-5}	8	1000			3048
fargo	100	10^{-5}	8			x	2708
fargo	100	10^{-5}	8		sg	x	589
fargo	100	10^{-5}	8	1000	sg		499
pluto	100	10^{-5}	8			x	857
pluto	100	10^{-5}	8	1000			0
fargo	100	10^{-4}	16			x	699
pluto	100	10^{-4}	16			x	797
fargo	100	10^{-4}	8			x	569
pluto	100	10^{-4}	8			x	478
fargo	100	10^{-3}	16			x	0
pluto	100	10^{-3}	16			x	0
fargo	100	10^{-3}	8			x	0
pluto	100	10^{-3}	8			x	0

Notes. [a] Planet introduction time. 100 orbits if empty. [b] Special propertiy of the model. "sg" if self-gravity is included. "vort rem" and "art vort" refer to the models discussed in Sect. 6.1 with the removed vortex and the artificial vortex, respectively. [c] Lifetime of the vortex in planetary orbits. ">" indicates that the vortex still exists at the end of the simulation.