UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title
Monte Carlo Radiation Transport Simulations of Asymmetric Neutron Star Mergers

Permalink
https://escholarship.org/uc/item/18x9790q

Author
Klion, Hannah

Publication Date
2021

Peer reviewed|Thesis/dissertation

Monte Carlo Radiation Transport Simulations of Asymmetric Neutron Star Mergers

by

Hannah Klion

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eliot Quataert, Chair
Associate Professor Daniel Kasen
Professor Joshua S. Bloom

Summer 2021

Monte Carlo Radiation Transport Simulations of Asymmetric Neutron Star Mergers

Abstract

Monte Carlo Radiation Transport Simulations of Asymmetric Neutron Star Mergers

by

Hannah Klion

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Eliot Quataert, Chair

The merger of two neutron stars produces a neutron-rich outflow of $> 10^{-2}\,\mathrm{M_\odot}$. This ejecta is a prime site for r-process nucleosynthesis, which will produce a range of radioactive heavy nuclei. Within a second of merger, an accretion disc forms around the central remnant. Material from this disc accretes onto the central compact object, launching a relativistic jet. The disc is also the source of a mildly-relativistic radioactive wind. Heating by r-process decays can accelerate accelerate portions of this wind by $\sim 0.1\,c$. Within a few hours, emission from the radioactive material gives rise to an ultraviolet, optical, and infrared transient: a kilonova. These ejecta are not expected to be spherically symmetric, so the emission will depend on the multi-dimensional ejecta structure as well as the viewing angle of the observer. Calculating this emission requires solving the time-dependent equation of radiation transport. To do so, I use Sedona, a parallel multi-dimensional, multi-frequency Monte Carlo radiation transport code. I analyze the parallel structure and performance of Sedona on modern supercomputers and discuss the performance and usability improvements I have made to the code. I use the endstates of a suite of 2D relativistic hydrodynamic simulations of jet-ejecta interaction as initial conditions for Sedona calculations, producing viewing angle-dependent light curves and spectra starting at $1.5\,\mathrm{h}$ after merger. I find that on this time-scale, jet shock heating does not affect the kilonova emission for the jet parameters surveyed. However, the jet disruption to the density structure of the ejecta does change the light curves. The jet carves a channel into the otherwise spheroidal ejecta, revealing the hot, inner regions, and making polar emission brighter and bluer. The winds launched from post-merger accretion discs are also aspherical. I continue a 3D general relativistic magneto-hydrodynamic simulation of a disc wind in 2D hydrodynamics, and follow it until the flow is self-similar. I find that the inclusion of r-process heating doubles the mass-weighted median velocity, making the light curves brighter and bluer at $\sim 1\,\mathrm{d}$ post-merger.

To my parents

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation and the years of work that led up to it would not have been possible without my family — my parents, Reid and Laurie Klion, my grandparents, Jack and Doris Mann and Lila and Jerry Klion, and my uncle Phil Klion — who have supported of all of my educational goals, starting when I declared at eight years old that I wanted to get a PhD. (Did I really understand what that was? No. Did I know what it would entail? Definitely not.) From ISI, to Caltech, to Berkeley, they have provided constant love and support. A special thanks to my grandfather Jerry Klion for buying me my first scientific calculator in the sixth grade and encouraging me to forge ahead with my newfound interest in math.

To my advisor, Eliot Quataert, thank you for your guidance and support throughout the last seven years, and for pushing me to see the bigger picture when I got bogged down in the details. I also deeply appreciate the space you gave me to figure out my career goals and the direction I wanted to take my work.

To Dan Kasen, thank you for being my unofficial co-advisor. Our very frequent meetings at the start of quarantine helped so much with the adjustment to isolation and working from home. Thank you for letting me have free rein over SEDONA's workflow and code. Making it run faster and better has been such a rewarding project.

My PhD work was largely supported by the Department of Energy Computational Science Graduate Fellowship. Interacting with fellows and alumni has been one of the highlights of graduate school. My CSGF practicum at Oak Ridge National Lab was instrumental in helping me realize I wanted to focus on the "computational" part of "computational science." Thank you to Bronson Messer, Austin Harris, Judy Hill, Chloe and Michael Sandoval, and the rest of the Scientific Computing and Nuclear Astrophysics groups for making it a great experience.

I overlapped with many graduate students and postdocs, who enriched my day-to-day life in Campbell Hall. I have had the pleasure of collaborating with several of you on the work within this dissertation, side projects, and outreach around the Bay Area. Thank you to you all, especially to Ellianna Abrahams, Jennifer Barnes, Paul Duffell, David Khatami, Aashrita Mangu, Philipp Mösta, Abigail Polin, and Wren Suess, who have been wonderful friends as well as colleagues.

To all of the friends I've made in the last six years: I would not have made it through classes, conferences, and research frustrations without you. A special thank you goes to Roger Curley, Maia Werbos, Morgan Presley, and Ariel Hansen. Your friendship has been invaluable. Finishing my PhD while working from home amid a pandemic was an unexpected challenge. You all made it so much more enjoyable.

To Jen Karolewski and Kayla McCue, thank you for being constants in my life for the last ten years. Even though we're now on opposite coasts, I couldn't be happier to go through Caltech and now graduate school with the two of you.

And finally, I owe an endless debt of gratitude to my partner, Allison Kong. Your friendship, love, and support have made all of this possible.

Chapter 1

Introduction

1.1   Neutron Star Mergers

Six years ago, the Laser Interferometer Gravitational-Wave Observatory (LIGO) (Abbott et al., 2009) was in the midst of a decades-long quest for the first direct detection of gravitational waves. The main target of the LIGO project was the inspiral and merger of neutron star (NS) and black hole (BH) binary systems. If the binary separation is small enough, these binaries lose appreciable orbital and angular momentum to gravitational radiation, and the orbits will decay within the life time of the universe. In the last six years, dozens of such systems have been detected, beginning with the first binary black hole coalescence in 2015 (Abbott et al., 2019, 2021).

GW170817 was the first gravitational wave detection of an NS-NS merger (Abbott et al., 2017a). It was followed by a short gamma ray burst (GRB) 1.7 s later (Abbott et al., 2017c; Goldstein et al., 2017; Savchenko et al., 2017). While the exact gamma-ray emission mechanism remains uncertain, subsequent radio observations have found that the afterglow appears to move superluminally across the sky and is consistent with viewing a collimated jet at 20° from the binary's orbital axis (Mooley et al., 2018a).

An ultraviolet, optical, and infrared counterpart, AT 2017gfo (also SSS17a), was detected hours later, consistent with a $v \sim 0.1 \, c$ outflow of $> 0.01 \, \mathrm{M_\odot}$ of radioactive material with an average opacity of 1–$3 \, \mathrm{cm^2 g^{-1}}$ and a total kinetic energy of $\sim 10^{51} \, \mathrm{erg}$ (Abbott et al., 2017b; Arcavi et al., 2017; Kasliwal et al., 2017; Drout et al., 2017; Villar et al., 2017; Coulter et al., 2017; Cowperthwaite et al., 2017; Chornock et al., 2017; Soares-Santos et al., 2017). The transient was rather red and evolved on a week time-scale. This was consistent with predictions for a NS merger counterpart, a "kilonova" (Li & Paczyński, 1998; Metzger et al., 2010; Metzger & Berger, 2012; Barnes & Kasen, 2013).

The red color suggests that the material producing the transient has a high opacity, which is characteristic of rapid- (r-)process nucleosynthesis. In the midst of a merger, the ejected material undergoes rapid decompression from nuclear densities. Once the material cools below $\sim 10 \, \mathrm{GK}$ and leaves nuclear statistical equilibrium (NSE), neutrons capture onto light

seed nuclei faster than the nuclei can undergo beta decays. This r-process nucleosynthesis produces a range of radioactive neutron-rich nuclei that beta decay to stability over weeks (Lattimer et al., 1977; Eichler et al., 1989; Metzger et al., 2010). The bulk of the heat released from r-process decays is deposited in the first few seconds after the material leaves NSE. It then falls off approximately as a power law, powering the kilonova (Li & Paczyński, 1998; Metzger et al., 2010).

The distribution of nuclei produced by the r-process primarily depends on the electron fraction $Y_{\rm e} \equiv n_p/(n_p + n_n)$ of the nuclear material. The proton and neutron densities are n_p and n_n, respectively. When $Y_{\rm e} \lesssim 0.25$, the r-process produces material enriched with Lanthanides and Actinides, which have uniquely high opacities. Their partially-filled f orbitals produce a high density of moderately strong atomic lines, which lead to a high quasi-continuum opacity ($\kappa \sim 10\,{\rm cm^2 g^{-1}}$), especially in the blue and ultraviolet (Kasen et al., 2013; Tanaka et al., 2020). Low $Y_{\rm e}$ ejecta would therefore produce a red, dim, slowly-evolving transient that is bright in the near-infrared (Barnes & Kasen, 2013). By contrast, higher-$Y_{\rm e}$ material gives rise to lighter r-process elements, which primarily occupy the second row of the d-block (i.e. transition metals). These elements have lower average opacities $\kappa \sim 0.5$–$1\,{\rm cm^2 g^{-1}}$ that are again highly wavelength dependent.

The light curve of AT 2017gfo can be modeled by (at least) two distinct components with opacities that correspond to light and heavy r-process products (Cowperthwaite et al., 2017; Kasen et al., 2017; Perego et al., 2017; Villar et al., 2017). Both components are consistent with being heated by r-process decays. At early times, emission is largely due to a "blue" component associated with material rich in elements near the first r-process peak. Within a few days, there is a transition to the "red" emission, from material enriched by high-opacity Lanthanides around the second r-process peak.

The association of a GRB with neutron star mergers also matched predictions. Short GRBs can release large amounts of energy ($\sim 10^{54}\,{\rm erg}$, isotropic-equivalent) in gamma rays, with variability on short ($\sim$ ms) time-scales. The time-scale and high energies require a highly relativistic outflow (Lorentz factor $\gamma > 100$). The merger of two neutron stars or a neutron star and black hole was proposed as the central engine (Paczynski, 1986; Eichler et al., 1989; Narayan et al., 1992). A compact binary origin for short GRBs is also consistent with their correlation with older stellar populations (Bloom et al., 2006; Leibler & Berger, 2010) and their larger galactic offsets relative to long GRBs and supernovae (Bloom et al., 2002; Fong & Berger, 2013). The GRB associated with GW170817 was a low-energy, off-axis burst (Abbott et al., 2017c).

GW170817 confirmed much of the theoretical picture of neutron star mergers. Advanced LIGO's third observing run, which ended in early 2020, detected one additional plausible NS-NS system (GW190425; Abbott et al., 2020). Due in part to a combination of the large localization area and the distance of the event, follow-up efforts were unable to identify a kilonova.

1.2 Asymmetry in Neutron Star Mergers

The upcoming observing runs of Advanced LIGO are expected to detect around 0.2 to 2 neutron star mergers per year, some of which will have detectable electromagnetic counterparts (Abbott et al., 2021). Understanding the expected diversity of kilonovae will assist in targeting counterpart searches and in the interpretation of subsequent events. Variations are expected in the observations due to intrinsic differences in the amount, composition, and distribution of the ejecta (Gompertz et al., 2018).

It is also expected that some components of neutron star merger ejecta are non-spherical, and may show latitudinal variation in density and opacity. A number of recent studies have examined how ejecta components with different geometry and opacity can affect kilonova light curves (Wollaeger et al., 2018; Kawaguchi et al., 2020; Darbha & Kasen, 2020; Korobkin et al., 2021; Nativi et al., 2021). Our viewing angle relative to the binary plane for each of these events will be different, so understanding how identical events could appear different at different viewing angles is an important aspect of the light curve diversity.

As an NS-NS binary merges, mass ejection occurs on dynamical ($\sim$ ms) time-scales. The neutron stars can tidally disrupt one another, and their collision can squeeze material out along the orbital axis. Numerical simulations predict that dynamical ejecta will consist of 10^{-4}–$10^{-2}\,M_\odot$ of material with escape velocities of $\sim$ 0.1–0.3 c (Bauswein et al., 2013; Hotokezaka et al., 2013a; Lehner et al., 2016; Bovard et al., 2017; Dietrich et al., 2017; Radice et al., 2018). The overall density distribution is expected to be an oblate ellipsoid (Hotokezaka et al., 2013a). Tidally-driven outflows are concentrated near the orbital plane, and are expected to have a low Y_e. The squeezed material has a wider angular distribution and is shock-heated, giving it a higher Y_e (Radice et al., 2018).

Some of the disrupted material is gravitationally bound and can form an accretion disc of up to $0.3\,M_\odot$ that evolves on longer time-scales (100 ms–10 s) (Oechslin et al., 2007; Hotokezaka et al., 2013b). Initially, neutrino cooling dominates because the disc is hot and dense (Popham et al., 1999; Narayan et al., 2001). As the disc cools and spreads, neutrino cooling becomes inefficient, and the disc becomes fully advective. Weak interactions freeze out, which can lead to a strong neutron-rich wind (Metzger et al., 2009). Recent 3D general relativistic magnetohydrodynamic (GRMHD) simulations have found that these winds can comprise half of the disc mass (Fernández et al., 2019; Christie et al., 2019). The winds tend to be prolate and show compositional stratification in both radius and meridional angle (e.g. Fernández et al., 2019; Kawaguchi et al., 2020).

R-Process Heating in Disc Winds

Many studies of post-merger accretion discs have relied on hydrodynamic simulations, where an artificial viscosity drives angular momentum transport (e.g. Fernández & Metzger, 2013; Just et al., 2015; Fujibayashi et al., 2018). Siegel & Metzger (2017, 2018) presented the first 3D GRMHD simulations of the post-merger accretion disc evolution. They track the system out to 400 ms, by which point about half of the wind has been launched. Fernández et al.

(2019, hereafter F19) followed the evolution of a post-merger disc system to 9 s, allowing all of the disc material to be accreted or ejected. Miller et al. (2019) evolve the wind to 130 ms with concurrent GR Monte Carlo neutrino transport, and with a full nuclear reaction network. They find that the resulting outflow primarily consists of light r-process elements, consistent with a blue disc wind.

R-process heating provides around 1–3 MeV per nucleon, most of which is deposited within 1 s. If this were completely converted to kinetic energy, it would boost a particle at rest to a velocity of 0.1–0.2 c. Prior work has found that r-process heating can have a particularly strong effect on the tidal component of the dynamical ejecta, causing it to inflate (Rosswog et al., 2014; Grossman et al., 2014; Fernández et al., 2015; Darbha et al., 2021). R-process heating has not yet been included in a 3D GRMHD accretion disc simulation, so the details of its effects on kilonova winds remains uncertain.

The very long-term evolution of disc winds is also uncertain. Due to the computational expense, most disc wind simulations are only evolved to $\lesssim$ 10 s, while the wind needs to expand for several times that in order to reach homology. Longer-term evolution of hydrodynamic winds has been studied by extracting wind properties at a large radius, and using that to set an inner boundary condition on a larger grid (e.g. Kasen et al., 2015; Kawaguchi et al., 2021). In Chapter 4, I apply this technique to the 3D GRMHD simulation of F19, while including the dynamical effects of r-process heating.

Jet-Ejecta Interaction

Long-term observations of the radio afterglow of GW170817 indicate that the merger launched a successful jet (Mooley et al., 2018b). Such a jet has, by definition, tunneled through the surrounding ejecta. The jet propagation will undoubtedly affect the structure of the slower radioactive ejecta ($v \sim 0.1\,c$), potentially leaving an imprint on the kilonova itself. Even if all neutron star mergers launch jets, engine energy, duration, and opening angle may all vary as well, contributing to the diversity of kilonova light curves.

Studies of jet effects on kilonova light curves have generally focused on the effect of shock heating on the distribution of thermal energy. Numerical simulations show that the shock-heated cocoon surrounding a jet may dominate the light curve in the first ~hour after merger (Kasliwal et al., 2017; Gottlieb et al., 2018). Piro & Kollmeier (2018) argue that the blue component of the kilonova emission could be entirely due to jet shock heating, as opposed to the radioactive decay of r-process elements.

The disruption of the ejecta by the jet can also affect the light curves. The dynamical ejecta may include a thin outer shell of high-opacity Lanthanides (Kasen et al., 2015). Nativi et al. (2021) find that a jet can disrupt this layer, brightening polar emission by allowing light from the inner disc wind to escape. The focus of Chapter 3 is on assessing the effects of jet propagation on kilonova emission, both due to heating and changes to the ejecta structure.

1.3 Dissertation Overview

The primary focus of this dissertation is understanding the effects of jet-ejecta interaction and dynamical r-process heating on neutron star merger ejecta. In particular, I am interested in the effects on the resulting kilonova emission. My primary tool has been SEDONA, a multi-dimensional, multi-frequency time-dependent Monte Carlo radiation transport code (Kasen et al., 2006; Roth & Kasen, 2015). Throughout this work, I have improved its parallel performance and added important features such as checkpointing and a new particle-based output mode. In Chapter 2, I discuss the approach SEDONA takes to the radiation transport problem, as well as its parallelization strategy. I present the results of shared- and distributed-memory parallel scaling tests, as well as tests of checkpointing and the particle-based output. Sample test scripts can be found in Appendix A.

Chapter 3 investigates the effects of jet-ejecta interaction on the viewing-angle dependence of kilonova light curves. I post-process representative hydrodynamic simulations from Duffell et al. (2018) with SEDONA, simulating the light curves and spectra. The shock heating from the jet interacting with the surrounding ejecta has no effect on the emission after $\sim 1.5\,\mathrm{h}$. The jet can evacuate a narrow cavity within the bulk of the ejecta, producing brighter and bluer light curves near the poles. The enhancement could be visible above the synchrotron GRB afterglow for slightly-off-axis observers.

Chapter 4 examines the long-term evolution of a 3D GRMHD disc wind simulation. The 3D model of Fernández et al. (2019) is used as the basis for the inner boundary condition of a 2D hydrodynamics simulation. I investigate the effect that r-process heating has on the dynamical evolution of the wind, approximately accounting for heating that occurs before the wind enters our domain. Heating boosts the mass-weighted median velocity of the wind from $0.06\,c$ to $0.12\,c$. SEDONA radiation transport simulations show that the dynamical r-process heating changes the shape of the light curves and also causes them to be brighter and bluer on $\sim 1\,\mathrm{d}$ time-scales.

In Chapter 5, I summarize my results and discuss potential interesting directions for observations and future work.

The first project I worked on in graduate school was on red giant asteroseismology. It is sufficiently unrelated to my work on kilonovae that I have not included it in the main text of this dissertation. I have, however, included it in Appendix B. In it, I present a diagnostic for the location of differential rotation in red giants. Observations have shown that solar-mass red giant cores spin more slowly than predicted by conventional models of angular momentum transport. Convection and/or instabilities caused by large-scale magnetic fields may play a role. A measurement of the location of differential rotation could be used to constrain the primary angular momentum transport mechanism. Red giant rotation profiles are difficult to measure through asteroseismic means alone. I find that by combining independent surface rotation measurements with a few asteroseismic measurements, it is possible to constrain the rough location of the differential rotation.

Chapter 2

Monte Carlo Radiation Transport with Sedona

This dissertation is primarily concerned with connecting hydrodynamic simulations of astronomical transients to their ultraviolet, optical, and infrared counterparts. Doing so requires solving the equations of radiation transport. I use SEDONA, a multi-frequency, multidimensional Monte Carlo radiation transport code to do so. Throughout my PhD, I have made performance and usability improvements to SEDONA, enabling its use for problems that require more Monte Carlo particles and/or larger spatial grids. In this chapter, I summarize SEDONA's approach to the radiation transport calculation (Section 2.1). I discuss how the code is parallelized and study its scaling behavior in Section 2.2. In Section 2.3, I present the improvements I have made, including OpenMP threading, particle-based output, and "dead" zones.

2.1   Sedona

SEDONA is a multi-frequency, multi-dimensional Monte Carlo (MC) radiation transport code written in C++ (Kasen et al., 2006; Roth & Kasen, 2015). Its primary application is the calculation of light curves and spectra from radioactively-powered astronomical transients such as supernovae and kilonovae. It solves the time-dependent equation of radiation transport, representing the spatial- and frequency-dependent radiation field with an ensemble of particles, or photon packets, that have a 3D position $\vec{x}$, energy E, and frequency ν. Each particle represents a set of $E/h\nu$ photons, where h is the Planck constant. Particles also have an associated time t, which represents where they are within a time step. Many (often $\sim 10^7$–10^9) packets are needed to produce a reliable estimate of the radiation field.

The ejecta properties are specified on a grid, with options for a 1D spherical, 2D cylindrical, or 3D Cartesian structure. An initial model provides a spatially-varying density, temperature, and composition at a time t_0. In the configuration of SEDONA used here, the ejecta are assumed to be in homology, that is that $r = vt_{\exp}$, where r is the radial coordinate,

v is the local velocity, and t_{exp} is the time the ejecta has been expanding. This is a good approximation for the hydrodynamics of the ejected material on the time-scales relevant for emission. The density ρ therefore evolves as

$$\rho(t) = \rho(t_0) \left(\frac{t_0}{t}\right)^3. \tag{2.1}$$

Frequency-dependent quantities such as opacities and emissivities are tabulated in $N_{freq} \approx 10^2\text{--}10^3$ frequency bins for each zone and time step. There are typically ~ 100 zones in each grid dimension, giving a total of $N_{zone} \approx 10^{2D}$ zones in a D-dimensional grid.

The system is evolved in a series of discrete time steps, during which ejecta properties are held constant. The SEDONA simulations of Chapters 3 and 4 consist of $\sim 10^2\text{--}10^3$ logarithmically spaced time steps between 300–900 s and ~ 10 d, and another $\sim 10^2\text{--}10^3$ evenly spaced time steps of a few hours until ~ 100 d.

Each time step consists of the following:

1. **Opacity:** The gas state is solved in each zone, including including ionization fractions and level populations. Then the absorptive and scattering opacities are computed for each zone and frequency bin.

2. **Transport:** Particles are created to represent radioactive heating. For each particle, including those that existed previously:

   a) Its frequency-dependent mean free path is calculated. This is used to sample the distance the particle will travel before it scatters or is absorbed, d_{sc}. The distance to the boundary in the direction of the particle's motion, d_{bn} is calculated, as is the distance the particle will travel before it Doppler shifts to a different frequency bin, d_{ds}. The code also calculates the distance the particle will travel in the time remaining in the time step, $d_t = (t_{end} - t)/c$, where t_{end} is the end of the time step, and t is the current particle time.

   b) The particle is propagated a distance $d = \min(d_{sc}, d_{bn}, d_{ds}, d_t)$ and its time incremented by d/c. If $d = d_{sc}$, the appropriate absorption or scattering event is calculated in the fluid frame. This captures adiabatic losses as well as relativistic effects.

   c) These steps are repeated until $t = t_{end}$, or until the particle is absorbed or escapes the grid. Particles that escape are tallied according to their observer-frame frequency, direction of motion, and escape time.

   Throughout the propagation steps, tallies are kept in each zone to calculate the radiative heating rate and mean (frequency-dependent) radiation field intensity J_ν.

3. **Temperature:** The temperature of each zone is calculated by equating the rate of thermal emission to the rate of radioactive heating plus the rate of photon absorption. The density is also updated according to equation 2.1.

2.2 Parallelism

Approach

Modern supercomputers and clusters are made up of multiple connected nodes. For example, NERSC's Cori Haswell partition consists of > 2000 nodes that have two CPUs, each with 16 physical cores. Each node has its own RAM (memory) that is accessible by either CPU. There is also a high-bandwidth network that connects the nodes to each other and to a file system.

In order to take advantage of Cori and similar machines, SEDONA must be able to use all of the cores on a single node and be able to distribute its work among several nodes. To do so, the code is parllelized using both MPI and OpenMP. OpenMP enables threading and shared-memory parallelism, while MPI allows a single program to have multiple instances ("ranks") with separate memory that communicate with one another via a network.

SEDONA's parallelization strategy is to decompose the problem by particle; each MPI rank maintains a copy of the spatial grid, and propagates a subset of the particles, which are never communicated. Grid and output spectrum information is communicated as needed. The parallelization and communication structure over N MPI ranks is:

1. **Opacity:** Each rank calculates the opacities for $1/N$ of the grid zones, and the work is divided between the threads. The results are communicated to all ranks with an MPI collective. This requires the communication of $\mathcal{O}(N_{\mathrm{zone}}N_{\mathrm{freq}})$ values.

2. **Transport:** Particle generation and propagation happen independently on every rank. The propagation step is parallelized with OpenMP. The radiation energy density, absorbed energy, and often J_ν are reduced via an MPI collective. If J_ν is being stored, then this step requires the communication of $\mathcal{O}(N_{\mathrm{zone}}N_{\mathrm{freq}})$ values. The other reduced radiation quantities are scalars, so only require $\mathcal{O}(N_{\mathrm{zone}})$ values. The communicated quantities are tallies, so the communication costs do not depend on the number of particles.

3. **Temperature:** Each rank calculates the temperature in $1/N$ of the grid cells, and this work is divided between the OpenMP threads. The result consists of $\mathcal{O}(N_{\mathrm{zone}})$ values and is communicated with an MPI collective.

Within each step above, grid- and particle-based calculations are trivially parallelizable. Essentially, each rank runs its own particle propagation step, and then the results are averaged. The spatio-frequency grids are expensive to communicate, given the $\mathcal{O}(N_{\mathrm{zone}}N_{\mathrm{freq}})$ size. By contrast, the temperature communication tends to be cheaper, since there is only one temperature per zone.

From a programming standpoint, the primary advantage of this approach to parallelism is the simplicity of the communication pattern. The main alternative would be to decompose the domain, such that each rank would operate on a portion of the grid. In that case, there would be no need to communicate grid data such as the radiation field and opacity.

However, particles would need to be communicated between ranks during the propagation step. In our current model, all of the communication is limited to MPI collectives in the grid calculation step. This approach to parallelization works well for problems where the primary cost is in particle propagation, since it is the step that does not require any communication. Most of Sedona's use cases are currently in this regime. As we move to larger spatial and/or frequency grids, the communication costs will increase, possibly making a domain-decomposed approach more advantageous. Such an approach could more easily hide particle communication costs by overlapping inter-rank particle transfers with continued propagation of other particles.

The other advantage to our current approach is in load balancing between ranks. Particles are preferentially created in regions with high energy density, which in a radioactivity-powered system, have high mass densities and high optical depths. Consequently, particles are not evenly distributed on the grid. Since each rank operates on the whole domain, particles are, on average, evenly distributed between ranks. The simplest grid-decomposition scheme would give each rank an equal number of zones, which would lead to a highly unbalanced distribution of particles between ranks. Effective grid decomposition would therefore require some kind of adaptive scheme that allows for blocks of zones to be divided if too many particles accumulate.

Scaling

To assess the parallel performance of Sedona, I use a modified version of the 2D `toy_type1a_supernova` test problem. The problem calculates the light curve and spectra of a toy Type Ia supernova model. The model has eight atomic species, and a 100-by-200, regularly-spaced cylindrical grid. For the timing test, I run for 10 time steps, adding 10^6 particles at every step (10^5 per step in the OpenMP-only test). This test is representative of a typical Sedona calculation performed in Chapters 3 and 4: millions of Monte Carlo particles propagating on a $\sim 10^4$-zone 2D cylindrical grid. In this problem, $N_{\mathrm{zone}} = 2 \times 10^4$ and $N_{\mathrm{freq}} = 2400$. In the particle propagation and radiation communication steps, we track the frequency-dependent J_ν rather than just the total radiation energy. While this is not necessary in our test problem, it is the current default option, and is needed for calculations outside of local thermodynamic equilibrium, as well as for the mode where the opacity is updated concurrently with the temperature.

In this section, I use the April 2021 development version of Sedona (commit `7f94e3b`). The scaling tests were performed on the Haswell partition of Cori at NERSC. Thread affinity is set so that each OpenMP thread is bound to a different physical core. I use the Intel compilers (version 19) and the environment variables `OMP_PLACES=threads` and `OMP_PROC_BIND=true`. I do not use any flags in the `srun` command. The number of ranks is specified via the `--tasks-per-node` and `--nodes` flags to `sbatch`. See appendix A for an example script. Note that if using the GNU compiler, different OpenMP settings are necessary to obtain the same thread layout.

The timings shown are for the calculation and communication steps outlined in Sections 2.1 and 2.2. These are measured using `MPI_Wtime` on the master rank. Ideally, we would time every rank to identify load imbalances and get average timings. However, the infrastructure to do so does not currently exist in SEDONA. Due to the parallelization structure, we do not expect load imbalances to be significant for this problem. With sufficient particles, each rank will have a very similar particle propagation time. The opacity calculation should be similarly well balanced because regardless of composition or location in ρ–T space, the same calculations are performed for every zone in a simulation. There are options for which opacities to include, but they are simulation-wide, rather than zone-specific. The same applies for the temperature calculation. In some problems, load imbalances may develop if the non-linear solves require vastly different numbers of iterations, especially in a case where the opacity is updated alongside the temperature. More sophisticated timing infrastructure would be a valuable tool for performance analysis.

The temperature calculation is not timed independently. The times shown here are calculated by subtracting the time for temperature communication from the time for the full temperature step (calculation and communication).

OpenMP

Particle propagation, opacity calculation, and the temperature calculation are all embarrassingly parallel and threaded with OpenMP. If a calculation takes time T_1 to execute on 1 thread, it will ideally take

$$T_N = \frac{T_1}{N} \tag{2.2}$$

when run using N threads. We test this for the 2D toy Ia problem. Fig. 2.1 shows strong scaling behavior with OpenMP, running the same problem on different numbers of threads. Particle transport is the most expensive part of the calculation, exceeding the cost of opacity and temperature by over 50 per cent. Because this calculation is on one MPI rank, communication costs are (essentially) zero. Threaded calculations dominate the run-time, so the total time for each step shows good strong scaling behavior, even though it includes times for serial regions of the code.

The scaling falls slightly short of ideal behavior, with an average wallclock time per step of 232 s on one thread, as compared to 11.1 s on 32 threads. Fitting the times to a power law finds that $T_N \propto N^{-0.86}$. Perfect strong scaling behavior would put an average time step at 7.3 s.

Performance does not appear suffer when the threads are distributed across the two CPUs on the node. Up to 16 threads, all of the threads are on a single CPU. For 32 threads, we can no longer fit them on a single CPU if each thread is allocated a physical core. Naively, we might expect a performance penalty associated with a single process controlling multiple CPUs.

Increasing to 64 threads shows slight gains in performance, to 9.4 s from 11.1 s on 32 threads. The CPUs used have Hyper-threading, which allows the CPU to switch back and
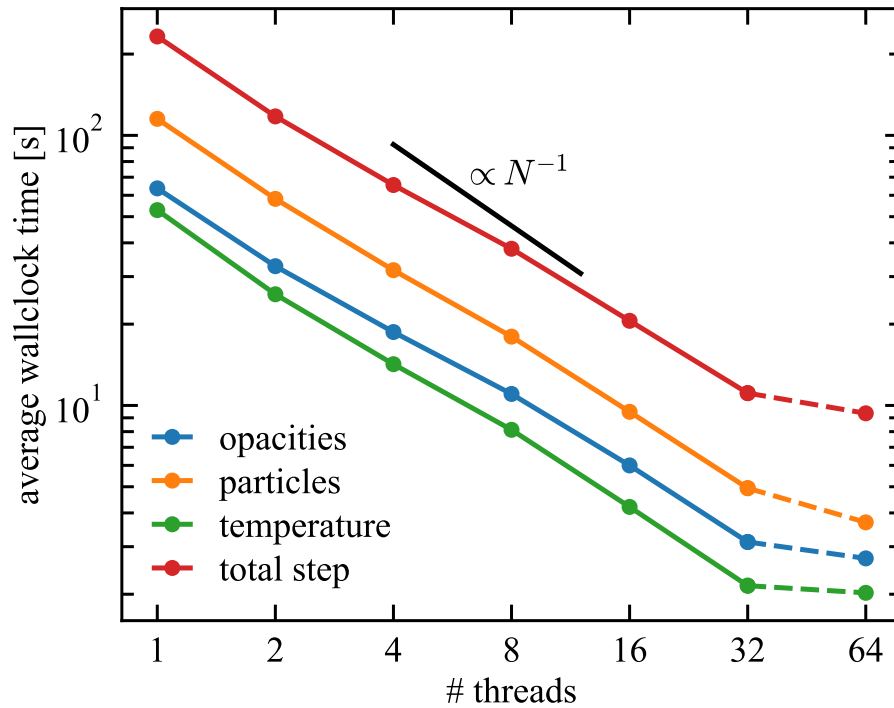
Figure 2.1: One-rank strong scaling behavior with OpenMP. Average wallclock time for main three portions of SEDONA calculation: opacity calculation (blue), particle transport (orange), and temperature calculation (green) are shown as a function of the number of threads used. The red line shows the total time spent on a time step, including parts of the calculation not plotted here. The black line shows the $\propto N^{-1}$ ideal strong scaling behavior, which the actual timings match reasonably well up to 32 threads. Up to 16 threads, all of the threads are on a single CPU. Up to 32, each thread is allocated its own physical core. At 64 threads, each is allocated to its own Hyper-thread, two of which are on each physical core. While this provides some speed-up due to the more efficient use of idle time, the gains are modest.
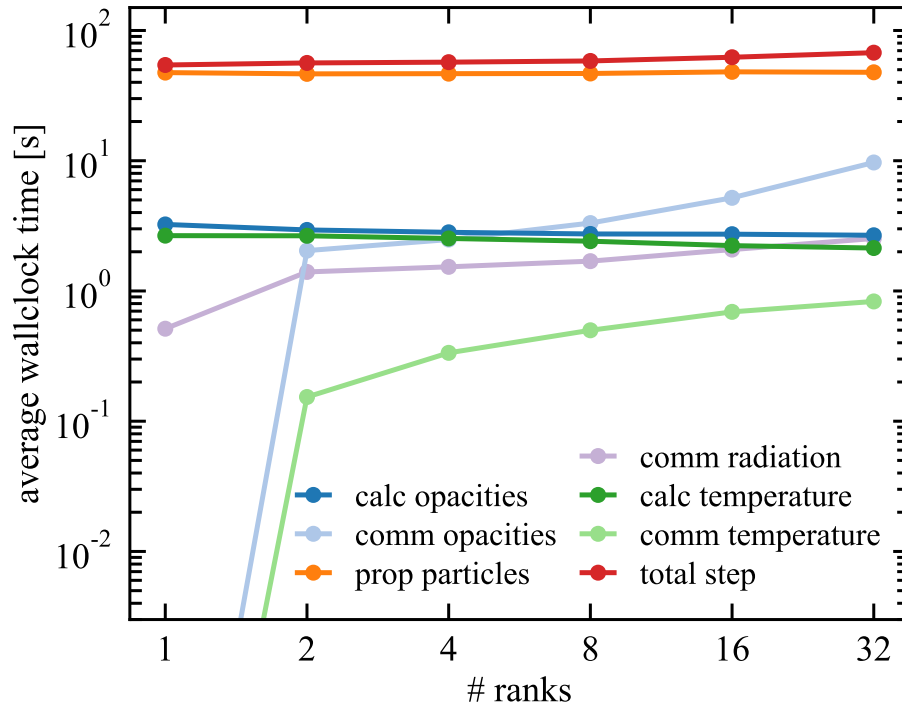
Figure 2.2: Average times per step for different hybrid MPI+OpenMP configurations on a single node. Each run has 32 threads total, evenly divided between the ranks. For example, the configuration with 4 MPI ranks has 8 threads per rank. Total step time is shown in red. Timings for the calculation-heavy portions of the time step are shown in darker colors, and communication-heavy steps are shown in light colors. Due to the large number of particles, particle propagation (in orange) dominates the run time. Opacity (blue) and temperature (green) calculation have comparable run times. The calculation-heavy steps have approximately constant run time as the number of ranks increases, though the communication time increases for all components (opacity: blue, radiation: purple, and temperature: green). Consequently, performance is best when there are as few ranks as possible per node.

forth between two threads while waiting for non-calculation steps such as cache reads or file system access. The small gains suggest that even without Hyper-threading, latency is not a substantial issue in this configuration of SEDONA.

This has implications for the optimal rank-thread configuration within a node. In order to fully utilize a node, we need to have 32 threads total, but they can be divided between any number of MPI ranks. The results of testing these configurations are shown in Fig. 2.2. The code runs fastest when running on a single rank and 32 threads. Steps that are computationally intensive have run times that are mostly insensitive to the number of ranks per node. While particle propagation is always the most expensive step, opacity and radiation communication both exceed opacity and temperature calculation in the configuration with 32 ranks and one thread per rank. The time to communicate temperature also increases, but remains subdominant, since there are $\mathcal{O}(N_{\mathrm{zone}})$ values to communicate. By contrast, the other two communication steps involve $\mathcal{O}(N_{\mathrm{zone}}N_{\mathrm{freq}})$ values.

The calculation steps are slightly faster on two ranks than on one (totaling 52.2 s versus 53.4 s), but the difference is small enough that the communication required in the two-rank run hides any speedup to the FLOP-heavy steps. A full time step takes 54.3 s on one rank and 56.3 s on two ranks. This trend may not hold for other problems. At two ranks the opacity communication has a run time comparable both to the temperature and opacity calculations. In a problem with a smaller spatial grid, the communication may be cheap enough that the two-rank configuration will be faster. Subsequent tests are performed using two ranks per node and 16 threads per rank; the choice of configuration is not expected to affect run time by more than a few per cent.

MPI

One of the biggest challenges for running SEDONA on large numbers of nodes (and consequently, large numbers of ranks) is the cost of communication. Strong scaling results for MPI up to 64 nodes (and 128 ranks) are in Fig. 2.3. The individual components of the calculation scale relatively well. The total run time per step plateaus due to the high costs of communication. At lower rank counts, particle propagation is the most expensive step, and is overtaken by opacity and radiation communication at 16 nodes (32 ranks). The run time for particle propagation nearly follows a $\propto N^{-1}$ relation, even on large communicators, so the code should scale well to more nodes for runs with more particles or fewer zones.

The opacity calculation scales similarly to particle propagation, though for this problem, the opacity communication quickly becomes many times more expensive than the calculation. Since the opacity calculation itself is split between the ranks, it rapidly falls off in run time. However, the communication cost increases with communicator size, doubling between one node and 64 nodes. The total time spent on opacity is 5.0 s on one node, decreases to a minimum of 3.4 s at 8 nodes, and increases again to 4.1 s by 64 nodes. There appears to be a sweet spot that compromises between calculation and communication costs, which for this problem is at 1250 zones per rank. A potential avenue for speeding the opacity step would be to split the ranks into multiple communicators such that regardless of the total number
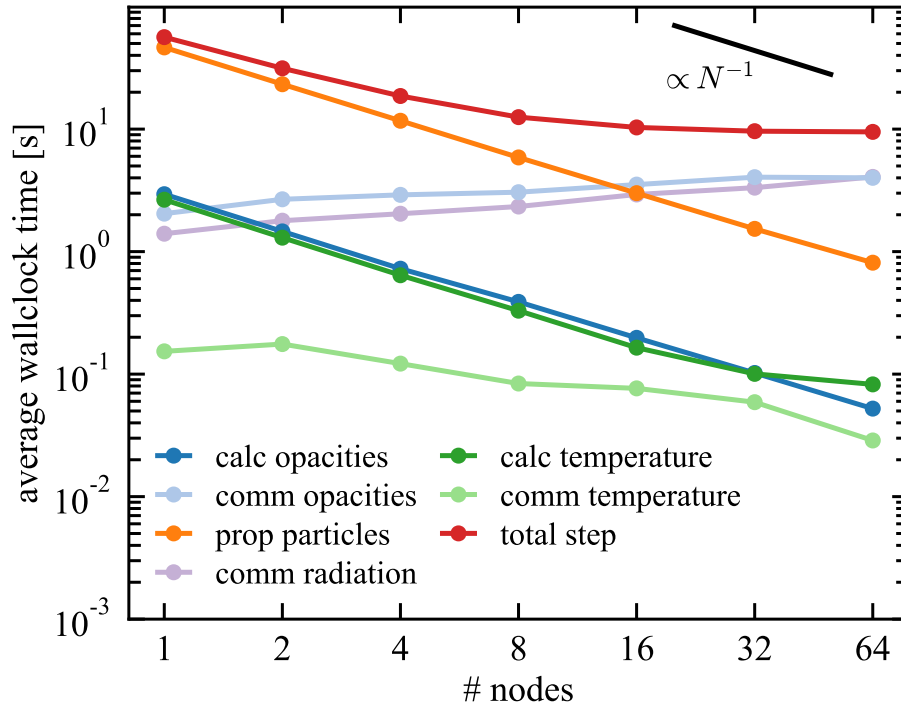
Figure 2.3: Strong scaling behavior for MPI parallelism. Average wallclock time per step is plotted as a function of the number of nodes (2 ranks per node). Ideal $t \propto N^{-1}$ behavior is shown in black. Total time per step is shown in red. Communication time is shown in lighter colors, and calculation steps are shown in darker ones. Communication cost for opacities (light blue) and radiation (light purple) increases with the number of nodes since the size of the MPI communicator also increases. Calculation-heavy steps nearly match the ideal behavior, so the code will scale well as long as particle propagation dominates the run time.

of ranks, each rank calculates opacities for approximately the same number of zones. If the 64-node run had the same opacity performance as the 8-node run, the total run time would decrease by 15 per cent.

The cost of communicating the radiation field is similar to that of the opacities, due to the similarly-sized amounts of data. Running without storing J_ν would somewhat reduce the total run time, since it would bring the radiation communication cost in line with the temperature communication. That said, opacity communication still handles $\mathcal{O}(N_{\text{zones}} N_{\text{freq}})$ values, so the performance improvements would be slight, rather than order-of-magnitude.

The temperature calculation follows a similar trend to the particle propagation and opacity calculation, approximately falling off as $\propto N^{-1}$. Starting at 32 nodes, the speed up is somewhat less than expected. At that node count, the temperature calculation is parallelized over 1024 OpenMP threads, with each thread doing the calculation for only 20 zones. A
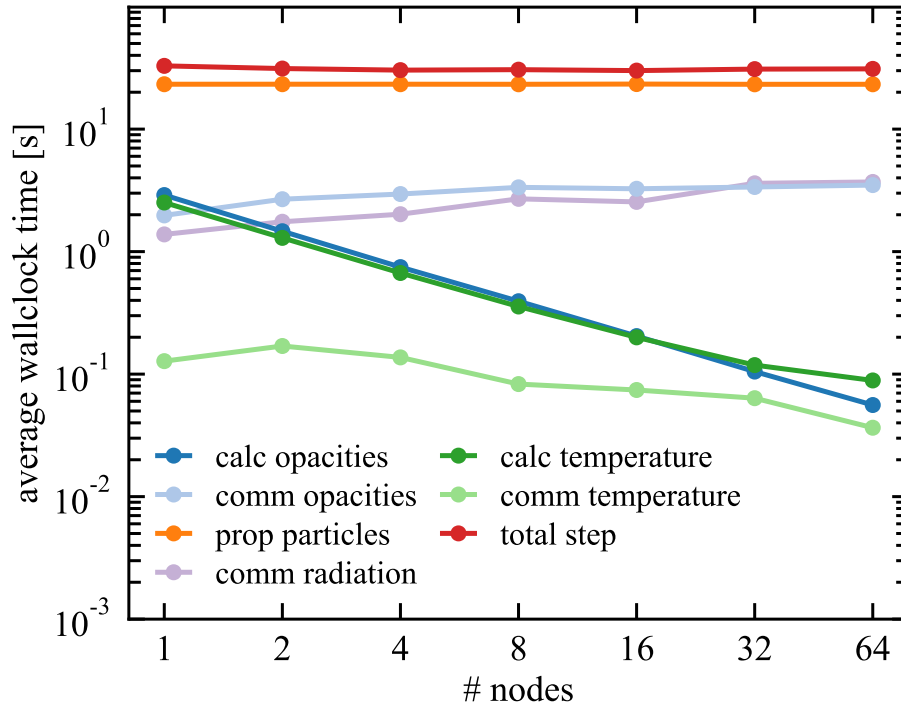
Figure 2.4: Weak scaling test, where the number of particles per node remains fixed at 5×10^5 as the number of nodes increases. Particle propagation time (orange) and the total time per step (red) stay approximately constant, which indicates good weak scaling behavior. The timings for the other components are the same as in the strong scaling tests, since those costs are independent of the number of particles.

temperature calculation for a single zone is much cheaper than calculating the opacity for all of the frequency bins within a zone. The small amount of work per thread may cause us to reach a point of diminishing returns for parallelizing the temperature calculation.

The communication cost for the temperature remains small, since there are only $\mathcal{O}(N_{\mathrm{zones}})$ to communicate. The decreasing cost with rank count is unexpected. It is possible that the smaller amount of communicated data allows the MPI implementation to make optimizations that are not possible with the opacity or radiation communication. The temperature communication is such a small portion of the total time that understanding its scaling behavior is not important.

The scaling behavior for particle propagation continues to be good even as the number of particles increases. Fig. 2.4 shows the results of a weak scaling test, where the number of particles per rank is held fixed, and the number of ranks increases. We use 5×10^5 particles per node, for up to a total of 3.2×10^7 particles on 64 nodes. The total cost of particle propagation remains mostly unchanged as the number of particles and ranks

increases. Because this is the most expensive step in this test, the total run time per step is also nearly constant. The run times for the other components do not depend on the number of particles, so the times for the other steps are the same as in the MPI strong scaling test.

GPUs

The fraction of supercomputing power from graphics processing units (GPUs) is increasing. Many of the current-generation machines and nearly all of the next-generation facilities have at least one GPU per node. Fully utilizing a node will require leveraging the GPU(s) in addition to the CPU cores. Individual components of the SEDONA calculation may be able to be efficiently offloaded to a GPU. For instance, the non-LTE calculation requires a large sparse matrix solve, which GPUs can do more effectively than CPUs.

Moving the transport step to GPUs may be a challenge. Effective use of GPUs requires that the work be divided into chunks that will execute the same instructions in lock-step. The number of times a particle interacts with the fluid (e.g. scatters, is absorbed, reaches a grid boundary) within a time step is highly variable. There is also frequent branching within a single thread. Moving particle propagation to GPUs while maintaining the current algorithm, where a full time step of particle propagation is handled by a single thread, would likely perform poorly. Effectively implementing transport on GPUs may require moving to an event-based algorithm, which have been shown to perform well on GPUs in the context of neutron transport (Hamilton & Evans, 2019). In these algorithms, the unit of work assigned to a thread is a single iteration of particle movement or interaction. This tends to perform better than the history-based approach that SEDONA currently uses because there is much less branching within a single unit of work.

2.3 Improvements to Sedona

I have made a number of performance and usability improvements to SEDONA, allowing for more efficient simulations and fewer wasted node-hours.

OpenMP Threading

SEDONA was initially only parallelized with MPI. In order to fully utilize each node, there was one MPI rank per core, leading to e.g. 32 ranks per node as in the right-most points of Fig. 2.2. As long as the spatial grid is small, this approach performs fairly well. Each rank must maintain its own copies of several spatio-frequency grids, which can be too large to fit in memory for multi-dimensional problems. For our test problem above, the zone data take up $> 3\,\text{GB}$ per rank, for a total of $\sim 100\,\text{GB}$ of RAM occupied by 32 ranks. Cori's Haswell partition has $128\,\text{GB}$ per node, which is only barely able to meet the memory needs of the application in this configuration. The utilization fraction would be much lower if this were a 1D simulation and with a factor of 100 fewer zones.

In our 2D test problem, performance with 32 ranks on a single node was slightly worse than with 32 threads and a single rank. This is due to the increasing cost of opacity and radiation communication with the number of ranks. With a smaller spatial grid, these communication steps would likely be substantially cheaper, making a pure-MPI approach a good choice.

However, moving to larger grids such as those of the test problem above, and those used in the subsequent chapters of this dissertation, a pure MPI approach begins to suffer. Firstly, with as few as 25 per cent additional spatial or frequency zones, the grid data alone would use all of the RAM on the node. Shared-memory parallelism is necessary to make full use of the node's CPUs while using large spatial and/or frequency grids. Particle propagation was threaded before I started using SEDONA, but the opacity and temperature calculations were not. These steps became a substantial bottleneck when running 2D simulations with non-grey opacities for Chapter 3. I added OpenMP threading to these steps, which sped them up considerably.

Each zone's calculation is independent, so parallelizing by zone was straightforward. The main concern was race conditions. The opacity and temperature calculations both fill an instance of the `GasState` class with the zone properties, and then make calls to `GasState` methods to solve for the opacity and temperature. Previously, the `transport` class had a single `GasState` instance that was reused for every zone calculation. Since threads would share this instance, this would have been a source of race conditions in the code. To resolve this, the `transport` class now has a vector of N_{thread} `GasState` classes. Each thread has its own instance, so race conditions are avoided. `GasState` only contains information for a single zone, so the memory cost of having multiple instances is minimal. Each instance also has a pointer to the class that holds the atomic data. There is only one `AtomicData` instance per rank, which limits its memory footprint for atoms with large numbers of lines. The atomic data is read in at the start and is not updated throughout the simulation, so there are not race condition concerns.

Checkpoints

In order to take advantage of modern supercomputers, SEDONA needs to be able to checkpoint and restart a simulation. Machines generally have wallclock limits for jobs, so the simulation needs to be able to exit gracefully before reaching the time limit, causing the scheduler to cancel the run. Prior to the implementation of checkpoints, it was not uncommon to lose hundreds of node-hours of progress when a simulation took longer than anticipated.

Checkpoint files are HDF5 (HDF Group, 2021) archives that consist of the following data:

- metadata from the SEDONA binary that generated the checkpoint file, such as the compile date-time and the git version

- random number generator states for every OpenMP thread on all MPI ranks
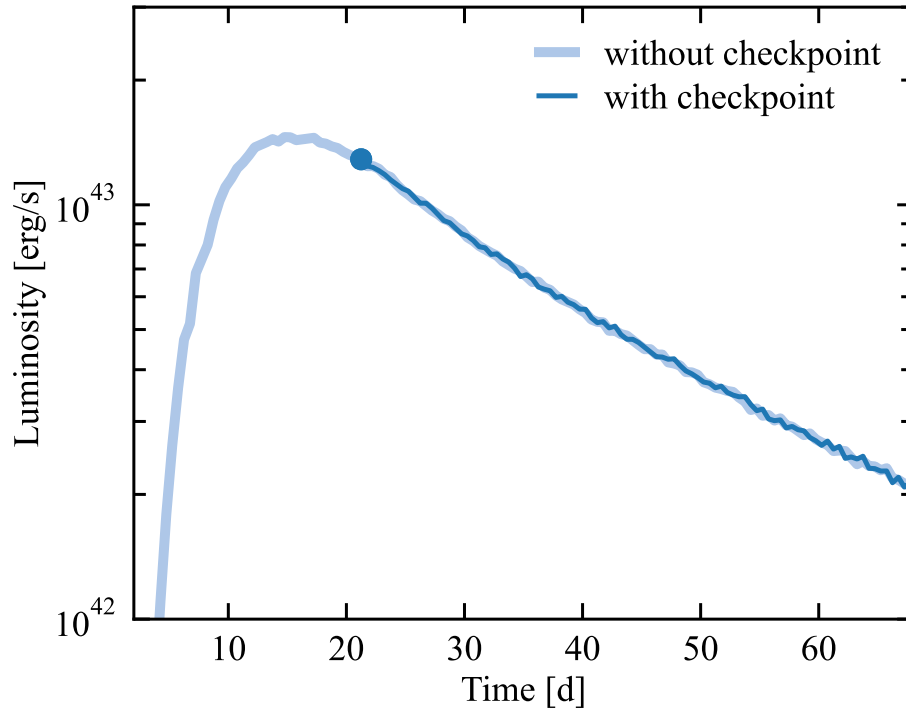
Figure 2.5: Results of checkpoint test with the Lucy supernova test problem. The light blue shows the result of the simulation run normally to its end. The dark blue shows the light curve resulting from the simulation restarted at around 20 days. The light curves match each other within Monte Carlo noise. While the random number generators were checkpointed, the results do not match exactly because the particles are not assigned to threads deterministically, and each thread has its own random number generator.

- output spectrum for optical photon packets and gamma rays

- grid structure

- zone properties (e.g. local velocity, composition, temperature, density)

- particle properties (e.g. position, energy, frequency, type)

The inclusion of the grid structure and zone information means that the model file used to initialize the simulation is unnecessary for a restart. Run-time parameters are specified on restart in the same way as they are at the start of a new run.

Checkpointing is tested in the nightly test suite. The `lucy_supernova` test is stopped and restarted partway through, and the result is compared to the result without checkpointing. The test is done in 1D, 2D, and 3D. The results of the 1D test are shown in Fig. 2.5. There is good agreement between the light curves calculated with and without checkpointing; any

differences are within the Monte Carlo uncertainties. On checkpoint, the state of the random number generators is saved out, and on restart they are re-initialized with the appropriate seed and state. This is sufficient to be able to produce the exact same result with or without a restart. However, each OpenMP thread has its own random number generator, and particles are not statically assigned to threads. From a performance standpoint, this is necessary since the time taken to propagate a single particle has a high variance. Consequently, the particle-thread parings are non-deterministic, and results will vary slightly. If we were running only with MPI, then the output files produced with and without the restart would be identical. Maintaining the same random number generator states across restarts improves the quality of the random numbers since there is not a risk of unnecessarily repeating portions of the random number history.

Particle-Based Output

When particles escape the grid, their energy is tallied in a multi-dimensional histogram with respect to frequency, observed time, and viewing angle. I created an optional output mode that creates HDF5 archives that list the particles that have escaped from the grid. When a particle escapes, it is removed from the vector of active particles and added to a vector of escaped particles. When particles scatter, their position is recorded, so we have the position of last scatter for escaped particles. Whenever the code prints output files, it also writes out a file with a list of all of the particles that have escaped thus far. The escaped particle list is cleared, and propagation continues. The periodic clearing is because the full list of escaped particles can reach dozens of gigabytes, so holding them in memory is impractical.

I wrote an accompanying C++ program on top of the SEDONA code base that can apply various filters to the escaped particles and construct output spectrum files from the reduced particle list. This can be used to determine which particles are observed by particular observers within different time intervals. We can also do the reverse, determining the contributions of particular regions of the ejecta to the light curve.

The left panel of Fig. 2.6 shows the energy-weighted locations of last scatter for all of the escaped particles in the dynamical r-process SEDONA calculation of Chapter 4. The most common last-scatter locations are in the bulk of the disc wind, at $v \sim 0.1\,c$. The right panel shows the same but only for those particles seen by a polar observer before four hours. The bulk of the emission at those angles and at that time is from the high-velocity tail of the ejecta.

Dead Zones

In multi-dimensional grids, coordinates are constrained by the requirement that there can be no point on the grid with a velocity greater than the speed of light. In the homologous version of SEDONA, the velocity vector of a grid cell is fixed at $\vec{v} = \vec{x}/t_{\mathrm{exp}}$, regardless of whether there is any material at that point. Therefore, a 2D cylindrical velocity grid that extends to $0.8\,c$ in the x- and z-directions would have cells with velocities up to $1.13\,c$. This
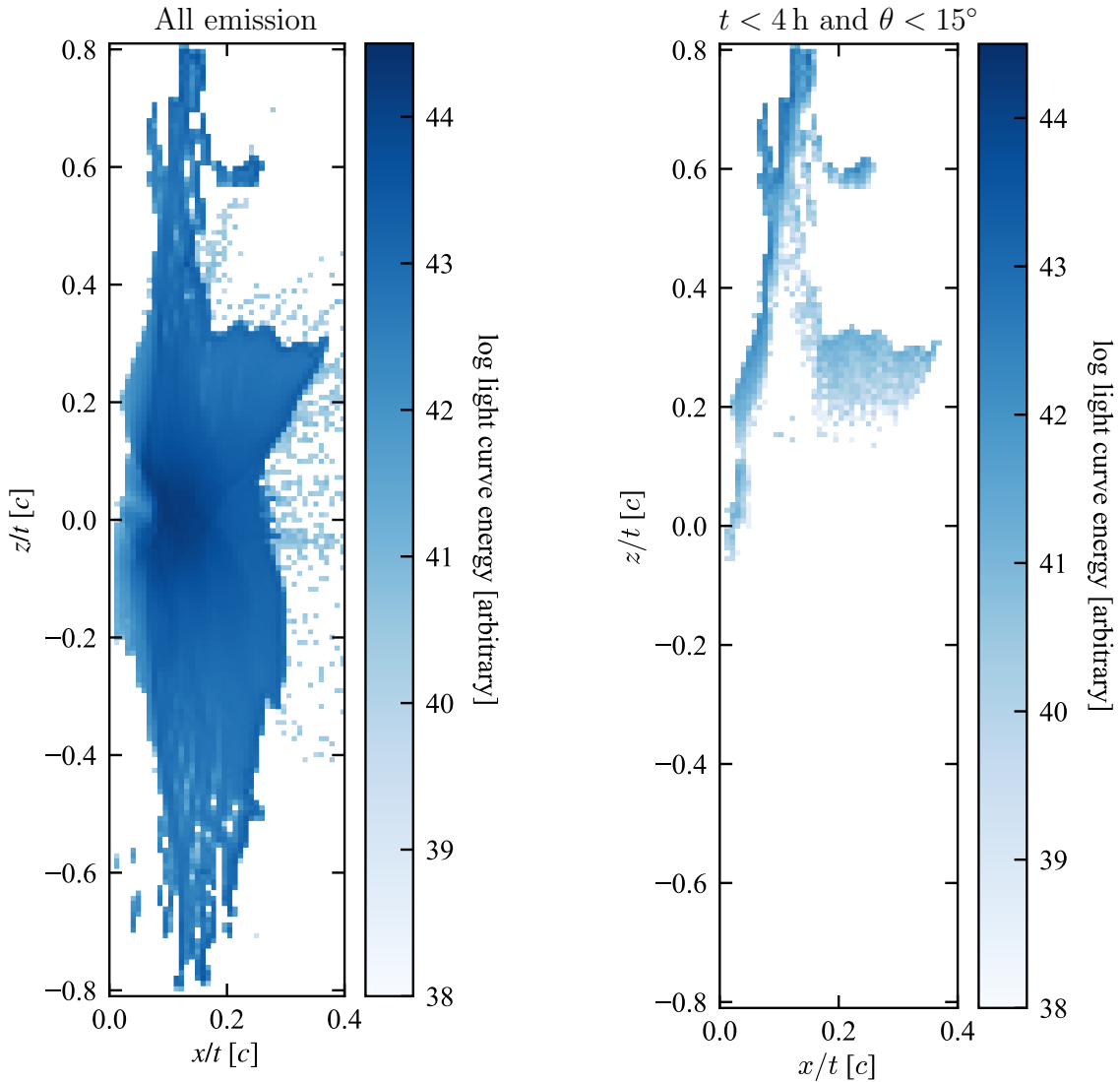
Figure 2.6: Example of results using particle-based output. On the left is a map of energy-weighted locations of last scatter for all of the escaped particles in the SEDONA simulation of the dynamical r-process model of Chapter 4. On the right is the same, but filtered to only show the particles seen by observers at the North pole in the first four hours. The locations are binned according to their coordinates in (homologous) velocity space. Most of the early emission is from the wall of the jet cavity, especially at high velocity.

causes problems in the calculation of both opacities and Lorentz transforms, even if the density is zero in those zones. This is also an issue in 3D Cartesian coordinates, but is avoided in the 1D spherical grid.

To extend grids to higher velocities, I added a "dead zone" feature to SEDONA. Zones have an additional flag that denotes whether or not they are dead; this is specified in the input model. Once a particle enters a dead zone, it is marked as escaped, and is tallied and removed from the simulation, just as if it had actually left the grid. Opacities are also not calculated in dead zones, and they should have densities of zero. Since there is no mass or energy density, the code will not generate any new radioactive particles in those zones, so it is as if the dead regions of the grid are not in the simulation in the first place. This has allowed me to extend the grids in Chapter 4 to 0.8–$0.9\,c$ in both directions, and will allow for the simulation of more multi-D, highly-relativistic transients.

2.4   Summary

Radiation transport has a number of applications, particularly in the domain of transient astrophysics. SEDONA is well-suited for calculating the thermal structure of and emission from supernovae and kilonovae. Accurate results require millions of Monte Carlo particles, which makes the simulations computationally expensive. Parallel methods are necessary to efficiently run on modern supercomputers. SEDONA takes a hybrid OpenMP+MPI approach that takes advantage of both shared- and distributed-memory parallelism. Parallel performance is good, especially in problems where the computational cost is dominated by particle transport. Communication costs are significant in multi-dimensional grids, and suggest that a domain-decomposed approach may be more suitable for some problems.

I have made performance and usability improvements to SEDONA, which have enabled the simulations performed in the chapters that follow. Without the threaded opacity calculation, the non-grey simulations of Chapter 3 would have been too expensive and memory-intensive. I have also added the ability to checkpoint and restart a simulation, which is critical for long simulations, and has made simulations more efficient. Inactive zones allow for larger spatial and velocity grids in multi-D simulations, which allowed for the high-velocity tail of the disc wind to be included in Chapter 4.

Chapter 3

The Effect of Jet-Ejecta Interaction on the Viewing Angle Dependence of Kilonova Light Curves

This chapter is adapted from Klion et al. (2021b) with permission from the co-authors.

The simulations in this chapter are based on a suite of 2D relativistic hydrodynamic simulations of a jet interacting with a homologously expanding outflow (Duffell et al., 2018). In these models, jet energy and opening angle vary, but it is assumed that there is no delay between the launch of the wind and the start of the jet. In this chapter, I select a representative sample of these models and calculate viewing-angle dependent kilonova light curves using SEDONA (Section 3.1). The results of these simulations are presented and discussed in Section 3.2. By 1 h after merger, the thermal energy deposited by the jet is negligible relative to the energy generated by r-process decays. The jet, however, substantially alters the density distribution of the ejecta near the pole. The channel carved by the jet can lead to brighter and bluer emission along the axis of the jet. I discuss observational implications in Section 3.3.

3.1   Radiation Transport Simulations

Initial Models

In Duffell et al. (2018) we studied the dynamics of a jet interacting with a homologously expanding outflow, varying jet energy and opening angle. Our calculations were in the limit where the delay between the launch of the outflow and the start of the jet engine is very small (10^{-4} s). This is much shorter than both the engine duration (0.1 s) and the time taken for the jet to break out of the ejecta (3×10^{-3} s in the fastest case). We performed 2D axisymmetric relativistic hydrodynamic simulations using JET, a moving mesh hydrodynamics code (Duffell & MacFadyen, 2011, 2013). We identified the conditions under
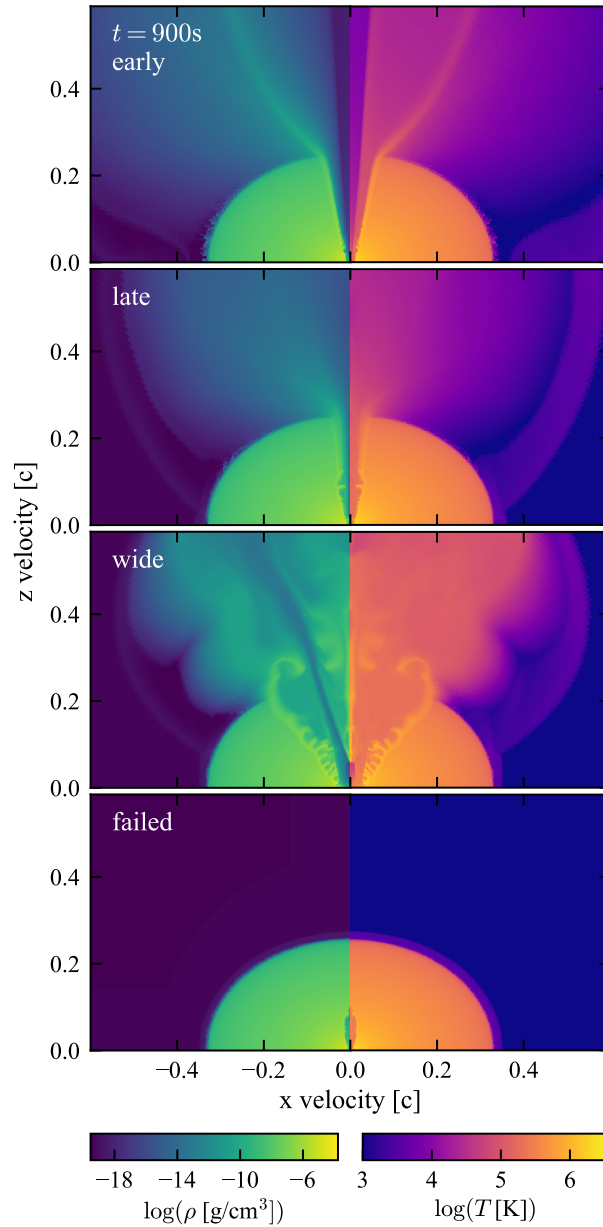
Figure 3.1: Starting conditions for our simulations, at $900\,\mathrm{s}$ ($15\,\mathrm{min}$) after homologous expansion and jet launching begin. The left and right panels show the logarithms of the mass density ρ and temperature T, respectively. Each set of panels corresponds to a different type of outcome from the jet-ejecta interaction, as identified in Duffell et al. (2018) and described in Table 3.1.

Table 3.1: Jet injection energies (E_j) and half-opening angles (θ_j) for the Duffell et al. (2018) models that we study in this chapter. The short names will be used to refer to these models throughout the chapter. The models are scaled so the total mass is $4 \times 10^{-2}\,\mathrm{M}_\odot$, and the jet is active for $0.1\,\mathrm{s}$. The jet energy is given both in ergs and as a fraction of the ejecta kinetic energy $E_\mathrm{ej} = 5.7 \times 10^{50}\,\mathrm{erg}$.

Model	Short Name	E_j [erg]	$E_\mathrm{j}/E_\mathrm{ej}$	θ_j [rad]
early jet breakout	early	3×10^{49}	5×10^{-2}	0.1
late jet breakout	late	5×10^{47}	8×10^{-4}	0.1
failed jet, breakout	wide	1×10^{50}	3×10^{-1}	0.4
failed jet, no breakout	failed	1×10^{47}	2×10^{-4}	0.1

which jets can break out of the surrounding ejecta, as well as scaling relations for the time of breakout and amount of energy thermalized. The initial structure of the ejecta follows numerical results of dynamical ejecta from neutron star merger simulations (Hotokezaka et al., 2013a; Nagakura et al., 2014) and is given in equations 15 through 19 of Duffell et al. (2018). This component is expected to make up a small portion ($\sim 10^{-3}\,\mathrm{M}_\odot$) of the total ejecta. However, for the sake of simplicity we assume that all of the ejected mass ($> 10^{-2}\,\mathrm{M}_\odot$) follows this structure. This is a crude approximation of NS merger ejecta, but our focus here is on the qualitative viewing-angle effects of a jet interacting with a radioactive outflow.

Depending on the opening angle and jet energy relative to ejecta mass, we identify four qualitatively different outcomes that can arise from a jet interacting with a homologously expanding outflow:

- early jet breakout (hereafter "early"): jet breaks out before the central engine turns off.

- late jet breakout ("late"): jet stays collimated but breaks out after the central engine turns off.

- failed jet, successful breakout ("wide"): energetic jet that fails due to its wide opening angle. The large amount of deposited energy still leads to a shock breakout, though not of a jet.

- failed jet, no breakout ("failed"): jet completely fails and there is no breakout.

In this chapter, we investigate the effect of the central engine on the kilonova light curves in each of these four cases. We choose one model from each category of outcome. Model parameters are shown in Table 3.1. We scale the hydrodynamic calculations such that the jet duration is $0.1\,\mathrm{s}$ and the ejecta mass is $0.04\,\mathrm{M}_\odot$. This corresponds to an ejecta kinetic energy $E_\mathrm{ej} = 5.7 \times 10^{50}\,\mathrm{erg}$.

The models from Duffell et al. (2018) have been evolved in JET for $t_\mathrm{J} \equiv 100\,\mathrm{s}$, until they are approximately homologous (i.e. $v \propto r$). After the JET calculation, we transform the

models onto an axisymmetric velocity grid, and exclude all material with $v > 0.6\,c$. There is a negligible amount of mass at these high velocities, so it will not affect the light curves on time-scales of an hour or more. The sole source of thermal energy in the JET models is the jet shocking and heating the surrounding ejecta. We denote the total of this thermal energy $E_{\rm sh}$, and the corresponding energy density $\varepsilon_{\rm sh}$. At $t_{\rm J}$, the optical depth of the material is very high, so starting the Monte Carlo radiation transport calculation at that time would be computationally infeasible. Fortunately, doing so is unnecessary. Assuming homology and adiabatic expansion, we can calculate the mass density (ρ) and temperature (T) structure of the ejecta when the SEDONA calculation starts at $t_{\rm S} \equiv 900\,{\rm s}$. We can relate ρ and $\varepsilon_{\rm sh}$ at $t_{\rm J}$ and $t_{\rm S}$ by

$$\rho(t_{\rm S}, v_r, \theta) = \rho(t_{\rm J}, v_r, \theta) \left(\frac{t_{\rm J}}{t_{\rm S}} \right)^3 \qquad (3.1)$$

and

$$\varepsilon_{\rm sh}(t_{\rm S}, v_r, \theta) = \varepsilon_{\rm sh}(t_{\rm J}, v_r, \theta) \left(\frac{t_{\rm J}}{t_{\rm S}} \right)^4, \qquad (3.2)$$

respectively, where v_r is the radial velocity coordinate, and θ is the latitude. Equation 3.2 accounts for the t^{-1} decline in total thermal energy due to adiabatic expansion.

The total thermal energy density at a point at $t_{\rm S}$ is the sum of the components from shock heating (equation 3.2) and from radioactive heating. The JET calculations do not account for the radioactive heating, so we add it in when constructing the SEDONA models. For a radioactive heating rate per unit mass $\eta(t)$, a fluid element of mass m will have a thermal energy due to radioactive heating, $e_{\rm rad}$ that evolves according to

$$\frac{{\rm d}e_{\rm rad}}{{\rm d}t} = \frac{-e_{\rm rad}}{t} + \eta(t)m. \qquad (3.3)$$

Therefore, the energy density due to r-process heating at a given point in space will be

$$\varepsilon_{\rm rad}(t_{\rm S}, v_r, \theta) = \frac{\rho(t_{\rm S}, v_r, \theta)}{t} \int_{t_0}^{t_{\rm S}} \eta(t) t \, {\rm d}t. \qquad (3.4)$$

The total thermal energy density at a given point at t_S is the sum of the shock and r-process contributions:

$$\varepsilon(t_{\rm S}, v_r, \theta) \equiv \varepsilon_{\rm sh}(t_{\rm S}, v_r, \theta) + \varepsilon_{\rm rad}(t_{\rm S}, v_r, \theta). \qquad (3.5)$$

We use equations 3.1 and 3.5 to evolve the hydrodynamics models from $100\,{\rm s}$ to $900\,{\rm s}$, when we begin the radiation transport calculation in SEDONA. During this phase, we adopt the r-process radioactive heating rate from fig. 3 of Metzger et al. (2010), assuming $Y_{\rm e} = 0.1$. Fig. 3.1 shows the mass density ρ and temperature T of our starting models at $15\,{\rm min}$. Note that for a given model, the density and temperature structures track each other closely. High mass density leads to a high radioactive heating rate. The total thermal energy from radioactivity greatly exceeds the thermal energy of shock heating by the jet, so the density structure dictates the temperature structure.
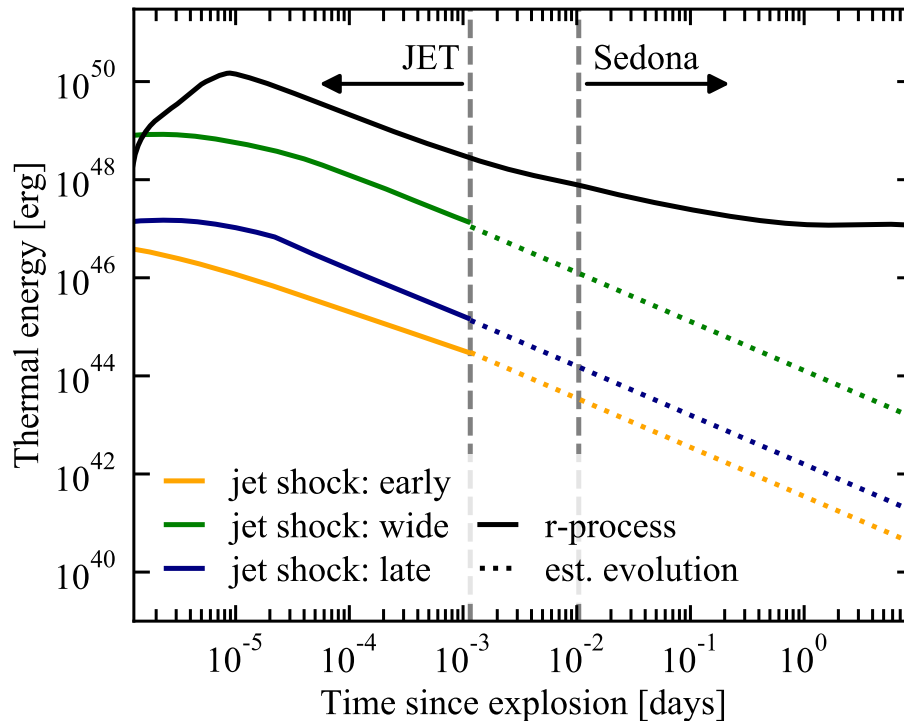
Figure 3.2: Estimated thermal energy of ejecta due to r-process heating (black, computed according to equation 3.4), and due to jet shock heating. Different colors correspond to our different models (orange: early, green: wide, blue: late). The dashed grey lines delimit the two phases of the calculation: the hydrodynamic simulations in JET (before $100\,\mathrm{s} \approx 10^{-3}\,\mathrm{d}$), and the SEDONA calculations that begin at $900\,\mathrm{s} \approx 10^{-2}\,\mathrm{d}$. The solid colored lines show the thermal energy in the JET simulations. The dotted lines show how the thermal energy would evolve while the ejecta are expanding adiabatically.

We quantitatively compare the thermal energy from jet shock heating and radioactivity in Fig. 3.2. We show the total thermal energy as a function of time during the JET simulations. Due to the short duration of the jet-ejecta interaction, the jet heating is completely over by $\sim 10^{-5}\,\mathrm{d} \approx 1\,\mathrm{s}$. Subsequently, the thermal energy decreases as $\propto t^{-1}$. Throughout, the energy from r-process heating exceeds that of the hottest jet-ejecta interaction model by nearly two orders of magnitude.

Fig. 3.3 shows the distribution of mass with radial velocity for the four models. Since we assume homologous expansion, the velocity distribution also corresponds to the spatial distribution of mass. The jet does not interact with the majority of the mass of the ejecta. The angle-averaged mass distribution (solid lines) of low-velocity ($\beta \lesssim 0.35$) material is very similar across all models. More energetic jets and those with wider opening angles impart more kinetic energy to the ejecta, leading to more high-velocity material.
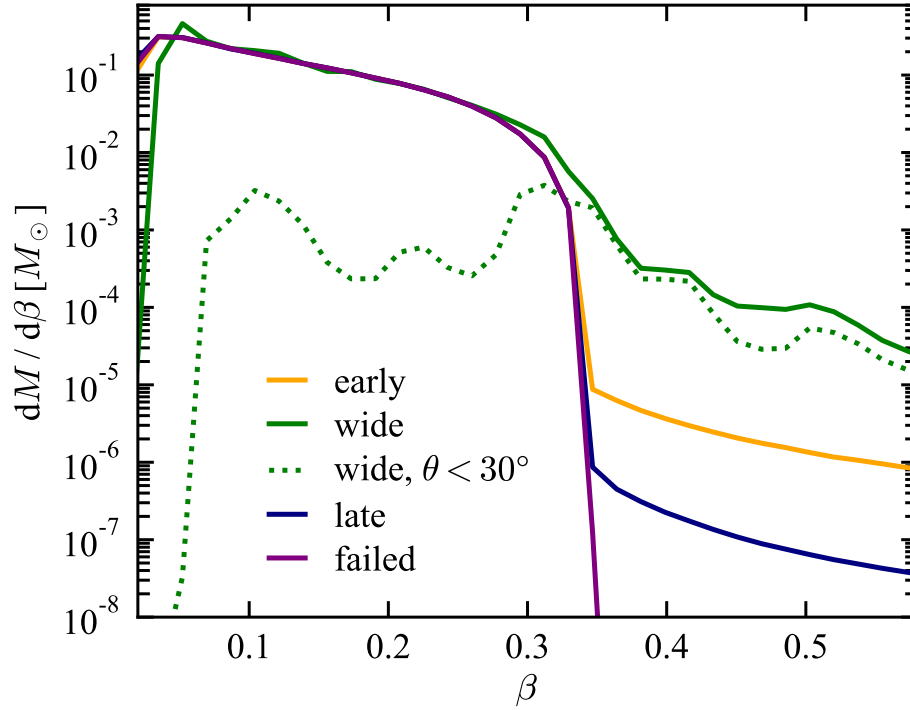
Figure 3.3: Distribution of ejecta mass $dM/d\beta$, as a function of velocity $\beta \equiv v/c$ for our ejecta models, denoted by different colors (orange: early; green: wide; blue: late; purple: failed). The dotted line only shows matter that lies within 30° of the pole. The solid lines show distributions for ejecta at all angles. In an angle-averaged sense, the primary differences between models is in the amount of high-velocity material. Near the pole, though, the jet can dramatically reduce the amount of low-velocity mass.

This high velocity material is confined to the polar regions. The dotted line shows the distribution of mass within 30° of the pole for the wide model. At $\beta \gtrsim 0.35$, the mass distributions for the full ejecta and polar region only are very similar. At those same latitudes, the amount of low-velocity material is reduced by orders of magnitude. The vast majority of the contribution to low-velocity material comes from equatorial latitudes. Our assumption of homologous expansion at fixed velocity (equation 3.1) neglects the fact that r-process heating per particle of 3 MeV can accelerate the gas up to $\sim 0.1\,c$. We may therefore somewhat underestimate the velocity of the low-speed part of the ejecta. By assuming homology, we also implicitly assume that hot ejecta will not expand into the cavity opened by the jet. In nature, magnetic fields may keep the channel evacuated. The fate of the jet cavity remains an open question. Assessing it would require long-term ($> 10\,\mathrm{s}$) magnetohydrodynamic simulations of neutron star merger ejecta.

Sedona

We use SEDONA, a time-dependent, multi-wavelength, multi-dimensional Monte Carlo radiation transport code (Kasen et al., 2006; Roth & Kasen, 2015), to determine the emission from the above models. The code tracks the emission of packets of radiant energy ("photons") with a given wavelength. SEDONA calculates their propagation through the ejecta, accounting for random absorption and scattering events in the moving background. All photon emission and scattering is done in the fluid frame. This accounts for relativistic effects such as Doppler shifting and beaming. It also ensures that the total energy of emitted photons is correct in the fluid frame. As the photons escape from the ejecta, they are tallied according to their escape time, wavelength, and propagation direction, giving time- and viewing-angle-dependent spectra and light curves. The ejecta are taken to expand homologously, such that the mass density in each cell falls off as t^{-3}. At each time step, the temperature of the cells is determined by equating the rate of thermal emission to the rate of radioactive heating plus photon absorption. Adiabatic losses emerge naturally from the scattering process. More details can be found in Chapter 2.

The heating due to radioactive r-process products is not dominated by a single nuclide. Instead, there is an ensemble of relevant nuclides with a wide distribution of half lives (Metzger et al., 2010). We do not track the decay and heating of all of these nuclides. We adopt the parametrized, time-dependent r-process heating rates of Lippuner & Roberts (2015), assuming an initial $Y_e = 0.13$, expansion time-scale $\tau = 0.84\,\mathrm{ms}$, and entropy $s = 32\,k_B$. We use the same parametrization throughout since the heating rate is largely independent of the microphysical parameters. Given the small impact of the jet on the thermal structure of the ejecta even on $\sim$second time-scales (Fig. 3.2), it is unlikely that the jet changes the composition of the bulk of the ejecta.

The r-process heating rate, η, is calculated using the time in the lab frame. In doing so, we neglect time dilation in the fluid frame. Our heating rate approximates the power law, $\eta(t) \approx At^{-1.3}$, so we underestimate the heating rate (energy deposited) by a factor of $\gamma^{1.3}$ ($\gamma^{0.3}$), where γ is the fluid Lorentz factor. Our fastest material has $\gamma = 1.25$, so the r-process luminosity we calculate is incorrect by at most 30 per cent. The effect on total energy deposited is even smaller, reaching only 5 per cent. The omission of time dilation in η will therefore have a small effect on our results.

Opacity

Unless otherwise stated, we assume all material has a grey opacity of $1\,\mathrm{cm}^2\mathrm{g}^{-1}$. This value is chosen to be roughly comparable to Planck mean line expansion opacities of first peak r-process elements (roughly, the second row of the d-block of the periodic table). Tanaka et al. (2020) find that at one day after ejection, these elements have mean opacities in the range of $\kappa \sim 10^{-2}$–$10\,\mathrm{cm}^2\mathrm{g}^{-1}$. By contrast, Lanthanide-rich ejecta are expected to have mean opacities an order of magnitude greater (Kasen et al., 2013; Tanaka et al., 2020). In this study, we focus on the relative effect of different jet parameters on the light curves at

varying viewing angles. As such, the exact choice of grey opacity does not affect our primary results.

To assess the possible impact of a more realistic, non-grey opacity, we re-run one of our models with temperature, density and frequency-dependent opacity. We include bound-bound, free-free, and electron scattering opacities. We use the line expansion formalism of Karp et al. (1977) to tabulate bound-bound opacities on our discrete frequency grid.

For this calculation, we use an isotropic composition of half calcium-90 and half iron-90. Ideally, we would instead use a mixture of first- or second-peak r-process products. However, this is not possible due to our early start time and therefore high initial temperature. At the start of our calculations, the hottest portions of the ejecta are at a temperature of 5×10^6 K and a density of $3 \times 10^{-5}\,\mathrm{g\,cm^{-3}}$. Using the Saha equation and ionization energies from Kramida et al. (2019) we can calculate the expected ionization state of a given element at that temperature and density. At that point in ρ–T space, we find that ruthenium (representative first-peak r-process product) is, on average, 41.7 times ionized, and that neodymium (representative second-peak/Lanthanide product) is 50.0 times ionized. Therefore, in order to calculate the opacity for a realistic composition, we would need atomic levels and lines for over 40 ions of each element included. These are not currently available. Opacities have been calculated out to temperatures of only $\sim 10^4$ K for the Lanthanides (Kasen et al., 2017; Fontes et al., 2020; Tanaka et al., 2020) and $\sim 10^5$ K for d-block elements (Banerjee et al., 2020).

We use calcium and iron because they have similar atomic structure to first peak r-process elements, and therefore similar opacity patterns, at least for atoms that are up to ten times ionized (Banerjee et al., 2020). We also have atomic level and line data for all ionization states of these elements. To make as complete a line list as possible, we combine the atomic data of the Chianti (Dere et al., 2019) and CMFGEN (Hillier & Lanz, 2001) atomic databases. We use a mean molecular weight of 90, which is typical of first-peak r-process products. When lines dominate the opacity, this choice more accurately reflects the actual number density of atoms in the ejecta. At the highest temperatures, though, this underestimates the opacity from electron scattering. The reason is that electron density from ionizing calcium and iron is artificially capped because they have fewer electrons than their counterparts one row below on the periodic table. The artificially lowered Thomson opacity will only affect the innermost parts of the ejecta at the earliest times. The photosphere is always at a low enough temperature (at the start, 2×10^5 K at a density of $1 \times 10^{-9}\,\mathrm{g\,cm^{-3}}$) that the degree of ionization of first and second row d-block elements are comparable.

3.2 Results

Radioactive Heating Dominates Over Jet Shock Heating

We simulate the viewing angle-dependent bolometric and broadband light curves for each of our four models. The bolometric light curves are shown in Fig. 3.4. Each panel shows a
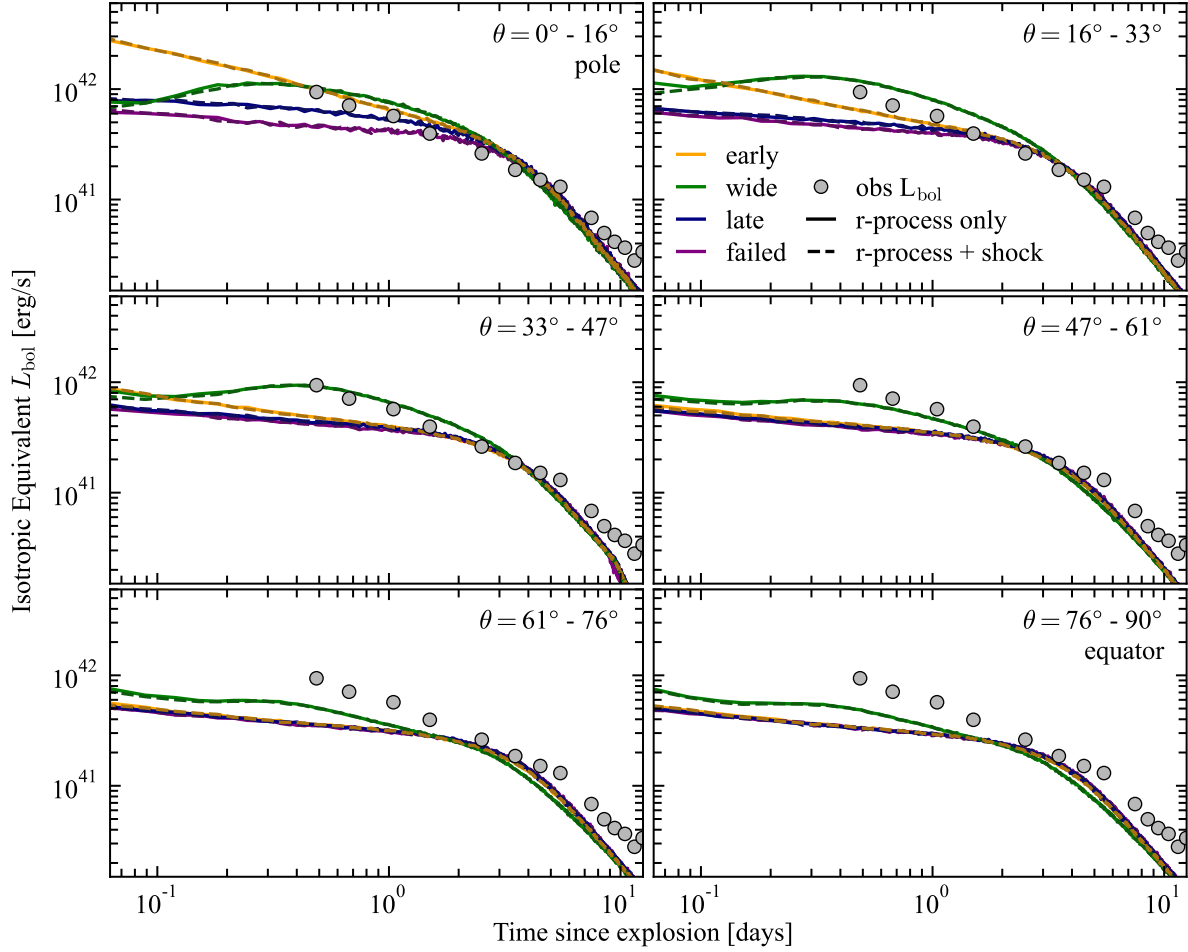
Figure 3.4: Isotropic equivalent bolometric luminosity at different viewing angles for different models. The bolometric luminosity of AT2017gfo (Drout et al., 2017) is overplotted in grey to guide the eye and demonstrate that our light curves are of approximately the correct luminosity and time scales. Each panel corresponds to a different viewing angle, arranged from polar to equatorial. Colors correspond to different jet-ejecta interaction cases, as described in Section 3.1 and as shown in Fig. 3.1 (orange: early; green: wide; blue: late; purple: failed). The light curves in solid lines consider only r-process heating, while the dashed lines also include jet shock heating. We find that the inclusion of jet shock heating does not appreciably affect the light curves. However, when the jet substantially alters the density structure of the ejecta, the polar light curves can be much brighter than on the equator. This is most pronounced in the early breakout case, where there is an energetic, collimated jet. By contrast, the model with a lower-energy jet and the same opening angle (late breakout) has a much less pronounced brightening on the pole.

different viewing angle, and the colors indicate the different models (orange: early; green: wide; blue: late; purple: failed). To guide the eye, the observed bolometric luminosity of AT2017gfo constructed in Drout et al. (2017) is overplotted in grey. We do not seek to exactly match the light curve of AT2017gfo, but do confirm that the time and energy scales in our models roughly match those of the event.

We find that the thermal energy imparted by the jet on the ejecta is negligible in comparison to the heating due to r-process decays. However, the kinetic energy of the jet is sufficient to change the density structure of the ejecta, leading to a viewing-angle variation in the radioactive transient.

The dashed lines in Fig. 3.4 show light curves calculated considering the thermal energy from both jet shock heating and r-process heating. The solid lines show curves from models that are identical to their dashed counterparts, other than that they include only r-process heating. The light curves with and without jet shock heating are very similar, even 1 h after merger. Since the jet shock heating in our models happens within the first several seconds of evolution, by 1 h the contributions have largely adiabatically degraded. Moreover, the thermal energy from shock heating is orders of magnitude less than that from r-process decays (Fig. 3.2).

This is true even in the "wide" case, which has the largest jet energy by a factor of three and the largest solid opening angle by a factor of 16. It also has the highest peak shock thermal energy, reaching $\sim 2 \times 10^{-2} E_{\mathrm{ej}}$. By contrast, the other models have peak thermal energies $< 10^{-3} E_{\mathrm{ej}}$ (Duffell et al., 2018). Even the large amount of thermal energy in the wide case is not sufficient to affect the light curve; radioactive heating dominates.

This demonstrates the difficulty of thermalizing sufficient energy for jet-ejecta shock heating to directly impact the kilonova light curve. We find that without a delay in jet onset, any effects of the jet on the light curve after one to two hours will be due to changes to the density profile of the ejecta. If the jet were delayed relative to the outflow, we would see more thermal energy from shock heating. Whether or not this would be sufficient to affect the light curve at one to two hours will depend on the length of the delay and is currently uncertain.

Jet Can Strongly Affect Density Structure

The jet has a minimal effect on the density or energy structure of the failed model. We use the failed model to study the light curves from our ejecta model in the absence of a jet, allowing us to isolate the effects of the jet-induced asymmetry from the inherent viewing angle dependence of the light curves from the ejecta. This is particularly important since the initial ejecta distribution is slightly oblate (axis ratio 1.3, (Duffell et al., 2018)). The polar light curves in the failed case are ~ 50 per cent brighter than those on the equator due to the slightly larger surface area seen by an observer at $0°$ (Darbha & Kasen, 2020).

The three other cases, however, show larger changes to the bolometric luminosity. The early and late cases show polar light curves that are brighter by factors of ~ 3–4 and ~ 2 respectively in the first ~ 12 h. After this, the light curves lose most of their viewing angle
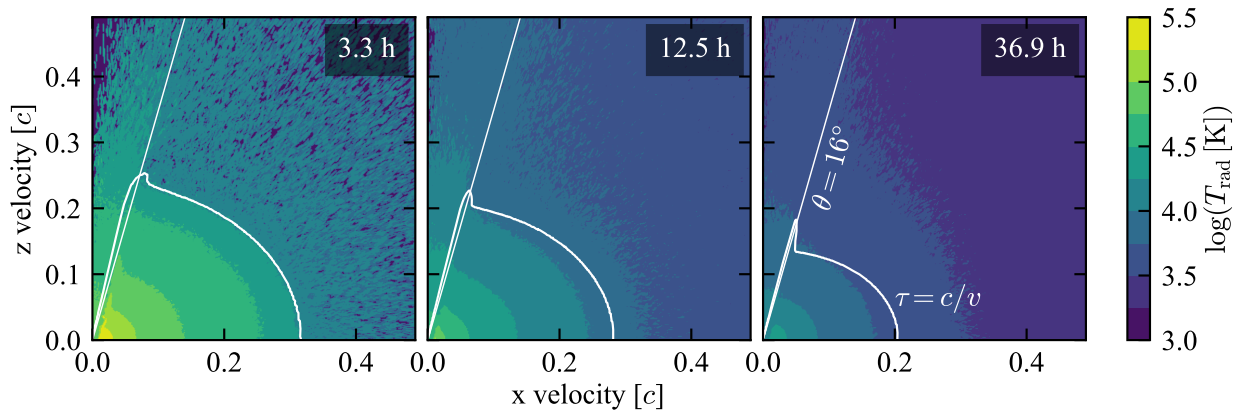
Figure 3.5: Slices of radiation temperature from radiation transport calculations of the early model at 3, 12, and 37 h after merger. The white contour marks where the radial optical depth $\tau = c/v$, where photons can efficiently diffuse out of the ejecta. This traces the density structure, showing the cavity carved by the jet. The hotter interior material becomes optically thin earlier than it would have in the absence of the jet. This leads to higher T_{rad} along the jet axis, giving brighter (Fig. 3.4) and bluer (Figs 3.6 and 3.7) emission on the poles than elsewhere. The diagonal white line marks $\theta = 16°$ from the pole, the maximum viewing angle included in the polar-most bin in our SEDONA results.

dependence. In these models, the jet has carved out a low-density tunnel along its axis. This cavity can be clearly seen in the density distributions plotted in Fig. 3.1. These low density regions have very little radioactive heating relative to the unperturbed regions of the ejecta that remain much more dense. The channel carved by the jet exposes the hot material that would otherwise remain optically thick until a few days after merger. The polar light curves are therefore brighter due to the hotter photosphere along the jet axis. While the jet cavity does slightly increase the projected surface area on the pole relative to the failed case, the effect is slight due to the small solid angle subtended by the jet. Instead, the primary cause of the increase in brightness is due to the higher photospheric temperature within the jet cavity.

This effect is most easily seen in temperature snapshots from the radiation transport simulations. Slices of the radiation temperature at 3.3, 12.5, and 39.6 h are shown Fig. 3.5. The white contour shows where the radial optical depth $\tau = c/v$. Temperatures at a given τ are higher at the poles than on the equator, and the ratio between polar and equatorial photospheric temperatures is greater at earlier times, dropping from 5 at 3.3 h to 3 at a day and a half. Throughout, there is a drop in temperature at around 15 degrees, corresponding to the angle where the jet-induced asymmetry ends. Temperatures are also higher slightly away from the jet cavity (at say $\sim 20°$) than on the equator.
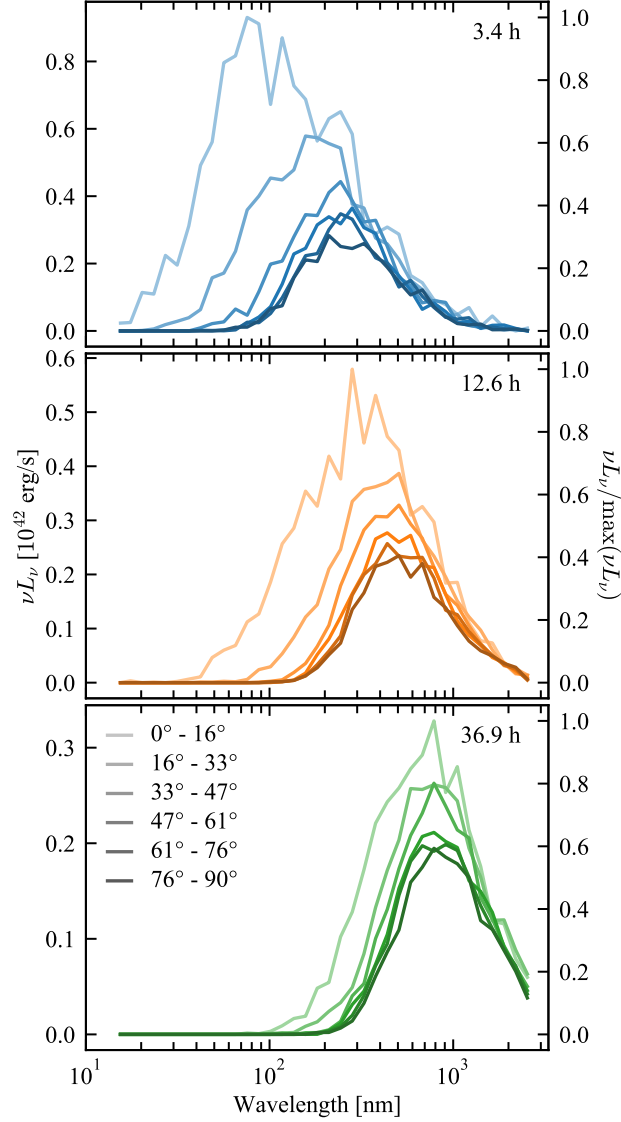
Figure 3.6: Viewing angle-dependent optical and UV spectra of the early breakout model. The times chosen approximately correspond to those shown in Fig. 3.5. The left y-axis shows absolute νL_ν, and the right axis is normalized to the maximum νL_ν across all spectra at that time. The viewing angles that are hottest in Fig. 3.5 are also substantially bluer in the spectra. At $\sim 3.4\,\mathrm{h}$, the polar spectrum peaks at $\sim 90\,\mathrm{nm}$, while the equatorial one peaks at $\sim 300\,\mathrm{nm}$. The jaggedness of the spectra is due to statistical noise and does not represent lines.

The hotter photosphere on the poles leads to emission that is both brighter and bluer. This is apparent in the viewing angle-dependent spectra shown in Fig. 3.6. For the first day of evolution, the spectral energy distribution (SED) peaks brighter and at shorter wavelengths on the pole than on the equator. At 3.4 h, the polar SED peaks at ~ 100 nm as compared to ~ 300 nm on the equator. Peak νL_ν is at around 1×10^{42} erg/s on the pole, as compared to 2×10^{41} erg/s on the equator. The polar spectra are wider than expected for a single-temperature black-body, indicating that the observer sees both the hotter polar region and cooler off-axis material. This highlights the need for multi-dimensional radiation transport simulations when predicting the light curves of non-spherical transients. At later times, the differences between polar and equatorial spectra become less pronounced, corresponding to the reduced temperature contrast between the two angles. Since the ejecta are also cooling, the spectra at all angles shift to redder wavelengths and are dimmer at later times.

Bluer spectra at earlier times near the poles correspond to brighter ultraviolet light curves at those times. Fig. 3.7 shows *Swift* UVOT W2-band (central $\nu = 193$ nm) and Cousins R-band (central $\nu = 635$ nm) light curves for our four models and compares them to observations (Kasliwal et al., 2017). The UV light curves show a similar trend to the bolometric luminosities. The early breakout model is brightest on the pole, and remains slightly brighter than the late and failed cases until about $45°$ from the pole, where all models other than wide become indistinguishable. Since the wide case has the largest impact on the density structure, it also has the greatest effect on the light curves. At all viewing angles other than the pole, the wide breakout case is brighter than the other models in the UV. By contrast, the effect of the jet cavity on the R-band light curves is much smaller. The higher temperatures on the pole enhance emission across the UV and optical spectrum, but the enhancement is much less on the Rayleigh-Jeans tail of the black body distribution, where flux is proportional to temperature, rather than near the peak where the scaling is stronger.

Comparison With Afterglow

Once the jet escapes the ejecta, it interacts with the surrounding interstellar medium, giving rise to synchrotron emission across the electromagnetic spectrum. Afterglows are brightest when viewed on-axis, and for energetic jets interacting with dense circumburst media (Sari et al., 1998; Van Eerten et al., 2010; Granot et al., 2018). We expect the optical afterglow to outshine the kilonova if the system is observed pole-on. Wu & MacFadyen (2019) fit the physical properties of the observed off-axis afterglow of GRB 170817A and use these parameters to predict what the event would have looked like on-axis. We convert their X-ray light curve to optical bands using the fitted spectral slope of the GRB 170817A afterglow, $L_\nu \propto \nu^{-0.6}$ (Margutti et al., 2018). This predicts a UV luminosity of -15 mag at one day, which is brighter than any of our on-axis kilonova models. This is consistent with predictions that, in general, an on-axis kilonova will be hidden from view by the GRB afterglow (Metzger & Berger, 2012). In the redder bands, the kilonova will stay brighter for longer; it is possible that the kilonova will exceed the afterglow several days after the burst, at which point the early-time effects of the jet-induced asymmetry on the kilonova will have dissipated.
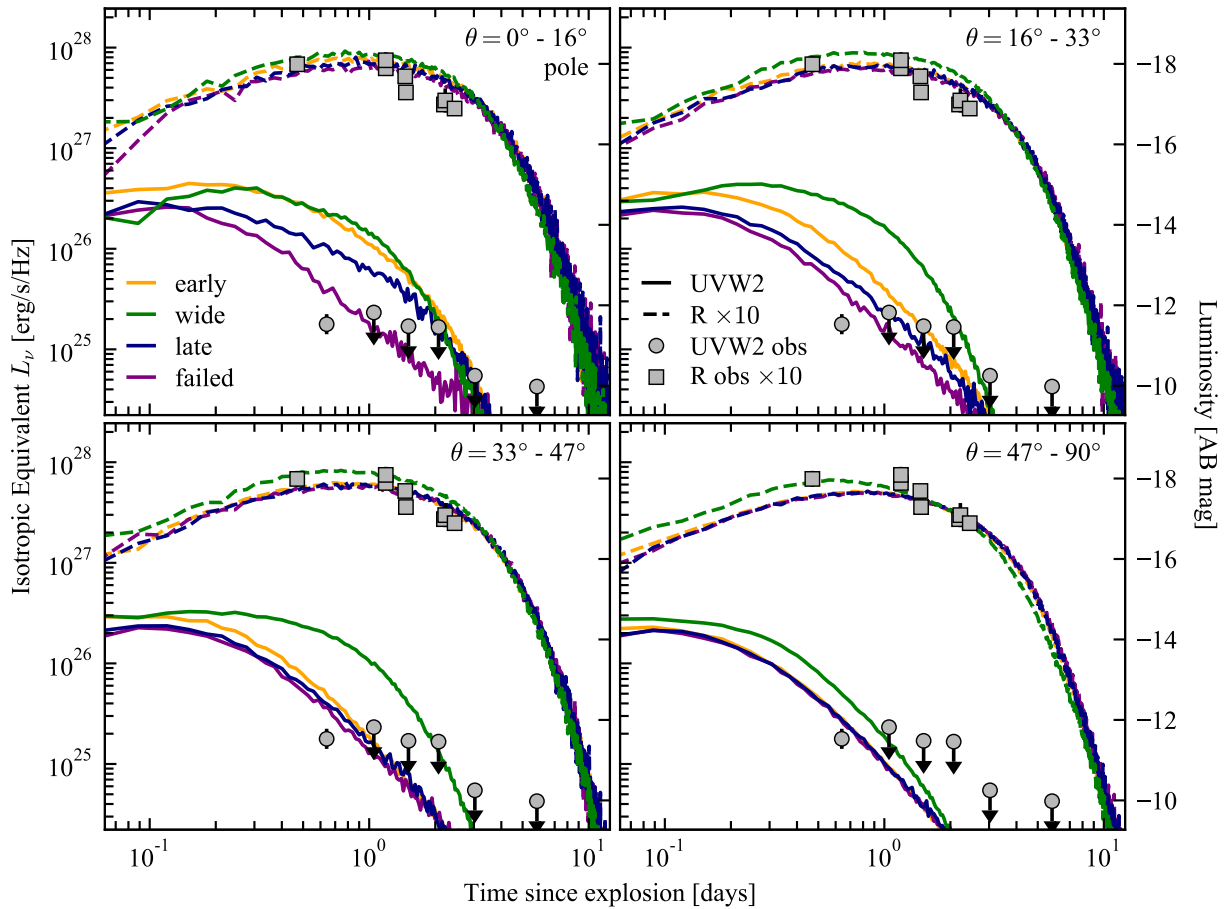
Figure 3.7: Isotropic equivalent luminosity in *Swift* UVOT W2 (solid line) and Cousins R (dashed line, ×10) bands at different viewing angles. These light curves are calculated in radiation transport calculations that account for both jet shock and radioactive heating of the ejecta. Different colors indicate different ejecta configurations, as described in Section 3.1 and shown in Fig. 3.1 (orange: early; green: wide; blue: late; purple: failed). Symbols show observed band luminosities of AT2017gfo (Kasliwal et al., 2017). Arrows indicate upper limits. Jet-ejecta interaction disrupts the density structure near the poles, leading to higher polar photospheric temperatures, particularly at early times. Thus, the UV light curves are more strongly affected by the jet-induced asymmetry than those in *R*-band.
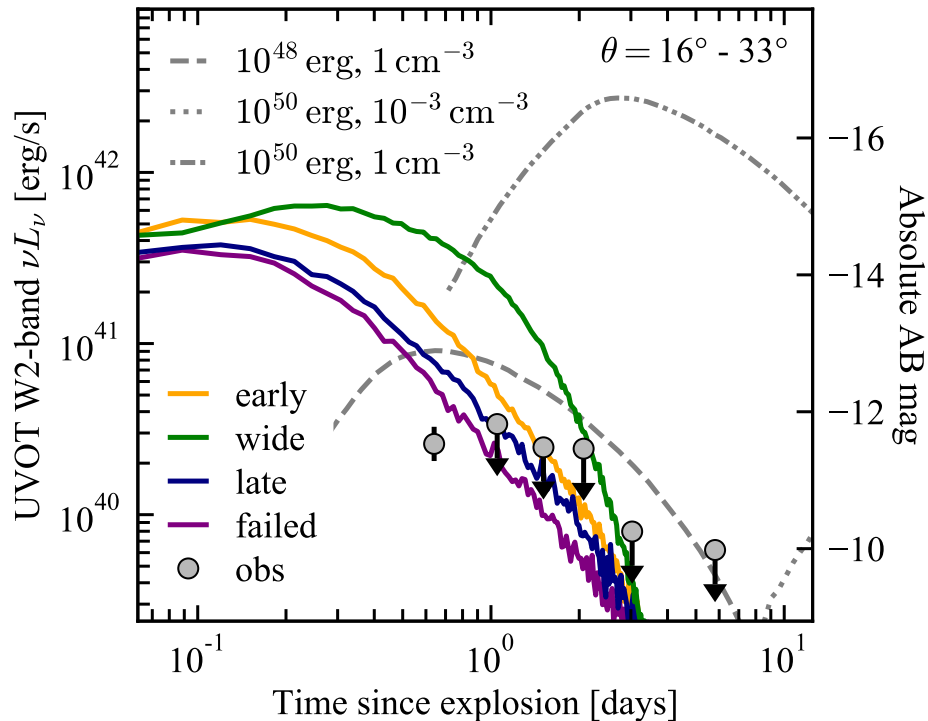
Figure 3.8: Isotropic equivalent luminosity compared to estimates of off-axis GRB optical afterglows. Colored lines show *Swift* UVOT W2 band light curves for different models just off-axis ($\theta = 16°$–$33°$). Off-axis afterglow models from Van Eerten & MacFadyen (2011) are overplotted in grey. We show estimates for combinations of different bipolar jet energies ($E \in \{10^{48}, 10^{50}\}$ erg) and ISM number densities ($n \in \{1, 10^{-3}\}$ cm^{-3}).

If observed off-axis, the kilonova may be brighter than the afterglow in UV bands for around one day. In Fig. 3.8 we compare our kilonova light curves with simulated off-axis afterglow light curves from Van Eerten & MacFadyen (2011). We consider combinations of circumburst number density ($n \in \{1, 10^{-3}\}$cm^{-3}) and jet energy ($E_{\mathrm{j}} \in \{10^{48}, 10^{50}\}$ erg). The densities are typical of short GRBs (Berger, 2014), and the jet energies overlap with our more energetic models (early and wide). The brightest afterglow model we consider, (10^{50} erg, 1 cm^{-3}), is likely to be an upper bound on the off-axis afterglow from one of our jet models. Even in that case, the UV kilonova will outshine the off-axis afterglow for the first half day of evolution.

Non-grey opacities

All of the calculations that we have discussed thus far make the simplifying assumption that the ejecta have a constant grey opacity of $\kappa = 1$ cm^{2}g^{-1}. This assumption is not realistic,
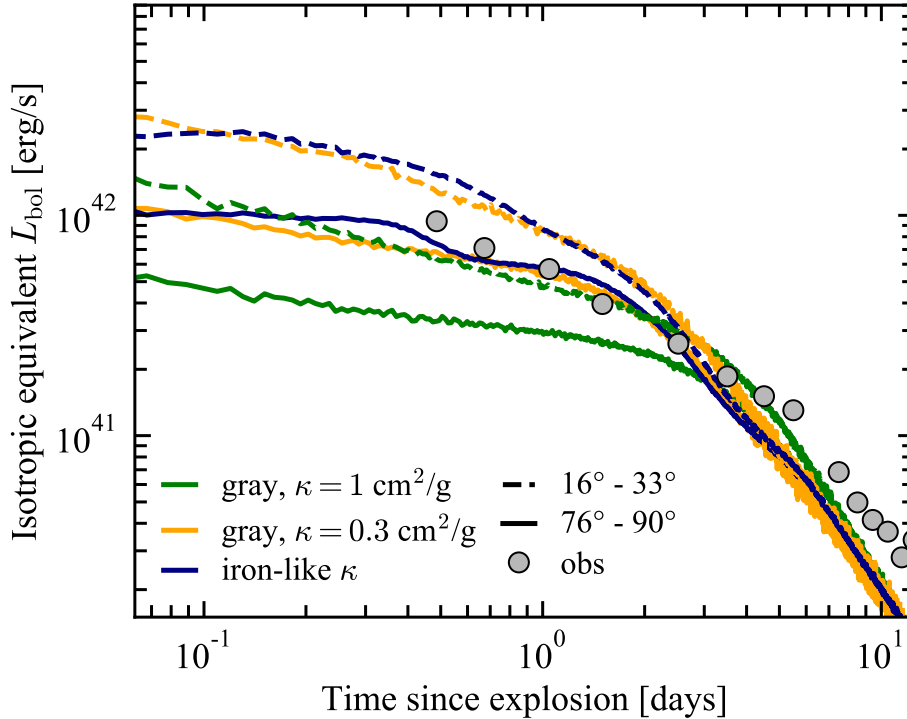
Figure 3.9: Isotropic equivalent bolometric luminosity at different viewing angles for the early breakout case. We compare iron-group-like opacities (blue, representative of first peak r-process elements), to bolometric luminosities for grey opacities of $1\,\mathrm{cm^2g^{-1}}$ (green) and $0.3\,\mathrm{cm^2g^{-1}}$ (orange). Solid lines show equatorial viewing angles, and dashed lines show a near-polar viewing angle. The bolometric luminosity of AT2017gfo is overplotted in grey circles (Drout et al., 2017). All models have a similar viewing angle-dependence: they are brighter as seen from the pole than they are from the equator. The results for iron-group-like opacities are very similar to the grey $\kappa = 0.3\,\mathrm{cm^2g^{-1}}$ results.

since kilonova ejecta are expected to be rich in r-process elements, whose opacity is dominated by a dense forest of atomic lines. There is therefore a concern that the brightening due to the jet-induced asymmetry could be obscured by atomic lines from d- or f-block elements that can blanket the blue and UV portions of the spectrum.

To assess the effect of non-grey opacity, we re-simulate the "early" model but use a composition of half 90-Ca and half 90-Fe to calculate the opacity. We refer to this model as "KFe." This composition approximates the opacity of the first peak of the r-process (Section 3.1). The isotropic equivalent bolometric luminosity is shown in Fig. 3.9. We compare these light curves with those of models with grey opacity $\kappa = \{0.3, 1\}\,\mathrm{cm^2g^{-1}}$ (K0.3 and K1.0, respectively). The figure shows the light curves seen by near-polar and equatorial observers. We do not show a pole-on kilonova light curve, since at those angles, it would be much

dimmer than the accompanying afterglow. The bolometric light curves of the KFe model roughly match those from K0.3; both KFe and K0.3 are brighter than K1.0, but have the same shape. The viewing angle-dependence remains in all models; near-polar viewing angles are brighter than their equatorial counterparts.

While bolometric light curves of grey (K0.3) and more realistic (KFe) models are comparable, the band luminosities differ, particularly in the UV. Fig. 3.10 compares *Swift* UVOT W2 and Cousins R-band light curves for the KFe and K0.3 models. The primary effect of a more realistic opacity is that the UV light curve evolves more quickly, peaking earlier and brighter and dimming more rapidly. This is most pronounced near the equator, where the KFe model peaks at -15.4 mag, and dims by 4.8 mag within the first day of evolution. By contrast, at the same viewing angle, K0.3 peaks slightly dimmer at -15.0 mag but only dims by 2.5 mag by the end of the first day.

The dimming in the UV is due to the rise in bound-bound opacity at ultraviolet wavelengths as the ejecta cool and recombine. Consequently, this dimming is less pronounced near the pole, where the temperature is higher. As seen from near the pole, KFe looks very similar to K0.3 for the first ~ 20 h, before there is a sharp turnover in the KFe UV light curve that is not matched by K0.3.

When UV absorption increases, there is a corresponding increase in re-emission across the spectrum. Photons are effectively scattered from high to low frequency. This leads to a slight enhancement to the R-band light curve after about a day. Overall the R-band light curve is largely unaffected by the choice between iron-like and comparable grey opacity, reflecting the low bound-bound opacity in the infrared and red bands.

3.3 Discussion

A relativistic jet in binary neutron star or neutron star-black hole mergers can affect the light curve seen from a homologously expanding, radioactive outflow. These changes are due to the impact of the jet on the density structure of the ejecta. The jet evacuates a region around its axis, which exposes the hotter inner material that would otherwise remain optically thick. The photosphere in those areas is hotter, and the emission from near the pole is brighter and bluer. The bolometric (ultraviolet) light curves can be enhanced by factors of a few (several) at near-polar viewing angles on $\sim$ day time-scales (see first panels of Figs 3.4 and 3.7). The effect is less pronounced 15°–30° from the pole, though still noticeable. Past 30° and in (infra)red bands, light curves are largely unchanged. It is unlikely that an off-axis afterglow would outshine the kilonova and hide this effect (Fig. 3.8). The degree of brightening depends on the jet parameters. Jets with larger energies and opening angles create larger cavities within the ejecta, so the light curves are more strongly affected. The effects are also not as confined to polar viewing angles (see Figs 3.4 and 3.7).

Our approximate treatments of both the homologous outflow structure and the opacity of the ejecta introduce uncertainty into our quantitative predictions. However, the qualitative difference between the polar and equatorial light curves of a given model and the rough
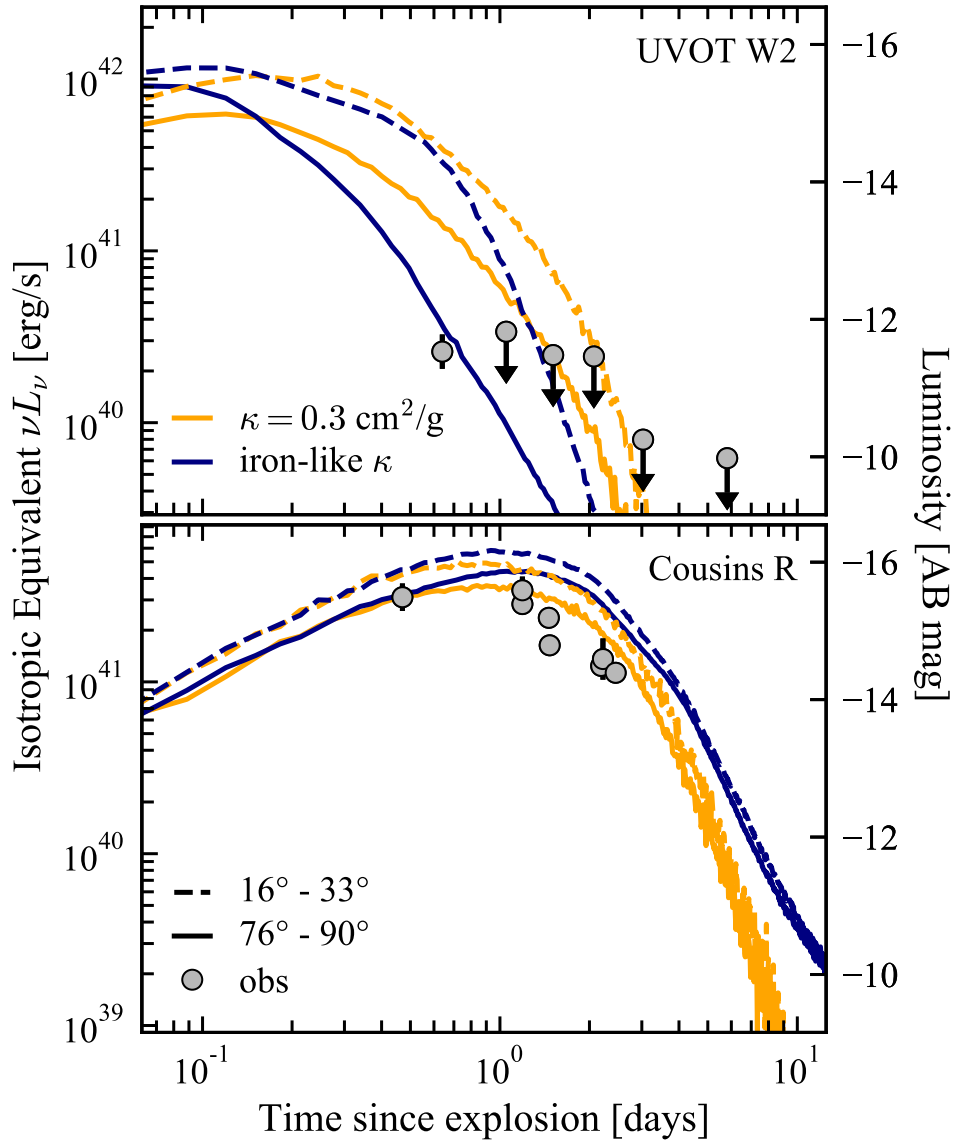
Figure 3.10: Isotropic equivalent band luminosity (*Swift* UVOT W2 and Cousins R) at different viewing angles for iron-like opacity (blue), which we take to be roughly representative of the first r-process peak. We compare these light curves with those from a model with a grey opacity of $0.3\,\mathrm{cm^2g^{-1}}$ (orange). Observations of AT2017gfo are shown in grey circles; arrows indicate upper limits (Kasliwal et al., 2017). We show light curves as seen from the equator (solid) and near the pole (dashed). The UV light curve falls more rapidly with more realistic opacities than the grey model, particularly on the equator. This is due to falling temperatures, which cause the iron and calcium to recombine, increasing the bound-bound opacity. The R-band light curve is largely unaffected by the choice of opacity.

magnitude of the effect should be robust, provided that a cavity exists. We have shown that an evacuated region within an otherwise spheroidal outflow should lead to brighter and bluer emission from inside the cavity. This result should generalize to other similar ejecta configurations, such as magnetically-driven outflows from NS merger accretion disks.

The hydrodynamic simulations that underlie our work are axisymmetric and ignore magnetic fields, which may result in an artificially underdense cavity. In 3D magnetohydrodynamics (MHD), jets may develop kink instabilities, which can disrupt their propagation (Bromberg et al., 2019). There may also be mixing between the jet and the surrounding cocoon (Gottlieb et al., 2021). Both of these effects could increase the density within the jet cavity and reduce the polar brightening. There is also a question of whether hot ejecta will expand latitudinally into the jet cavity, reducing its size or erasing it entirely. Magnetic pressure within the jet cavity may stabilize the channel, but the circumstances under which the cavity is stable are uncertain. Further study of the long-term ($> 10\,\mathrm{s}$), 3D MHD evolution of neutron star merger ejecta is needed to better understand the jet cavities that may form and persist in nature.

Because the impact of the jet cavity is largest in the blue and ultraviolet (see Figs 3.6 and 3.7), our results point to the importance of early UV followup of gravitational wave signals and highlight the value of wide-field UV surveys.

This picture is largely insensitive to our choice of opacity. Emission is brighter and bluer near the pole than on the equator for models that have a small grey opacity as well as for models that have an opacity representative of first peak r-process elements (see Figs 3.9 and 3.10). The UV light curve in this model falls off more quickly than in the grey case. As the ejecta cool and recombine, the bound-bound opacity increases in the blue and ultraviolet. The ejecta are hotter on the pole than on the equator, so the UV light curve stays bright for about a day near the pole, as compared to a few hours further off-axis. We also expect our qualitative results to be independent of any latitudinal variation in opacity, since the effects of a jet cavity on light curves are largely confined to near-polar viewing angles. Material near the poles is expected to have lower rather than higher opacity (Miller et al., 2019). Polar light curves may generally be brighter and bluer, even without a jet due to the lower opacity and larger projected surface area (Korobkin et al., 2021). A jet cavity would enhance this effect.

Our results complement recent work that has studied the effect of a jet on a layer of high-opacity material that may form in a neutrino-driven wind from the NS merger accretion disk (Nativi et al., 2021). Disrupting this layer of occluding material brightens the early kilonova as seen from near the pole. The effects of a jet on the long-term evolution of NS merger ejecta, and therefore on kilonova light curves, remains a rich and interesting direction for future work.

While some of the jet energy is thermalized in the ejecta via shocks, this contribution to the total thermal energy of the ejecta is negligible compared to that from radioactive heating at all of the times we study ($\gtrsim 1\,\mathrm{h}$). These results are consistent with those of Gottlieb et al. (2018), who find that cooling emission from jet shock heating becomes subdominant by $1\,\mathrm{h}$ for opacities typical of Lanthanide-poor material.

Our models assume a minimal delay between the ejection of the more spherical outflow and the start of the jet engine. In our highest jet energy per solid angle case ("early" in Table 3.1), the jet breaks through the ejecta at a radius of $\sim 10^7$ cm. In other cases we consider, ("late" and "wide" in Table 3.1) the jet/shock breaks out at $\sim 10^{10}$ cm, after about a second of expansion. After breakout, the deposited thermal energy falls off as the reciprocal of the radius of the ejecta. By delaying the start of the engine and using a wide jet opening angle, some prior work has suggested that the cooling emission from shock-heated regions can reach luminosities in excess of 10^{42} erg/s on time-scales of several hours to days (Kasliwal et al., 2017; Piro & Kollmeier, 2018). However, the energies supplied by jet heating in our calculations — which have jet opening angles and energies typical of short GRBs (Fong et al., 2015) — are factors of 100–1000 less than that supplied by the r-process (Fig. 3.2). In order for shock heating to be significant for the kilonova light curve, the jet energy would need to be increased by a large factor ~ 100, the "jet" would need to be relatively spherical, and/or there would need to be shocks between the jet and ejecta at distances much greater than the 10^{10} cm found here. We regard these as unlikely.

Chapter 4

The Impact of R-Process Heating on the Dynamics of Neutron Star Merger Accretion Disc Winds and Their Electromagnetic Radiation

This chapter is modified from Klion et al. (2021a), with permission from the co-authors.

In this chapter, I focus on understanding how r-process heating affects the structure of a 3D MHD-driven disc wind and, consequently, the resulting kilonova light curves. I construct an inner boundary condition for 2D GR hydrodynamics (GRHD) simulations based on the wind formed in the 3D GRMHD simulation of Fernández et al. (2019, hereafter F19), approximately accounting for r-process heating that occurs before injection. I perform a second 2D GRHD simulation without r-process heating. I present a formalism for approximately including r-process heating in Section 4.1, and discuss the details of my 2D GRHD simulations in Section 4.1. After 60 s of evolution, I pass the resulting density and temperature structures to SEDONA, a multi-dimensional, multi-frequency radiation transport code (Sections 4.1 and 4.1). In Section 4.2, I compare the results of our HARM simulations with and without r-process heating to assess the effect of r-process heating on the dynamical evolution of kilonova disc winds. Our predictions for disc wind light curves, with and without dynamic r-process heating, are found in Section 4.3. I conclude and discuss future research directions in Section 4.4.

4.1 Methods

Our full calculation consists of the following steps:

1. Simulate a post-merger black hole accretion disc in 3D GRMHD to 10 s. This model is the one presented in F19, and does not include r-process heating.

2. Extract wind properties from F19 at a radius of $r_b = 2 \times 10^4$ km, and use them to set an inner boundary condition for a 2D GRHD simulation. There are two versions of these conditions: one that is simply axisymmetrized, and another that is axisymmetrized and then modified to approximately account for r-process heating that occurred within the domain of the 3D simulation, prior to the gas reaching r_b.

3. Evolve two 2D GRHD simulations, each with one of the sets of boundary conditions described above. The simulation "with r-process" uses the r-process-adjusted inner boundary conditions and self-consistently includes an r-process heating source term during evolution. Meanwhile, the simulation "without r-process" does not include an r-process heating term, and uses the axisymmetrized boundary conditions (not modified to include r-process heating). We evolve both of these models until 60 s, at which point they are mostly in free expansion.

4. Calculate the optical emission from both of the 2D GRHD simulations using SEDONA, a Monte Carlo radiation transport code. In this phase of the calculation, both models include r-process heating, but the evolution is assumed to be homologous. The velocity structure is fixed, so the heating only affects the light curve. This allows us to isolate the effects of r-process heating on the dynamics of the ejecta.

2D GR Hydrodynamic Initial Conditions

We continue the 3D GRMHD simulation of F19 in 2D GRHD. Unless otherwise stated, we take $G = c = 1$. F19 initialize a torus of mass $0.033 \, \mathrm{M}_\odot$ with a strong poloidal magnetic field around a black hole of mass $3 \, \mathrm{M}_\odot$ and spin parameter $a = 0.8$. The simulations of F19 were performed before the announcement of GW170817/AT 2017gfo. Subsequent modeling has inferred that the initial torus mass was larger by a factor of ~ 3 (Shibata et al., 2017). We extract the time-series of the primitive variables at a radius of $r_b \equiv 2 \times 10^4$ km $= 4452 \, r_g$ in the 3D simulations: rest mass density ρ, energy density ε, four velocity in Kerr-Schild coordinates $\{u^t, u^r, u^\theta, u^\phi\}$ and composition (electron fraction Y_e and abundances of protons, neutrons, and alpha particles). This radius is small enough that the entire wind crosses through the surface before the end of the 3D simulation, but large enough that no bound material falls back through.

We make the simplifying assumption that the magnetic fields are zero. While they are critical for the jet launch and wind ejection, both of these occur interior to the inner boundary of our simulation. We are primarily interested in the baryon-rich wind, and do not expect the exclusion of magnetic fields to significantly affect our results.

We axisymmetrize the density by averaging in ϕ:

$$\langle \rho \rangle = \frac{\int \rho \sqrt{-g} \, \mathrm{d}\phi}{\int \sqrt{-g} \, \mathrm{d}\phi}, \tag{4.1}$$

and calculating mass-weighted averages for the other variables, X:

$$\langle X \rangle = \frac{\int X \rho \sqrt{-g} \, \mathrm{d}\phi}{\int \rho \sqrt{-g} \, \mathrm{d}\phi}. \tag{4.2}$$

The determinant of the metric is g. These axisymmetrized values are then used as a time-dependent inner boundary condition in 2D GRHD simulations. We perform two simulations: one with and one without r-process heating. When we do not include r-process heating, we directly use the axisymmetrized primitives at r_b as the boundary condition.

The majority of the thermal energy input by the r-process will be converted to kinetic energy via adiabatic expansion of the hot gas. The decay of r-process elements deposits $\sim 4 \times 10^{18} \, \mathrm{erg \, g^{-1}}$ on a time-scale of $\sim 1 \, \mathrm{s}$, which corresponds to a velocity increase of $\sim 0.1 \, c$ for a particle at rest. Of the mass crossing r_b in F19, $8.7 \times 10^{-3} \, \mathrm{M_\odot}$ (68 per cent) has a velocity $< 0.1 \, c$, so we expect acceleration from this burst of energy to be significant. The effects of Y_e on heating rates are generally modest (Lippuner & Roberts, 2015), so we assume a spatially uniform, time-dependent heating rate. We adopt the $Y_\mathrm{e} = 0.1$ r-process heating rate from Metzger et al. (2010). We approximate the time-dependent heating rate per mass $\eta(t)$ as a broken power law:

$$\eta(t) = \begin{cases} 6.7 \times 10^{17} \, \mathrm{erg/g/s} & t_1 < 0.01, \\ 1.1 \times 10^{19} t_1^{0.6} \, \mathrm{erg/g/s} & 0.01 \leq t_1 < 0.5, \\ 7.0 \times 10^{17} t_1^{-3.33} \, \mathrm{erg/g/s} & 0.5 \leq t_1 < 4, \\ 4.2 \times 10^{16} t_1^{-1.3} \, \mathrm{erg/g/s} & t_1 > 4. \end{cases} \tag{4.3}$$

where $t_1 = t/1 \, \mathrm{s}$. This rate assumes that all of the nuclear energy emitted by radioactive decays is converted to thermal energy. In the first several seconds, neutrinos are the only decay product that can easily escape, while the remainder are thermalized. Neutrinos likely carry away $\sim$ 20–30 per cent of the total energy (Metzger et al., 2010; Barnes et al., 2016). This adjustment would be less than the uncertainty in the r-process heating rate ($\sim$ factor of a few). We therefore assume complete thermalization during the hydrodynamic phase of the calculation. We take $t = 0$ for the r-process to be at the start of the simulation of F19. Physically, the r-process begins once the material leaves NSE, which will occur when the temperature drops below $\sim 10 \, \mathrm{GK}$. These temperatures are only achieved within a radius of $50 r_g = 225 \, \mathrm{km} \ll r_b$, and at early times $t \lesssim 300 \, \mathrm{ms}$. By contrast, the median r_b-crossing time is $1.7 \, \mathrm{s}$. By the time the wind enters our domain, it has been out of NSE for some time, and it is a reasonable approximation that it left NSE at $t = 0$.

Before we axisymmetrize, we adjust the velocities and internal energy to include the effects of r-process heating. To calculate the modified boundary condition at time t_b, we consider a fluid parcel of mass m that is crossing r_b at time t_b. We assume that it is expanding homologously and has been doing so since the start of the 3D simulation. This is an approximation since the parcel will accelerate, so its evolution is not actually homologous.
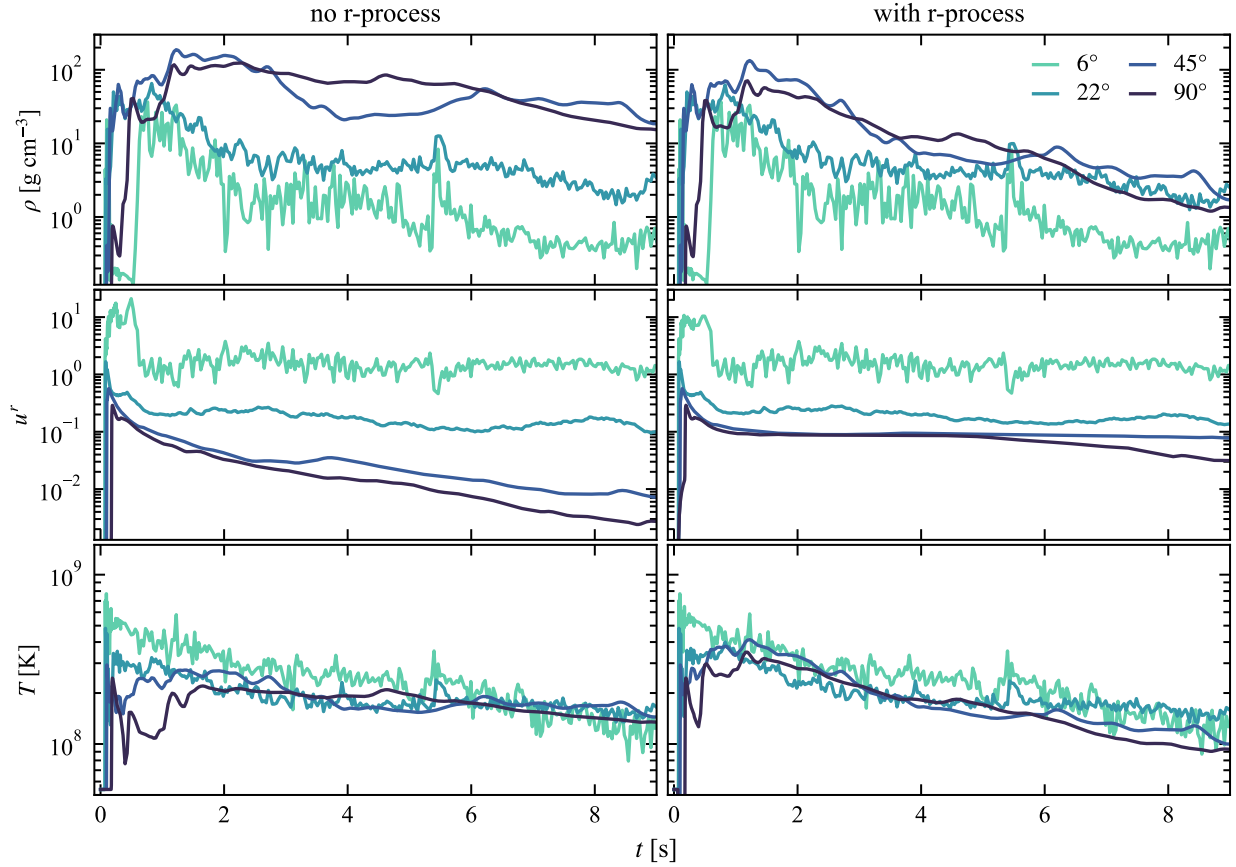
Figure 4.1: Inner boundary condition at $r_b = 2 \times 10^4$ km in the 2D GRHD simulations of long-term disc wind evolution (see Section 4.1), derived from the 3D GRMHD simulation of F19. The left (right) column shows the conditions used for the simulation without (with) r-process heating. The conditions on the right include approximate effects of r-process heating that occurs before injection. We show rest mass density (top panel), radial four-velocity (middle), and radiation temperature (bottom) as a function of time for a representative selection of angles. Temperature is calculated from internal energy density ε assuming radiation pressure dominates, $T = (\varepsilon/a)^{1/4}$, where a is the radiation constant. The conditions are largely symmetric across the equator, so we only show angles in the Southern hemisphere; angles are measured from the South pole. The effects of r-process heating are more pronounced toward the higher-density, lower-velocity equatorial region.
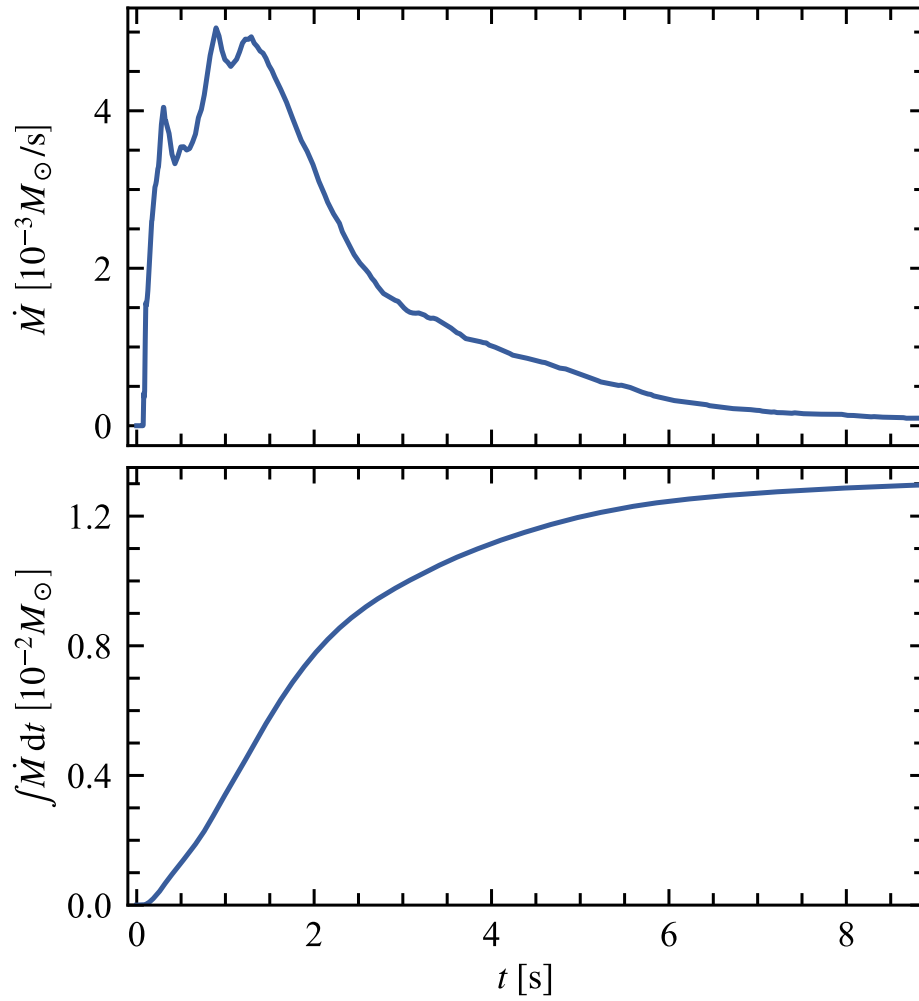
Figure 4.2: Mass injection rate $\dot{M}$ (top panel) and total injected mass (bottom panel) in our GRHD simulations as a function of time. The majority of the mass injection occurs within the first few seconds of the simulation, with a median injection time of 1.6 s. By construction, the mass injection rate is the same in the HARM runs with and without r-process heating.

Also, it would have been ejected from the disc at some time after the start of the 3D calculation. None the less, this should give a reasonable estimate of the distribution of r-process energy into thermal and kinetic. Improving on this approximation would require the fluid trajectories, which we do not have. The parcel has volume $V(t)$ at time t. Unless otherwise indicated, we work in the fluid frame. The gas has been heated radioactively since $t_0 = 0\,\mathrm{s}$ by $\eta(t)$. Its thermal energy due to radioactive heating e_{rad} will evolve as

$$\frac{\mathrm{d}e_{\mathrm{rad}}}{\mathrm{d}t} = \frac{-e_{\mathrm{rad}}(t)}{t} + \eta(t)m. \tag{4.4}$$

The first term on the right-hand side accounts for thermal losses due to adiabatic expansion. Solving the above differential equation and dividing by the volume $V(t)$ gives the internal energy density due to radioactive decays $\varepsilon_{\mathrm{rad}}$,

$$\varepsilon_{\mathrm{rad}}(t) \equiv \frac{e_{\mathrm{rad}}(t)}{V(t)} = \frac{\rho(t)}{t} \int_{t_0}^{t} \eta(t')t'\,\mathrm{d}t'. \tag{4.5}$$

We add $\varepsilon_{\mathrm{rad}}(t_b)$ to the internal energy density at r_b. We assume that the remainder of the total r-process energy is converted to kinetic energy, giving the parcel a boost Δe_{kin}:

$$\Delta e_{\mathrm{kin}}(t_b) = e_{\mathrm{rad,tot}}(t_b) - e_{\mathrm{rad}}(t_b), \tag{4.6}$$

where

$$e_{\mathrm{rad,tot}}(t) = m \int_{t_0}^{t} \eta(t')\,\mathrm{d}t' \tag{4.7}$$

is the total energy deposited by r-process heating up to time t. The kinetic energy boost $\Delta e_{\mathrm{kin}} \geq 0$ because the thermal energy attributed to radioactive decays can never exceed the total heating by radioactivity. The conversion of non-radioactive thermal energy to kinetic energy is already handled in the underlying 3D simulation.

The energies and masses in this calculation are in the fluid frame. Ideally, we would use the fluid proper time to calculate the heating rate. Calculating the proper time for each parcel of fluid is infeasible since we do not have information about the trajectory of individual fluid elements. For the sake of consistency, we will use the observer time throughout this chapter when evaluating $\eta(t)$. We do not expect this to have a substantial effect on our results. The vast majority of the mass has $v < 0.2\,c$, so proper and observer time will generally be equal. Additionally, the bulk of the heating occurs in the first few seconds, so the total energy deposited over the $\sim 10^2\,\mathrm{s}$ of the hydrodynamic simulation will be the same to within a few per cent. Effects on the dynamical evolution of the wind should therefore be negligible.

The combined rest mass and kinetic energy of the parcel is γm, where γ is the Lorentz factor. If we add Δe_{kin} in kinetic energy, the new energy of the parcel becomes

$$\gamma_{\mathrm{new}}m = \gamma_{\mathrm{old}}m + \Delta e_{\mathrm{kin}}(t_b), \tag{4.8}$$

and its new Lorentz factor is

$$\gamma_{\mathrm{new}} = \gamma_{\mathrm{old}} + \Delta e_{\mathrm{kin}}(t_b)/m. \tag{4.9}$$

The magnitude of the new 3-velocity is v_{new}, which can be calculated directly from γ_{new}. We scale the components of the 3-velocity by $v_{\mathrm{new}}/v_{\mathrm{old}}$. While we set our boundary conditions in terms of densities, e.g. ρ and ε, the physically significant quantity is the flux of mass and energy onto the grid. Accordingly, we scale the boundary ρ and ε by the ratio of the radial 4-velocities: $u^r_{\mathrm{old}}/u^r_{\mathrm{new}}$. This preserves the homologous estimate of the kinetic-thermal energy distribution.

Fig. 4.1 shows the time-dependent boundary condition used in our axisymmetric GRHD simulations for four representative angles. We show rest mass density, radial four-velocity (u^r) and temperature. We calculate the temperature from the internal energy density ε assuming that the gas is radiation-dominated, $T = (\varepsilon/a)^{1/4}$, where a is the radiation constant. Due to the jet, the polar regions have high velocity, particularly in the first $\sim 0.5\,\mathrm{s}$, where $u^r > 1$. The high velocity in the polar regions is largely unaffected by the addition of r-process kinetic energy. However, the slower material in the bulk of the wind is accelerated to a velocity of $\sim 0.1\,c$. We hold the mass flux constant, so the density of the wind is correspondingly lower in the case with r-process heating.

The r-process boosts the temperature of the early wind. At later times, most of the energy has been converted to kinetic, and the thermal energy increases due to r-process heating are modest. While the thermal energy flux is larger in the model with r-process heating, the energy density, and therefore temperature, is lower due to the much higher velocity of the material.

By construction, the mass injection rate $\dot{M}$ is the same in both GRHD simulations. $\dot{M}$ and its integral are shown in Fig. 4.2. The bulk of the mass enters the grid in the first 2–3 s.

We use the boundary data calculated from F19 for the duration of their GRMHD simulation ($\sim 9\,\mathrm{s}$). We interpolate the boundary conditions linearly in time and meridional angle. After 9 s have elapsed, we linearly taper the velocity at the boundary to zero, while setting the other variables equal to the floor values.

GRHD Simulations in 2D

We use HARMPI[1] (Tchekhovskoy, 2019), a parallel version of the code HARM (Gammie et al., 2003; Noble et al., 2006), to perform 2D axisymmetric GRHD simulations of the long-term disc wind evolution. We use the same Kerr metric as in F19, though the spacetime is approximately flat at the radii of interest. We work in modified spherical-polar Kerr-Schild coordinates. Our domain extends from $r_b = 4552r_g$ to $r_{\mathrm{out}} = 10^6 r_b$, and is discretized into 1024 radial points. The first 895 are spaced logarithmically between r_b and $r_t = 10^4 r_b$. The remaining 129 are spaced progressively more sparsely between r_t and r_{out}. This would allow us to run our simulation until $t \approx r_t/c = 680\,\mathrm{s}$ before encountering any artifacts due to

---

[1]Available at `https://github.com/atchekho/harmpi`

the grid boundary. The meridional grid is the same as in F19, but with double the number of cells. It covers the interval $[0, \pi]$, and consists of 512 cells, with resolution concentrated at the poles and near the equator. We employ an outflow boundary condition at r_{out}, and a reflective boundary condition in the meridional direction. The inner radial boundary condition is a time-dependent condition, as described in the previous section. In this phase, we perform two simulations, one that uses the axisymmetrized boundary conditions, and another that uses the r-process adjusted boundary conditions.

Throughout, we use an ideal gas equation of state with fixed adiabatic index $\gamma_{\mathrm{ad}} = 4/3$. Our simulations include the (anti)neutrino emission and alpha particle recombination models detailed in F19. However, the temperatures in our simulation are below the thresholds where these processes are significant, so the composition of the flow is a passive scalar and does not affect our results.

In the simulation that uses the r-process adjusted boundary conditions, we include an r-process heating term. During each time step Δt,

$$\Delta \varepsilon_{\mathrm{rad}} = \rho \eta(t) \frac{\Delta t}{u^t}, \tag{4.10}$$

is added to the fluid-frame internal energy in between updates to the conserved quantities. While we evaluate the heating rate per mass $\eta(t)$ (equation 4.3) in the lab frame, we still compute the added energy in the fluid frame. The factor of u^t arises because the fluid-frame time step is $\Delta t/u^t$.

We evolve our simulations to at least $t_H = 60\,\mathrm{s}$, when the flow has largely reached free expansion, and the structure has stopped evolving. At this point, regions of the ejecta with high Mach number suffer from numerical errors that artificially increase the internal energy. These are most prominent in the simulation with no r-process heating, whose thermal structure is not of interest. When constructing the SEDONA input model from the 2D HARM simulation with no r-process heating, we re-set the thermal energy density (but not the kinetic energy) to be that expected from radioactive decays plus adiabatic degradation,

$$\varepsilon(t_H) = \frac{\rho(t_H)}{t_H} \int_{t_0}^{t_H} \eta(t')t'\,\mathrm{d}t'. \tag{4.11}$$

Radiation Transport Initial Conditions

When the HARM models reach homology, the ejecta are still very optically thick. It is impractical to start radiation transport calculations on time-scales of $\sim 60\,\mathrm{s}$ due to the computational expense. Because we focus on accurately modeling the emission on time-scales of $> 1\,\mathrm{h}$, it is unnecessary to start radiation transport calculations before $t_S = 300\,\mathrm{s}$. Between t_H and t_S, evolution is well-modeled by homologous expansion. At the start of the SEDONA calculation, the density will be

$$\rho(t_S) = \rho(t_H) \left(\frac{t_H}{t_S}\right)^3. \tag{4.12}$$

R-process heating and adiabatic degradation of internal energy will continue as well, giving

$$\varepsilon(t_S) = \varepsilon(t_H)\frac{t_H}{t_S} + \frac{\rho(t_S)}{t_S} \int_{t_H}^{t_S} \eta(t')t' \, \mathrm{d}t'. \tag{4.13}$$

When constructing the input models for SEDONA, we exclude the high Mach-number regions that are susceptible to numerical errors in the internal energy. We apply density and radial cuts to the ejecta, only including cells where

$$\rho > \rho_{\mathrm{cut}} = 6.9 \times 10^{-7} \left(\frac{t}{60\,\mathrm{s}}\right)^3 \mathrm{g\,cm}^{-3} \tag{4.14}$$

and

$$\frac{r}{t} < 0.8c. \tag{4.15}$$

This contour is overplotted on density maps in Fig. 4.4. From 30 s onwards, this contour includes approximately 95 per cent of the mass on the grid in both simulations. Neither of these cuts affect the shapes of the light curves. Expanding the r/t cut to $0.9\,c$ does not affect the light curves. Lowering the value of ρ_{cut} by a factor of 10 increases the mass enclosed to 98 per cent and causes the overall luminosity of the early light curve to increase slightly (by 10 per cent).

Radiation Transport Simulations

To predict bolometric and broadband light curves from these disc winds, we use SEDONA, a time-dependent, multi-dimensional, multi-wavelength Monte Carlo radiation transport code (Kasen et al., 2006; Roth & Kasen, 2015). SEDONA tracks the emission and propagation of packets of radiation ("photons") through the freely-expanding ejecta. Interactions between the particles and fluid are calculated in the fluid frame, which naturally accounts for adiabatic losses as well as most relativistic effects. The one relativistic effect we neglect is in the evaluation of $\eta(t)$, where we do not distinguish between proper and lab-frame time (see Section 4.1). When particles leave the grid, they are tallied according to their wavelength, direction, and observer arrival time, providing time- and viewing-angle-dependent light curves and spectra.

The ejecta density and temperature are discretized on a cylindrical velocity grid. Since SEDONA requires homologous expansion, we reset the velocity from the HARM models to be purely radial and have magnitude $v = r/t$. The regions within our radial and density cuts are already mostly homologous so this should not significantly affect our results. Within the SEDONA run, density evolves as t^{-3}. The temperature is re-calculated at each time-step by equating the energy from photon absorption and radioactive heating with thermal emission. Adiabatic losses arise from the frame-shifting in the particle-fluid interactions.

The initial radiation field is represented by 10^7 particles. At each time step, 3×10^5 photons are created, representing the emission from r-process decays. We use a parametrized,

time-dependent r-process heating rate that assumes an initial entropy of $32\,k_\mathrm{B}\,\mathrm{baryon}^{-1}$, an expansion time-scale of $0.84\,\mathrm{ms}$, and electron fraction of 0.13 (Lippuner & Roberts, 2015). This rate is adjusted by a time-dependent thermalization fraction that ranges between $\sim$ 50–75 per cent (Barnes et al., 2016). The resultant net heating rate roughly agrees with equation 4.3 between t_H and t_S. As in the HARM calculation, we use the lab frame time to evaluate η. This may have a larger effect in the light curve calculation since the instantaneous luminosity is more important than the total energy budget. SEDONA uses a flat spacetime, so the time dilation factor is γ. For a power-law heating rate $\propto t^{-a}$, we underestimate the heating rate by a factor of γ^a, and the total deposited energy by γ^{a-1}. The fastest material on the SEDONA grid has $\gamma = 1.7$. After t_S, $a \approx 1.3$, so for the fastest material, we may underestimate the r-process luminosity by a factor of 2, and the total energy deposited within a time-step by 20 per cent. We expect the overall magnitude of this effect to be small since 98 per cent of the mass in the SEDONA simulation has $v < 0.5\,c$, for which luminosity is underestimated by less than 20 per cent.

We adopt a uniform grey (frequency-independent) opacity $\kappa = 1\,\mathrm{cm}^2\mathrm{g}^{-1}$. This is approximately the Planck mean opacity of a mixture of first-peak r-process elements at $1\,\mathrm{d}$ post-merger (Kasen et al., 2013; Tanaka et al., 2020). On a similar time-scale, a mixture of Lanthanides has a higher average opacity, $\sim 10\,\mathrm{cm}^2\mathrm{g}^{-1}$. We are primarily interested in the relative differences between ejecta with and without dynamic r-process, so the choice of a particular grey opacity will not affect the comparison, since the models including and excluding dynamic r-process heating will be affected in the same way.

4.2 Hydrodynamic Results

We follow our 2D GRHD simulations to $60\,\mathrm{s}$, by which point the flow is mostly in homology ($v \propto r$). Fig. 4.3 compares velocity profiles for selected angles against the profile that SEDONA imposes on the ejecta ($v = r/t$). Deviations from homology are small, especially in the lower-velocity regions that contain most of the mass. Further evolution in GRHD brings the polar regions in line with the velocity structure assumed by SEDONA, but allows for more numerical errors to accumulate in the internal energy in high Mach-number regions. After $60\,\mathrm{s}$, changes to the light curves due to these errors exceed those from further evolution of the density structure.

Density snapshots from our 2D GRHD simulations are shown in Fig. 4.4. In the first few seconds, there are minimal differences between the two simulations. The r-process energy has not yet caused the gas to expand or accelerate. By $10\,\mathrm{s}$, the effects become noticeable, as the heating blurs some of the small-scale structure in the wind. The broad structure stabilizes around 30 to $60\,\mathrm{s}$. R-process heating accelerates the slowest material to $\sim 0.1\,c$, which in this case, evacuates a sphere of radius $0.1ct$ at the center of the domain.

The initial and final $\mathrm{d}M/\mathrm{d}v_r$ and $\mathrm{d}M/\mathrm{d}\theta$ distributions are shown in Fig. 4.5. In both models, the vast majority of the mass is at relatively low velocity $\lesssim 0.2\,c$. Without r-process heating, the velocity distribution of the wind is mostly unchanged by further hydrodynamic
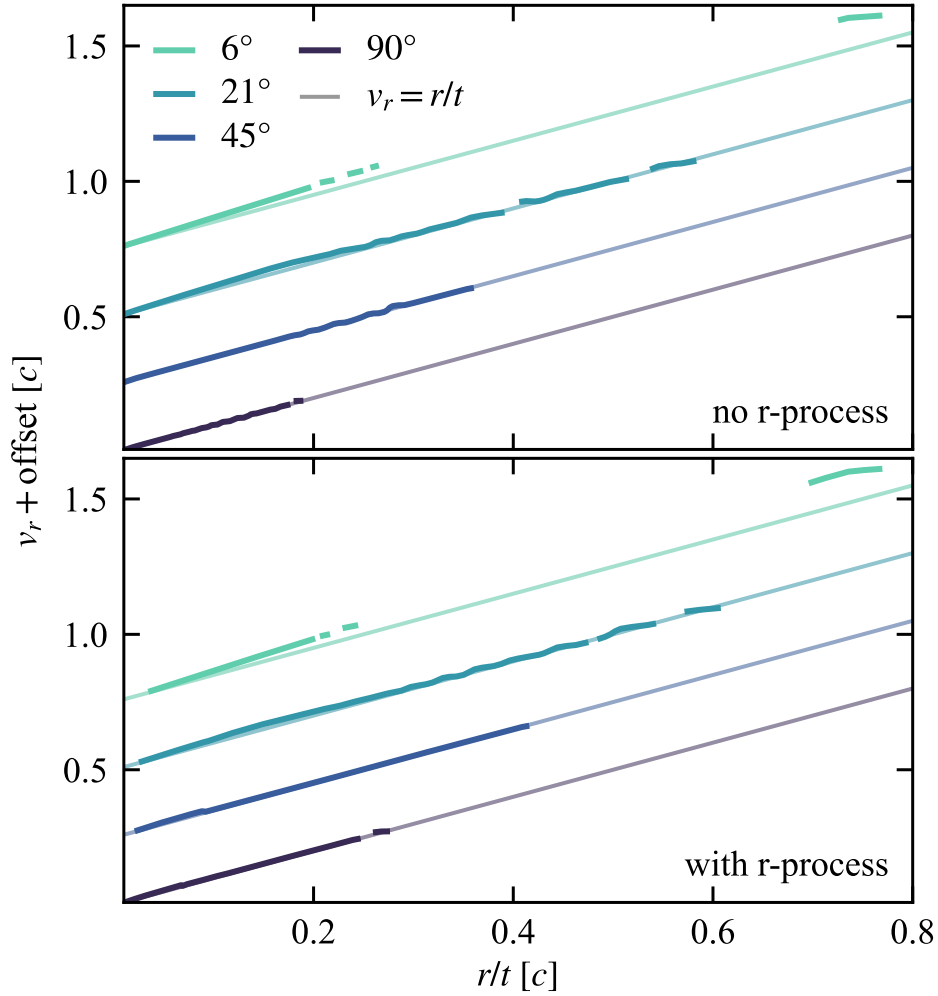
Figure 4.3: Radial profiles of radial velocity at a set of representative angles at $60\,\mathrm{s}$ (dark colors) in our long-term 2D GRHD simulations of NS merger disc winds. Angles are measured from the South pole. Different angles are separated by an offset. The light colored lines show homology ($v_r = r/t$) for a given angle, which is generally an excellent approximation to the simulation results; most of the motion is self-similar. The profiles are only shown where $\rho > \rho_{\mathrm{cut}}$ (equation 4.14). The top (bottom) panel shows results from the simulations without (with) r-process heating.
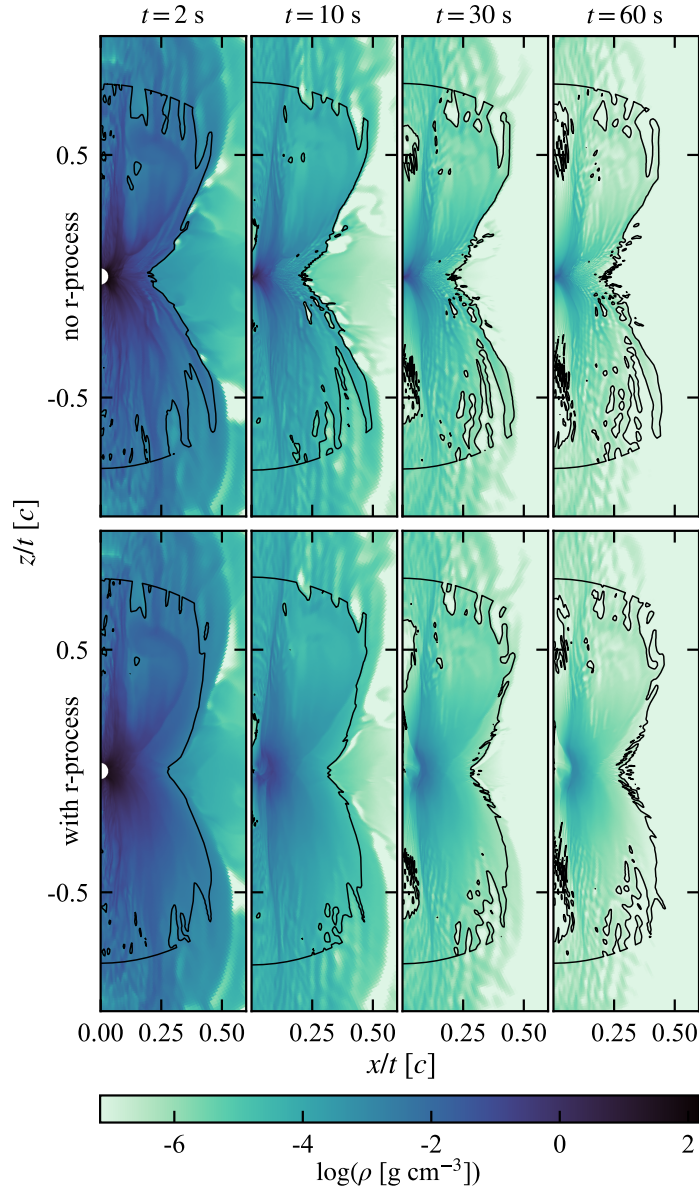
Figure 4.4: Snapshots of mass density after 2, 10, 30, and 60 s of evolution in the long-term 2D GRHD simulations discussed in Section 4.1. The top (bottom) panels show the runs without (with) r-process heating. The spatial bounds are scaled by the snapshot time. The black contour bounds the region included in the SEDONA radiation transport simulations ($r/t < 0.8\,c$ and $\rho > \rho_{\rm cut}$). The primary consequence of including r-process heating is that the low-velocity portion of the wind is accelerated to $\sim 0.1\,c$. Heating also smooths out the finer density structure. The effects also become more prominent as time progresses, and the gas thermal energy has time to be converted to kinetic energy.
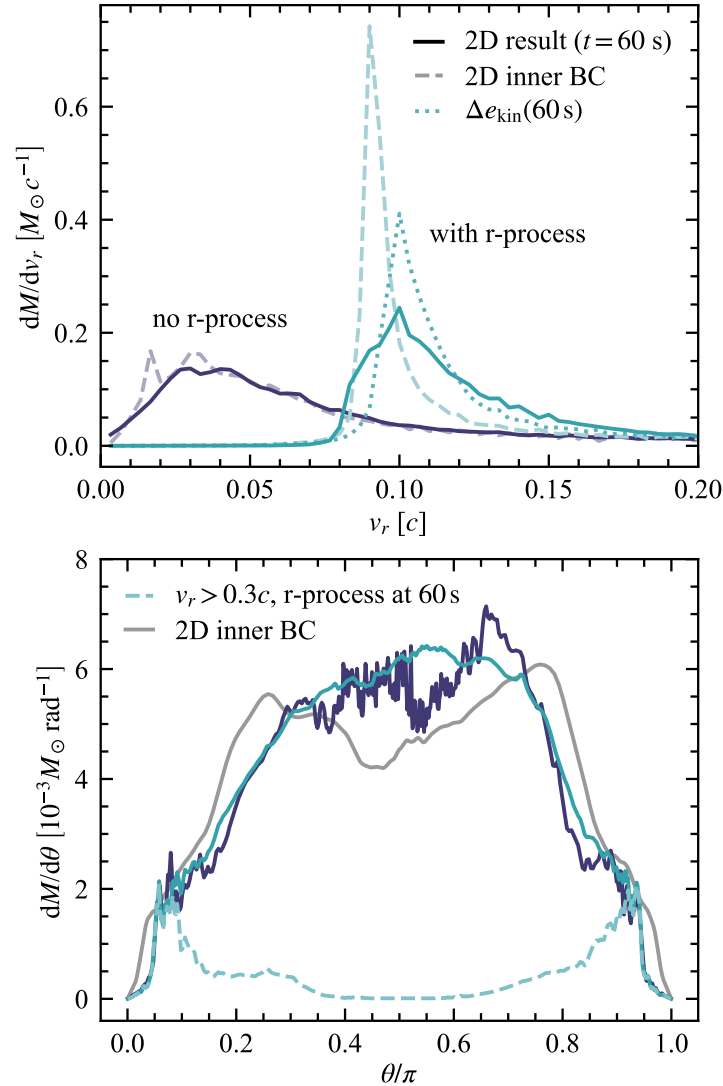
Figure 4.5: Distribution in v_r (top panel) and meridional angle θ (bottom panel) of the material at the end of the 2D GRHD simulations ($t = 60$ s; solid lines and darker colors). We compare the simulations with (teal) and without (purple) r-process heating. Top panel: The radial velocity distributions of material injected at the inner boundaries of the 2D simulations are shown in lighter colors and dashed lines. The dotted line shows the mass distribution obtained by applying the formalism of Section 4.1 but evaluating Δe_{kin} at 60 s. Bottom panel: The angular distribution of the input is the same in both simulations and is shown in grey in the bottom panel. The distributions remain similar at 60 s, though the inclusion of r-process heating smooths small-scale structure in the distribution. The material near the pole is almost all high-velocity ($v_r > 0.3\,c$, dashed line in bottom panel), which is not affected by r-process heating.

Figure 4.6: Same as the top panel of Fig. 4.5, but showing the high-velocity tail of the mass distribution, and only showing the resulting distributions at $60\,$s. The slopes of power-law fits to $\mathrm{d}M/\mathrm{d}v_r$ for both models are shown in black. We separately fit low-velocity $\lesssim 0.3\,c$ material and high-velocity matter, since the latter is dynamically unaffected by r-process heating.

evolution. There is much more of a change between the initial and final distributions with the inclusion of r-process heating. The li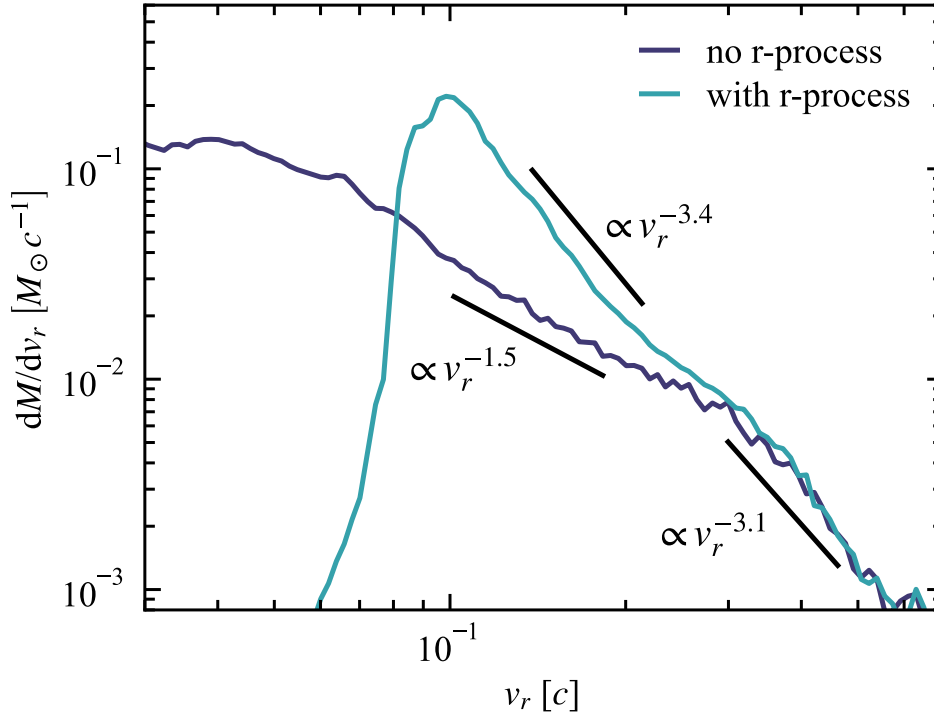ght teal line in the top panel of Fig. 4.5 is the inner boundary condition in our GRHD simulation with r-process heating. This shows a distribution that is much more sharply concentrated in velocity than the final configuration after the subsequent r-process heating and hydrodynamic evolution (dark teal line). The input distribution has almost no material with $v > 0.15\,c$. That is because the higher-velocity material is the first to enter the domain and has not yet a) experienced as much r-process heating, b) converted it to kinetic energy. A check of the formalism of Section 4.1 (equations 4.6–4.9) is to evaluate Δe_{kin} at $60\,$s, and compare the resulting distribution to the results of the 2D GRHD simulation at $60\,$s. The formalism accurately predicts the modal velocity of the wind, but underestimates the spread in the radial velocity distribution. The difference is especially pronounced for velocities below the mode ($v \lesssim 0.1\,c$).

It is also possible that our GRHD simulations with r-process heating underestimate the amount of mass with $v < 0.07\,c$. In the underlying 3D GRMHD simulation, there is

marginally-bound material that reaches a maximum radius $< r_b$, but would become un-bound with the additional boost from r-process heating. The radial cut at r_b excludes this material, in effect applying a total energy floor to the ejecta. This enforces an artificial minimum velocity on the wind, which could in turn suppress the central density in our r-process model. Ideally, r-process heating should be included from the start of the GRMHD simulations of NS merger discs.

As expected, dM/dv_r is unaffected by r-process heating above a velocity of $0.3\,c$ (Fig. 4.6). In that region, we find that $dM/dv_r \propto v_r^{-3.1}$. We exclude material where $\rho < \rho_{\text{cut}}$, but it is still possible that the slope of the mass distribution is affected by the density floor of the 3D GRMHD simulation, so this fit may not be that accurate. Below $0.3\,c$, r-process heating causes the slope to steepen, giving a rough power law fit $dM/dv_r \propto v_r^{-3.4}$, as compared to the distribution $dM/dv_r \propto v_r^{-1.5}$ found when fitting the distribution in the simulation without r-process heating. Fits are calculated over the interval $0.1\,c \leq v_r \leq 0.25\,c$ for the model with r-process heating, and $0.03\,c \leq v_r \leq 0.25\,c$ for the model without. The values of the slopes are unlikely to be that robust, but the steepening of the slope is a general prediction of r-process heating.

The angular distributions are also similar between the GRHD simulations with and without r-process. In both cases, the mass distribution is more equatorially-concentrated at the end of the 2D HARM simulations than it is at injection. This is due to the meridional component of the velocity at injection. The inclusion of r-process heating smooths some of the features in angular structure, and slightly broadens the structure. The poles remain relatively evacuated, which may be artificial, since 87 per cent of the energy from r-process is deposited off of the grid, i.e. prior to injection into the 2D GRHD simulation. The large fraction of the energy injected off-grid may lead our models to underestimate the extent to which the lower-velocity material will fill in the jet cavity. We account for the conversion of thermal to kinetic energy, but only apply this in the direction of the existing velocity vector. Our formalism does not account for meridional expansion that could occur from the early heating prior to injection.

That said, almost all of the polar material is moving faster than $0.3\,c$ (Fig. 4.5), which is a portion of the wind whose kinematics seem mostly unaffected by r-process heating. We also neglect magnetic fields, which will likely resist the gas's expansion into the cavity at the pole.

4.3   Radiation Transport Results

Using the Monte Carlo radiation transport code SEDONA, we calculate viewing angle-dependent bolometric and band light curves for our disc wind models with and without dynamical r-process heating. We include r-process heating in both SEDONA simulations, since instantaneous r-process heating at time $\sim t$ sets the luminosity at time $\sim t$. The code assumes homologous expansion, which is a good approximation at this stage; the heating does not affect the dynamics and only powers the kilonovae.
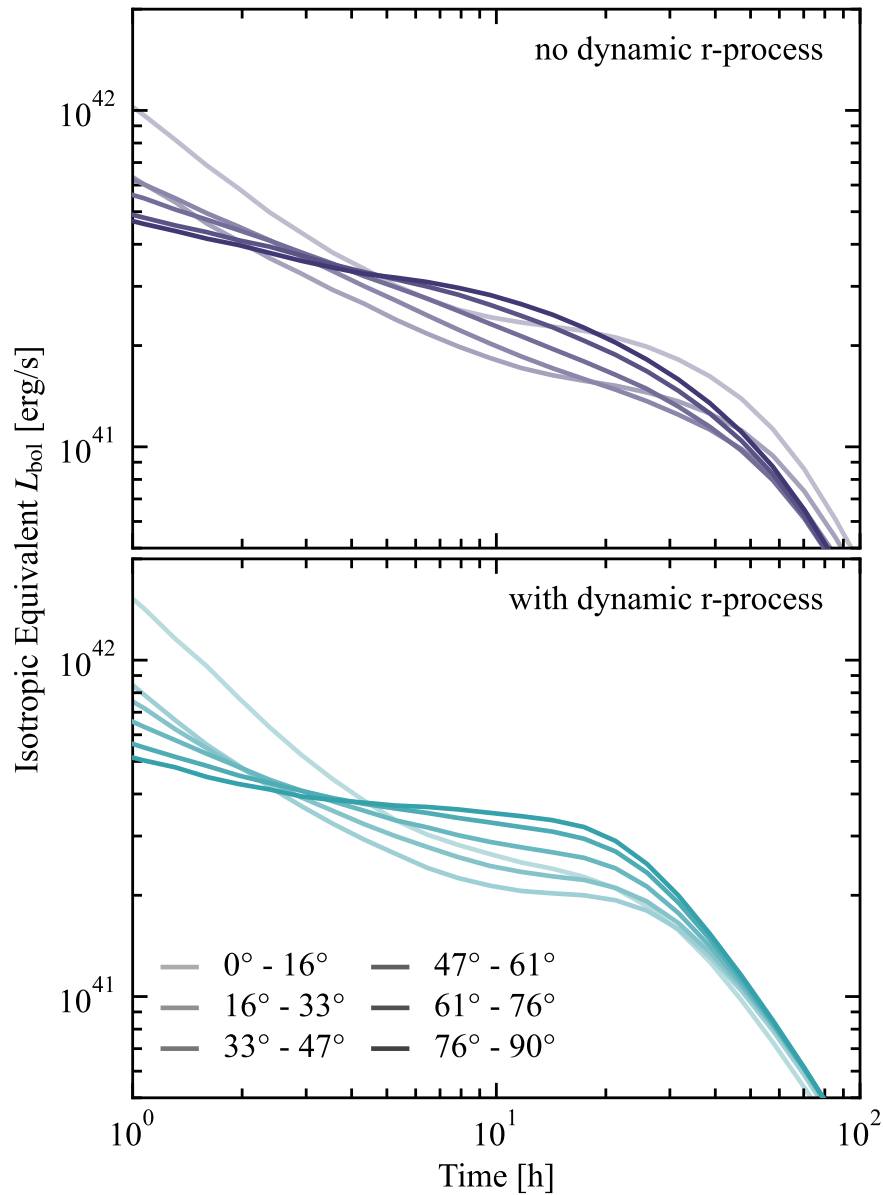
Figure 4.7: Viewing angle-dependent isotropic equivalent bolometric light curves for models without (top) and with (bottom) r-process heating during the hydrodynamic calculations. Both models include r-process heating in the light curve calculation. The light curves shown are averages for observers within the given angular ranges. Polar angles are shown in lighter colors, and equatorial angles in darker ones. For clarity, only Southern hemisphere viewing angles are shown; angles are measured from the South pole. Polar angles are brighter in the first few hours, after which equatorial viewing angles are brighter by a factor of ~ 2. The latter is likely due to the prolate wind structure.
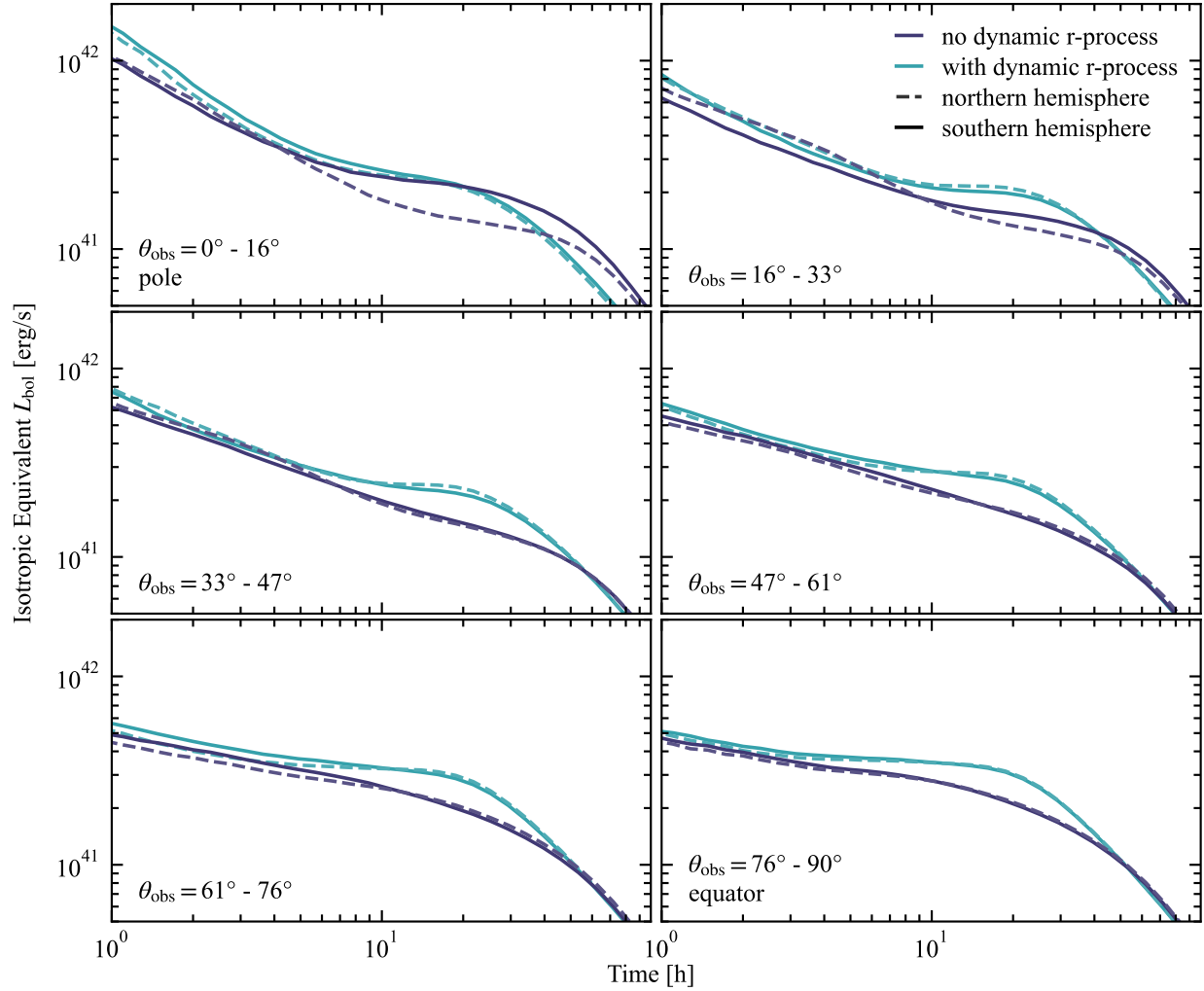
Figure 4.8: Isotropic equivalent bolometric luminosity light curves, as in Fig. 4.7, but each panel shows a different viewing angle range. Light curves from the model with (without) dynamic r-process are shown in teal (purple). Southern hemisphere light curves ($\theta_S = 180° - \theta_N$) are shown as solid lines; Northern hemisphere light curves are dashed. The main effect of dynamic r-process heating is to brighten and flatten the light curve from 0.3–3 d. The light curve also declines somewhat more quickly, most noticeably near the poles. This is due to the change in characteristic velocity from $\sim 0.03\,c$ to $\sim 0.1\,c$. The high-velocity material is minimally affected by r-process heating, so the early time portion of the light curve is nearly the same. The South pole is brighter than the North when there is no dynamical r-process heating, likely due to a non-general feature in the 3D GRMHD simulation. Dynamical r-process heating eliminates this difference.

Our results are shown in Figs 4.7 and 4.8. In both models, polar viewing angles are brighter in the first 2–3 h. Afterwards, the equator brightens relative to the polar regions. While there is not much high-velocity mass, it is concentrated around $\sim 15°$ from the poles, and is optically thick in the first few hours of evolution. Photons from a fast-moving photosphere near the pole are Doppler shifted towards the observer, resulting in particularly bright early polar emission. Due to the low density directly on the poles, polar observers are also able to see "deeper" into the ejecta, which can lead them to see a hotter and brighter photosphere on the poles, even if said photosphere is at a lower velocity (Klion et al., 2021b).

The slope of the high-velocity mass distribution, especially at the pole, may be sensitive to detailed physics not included in our calculations (e.g. equation of state, neutrino transport, and the exact magnetic field structure in the remnant disc). However, the general correlation between high-velocity material and bright, early light curves is plausibly robust. At later times, $\gtrsim 1\,\mathrm{d}$, the polar light curves continue to fall more sharply, while the equatorial light curves remain flat. Between 4 h and 1 d, the equatorial light curves are brighter than on the poles due to the prolate structure of the wind. The greater projected surface area on the equator causes equatorial light curves to be somewhat brighter (Darbha & Kasen, 2020; Korobkin et al., 2021).

The effects of dynamical r-process heating are most clearly seen when comparing light curves from the same viewing angles (Fig. 4.8). We also compare light curves in the Northern and Southern hemispheres, which gives an estimate of the minimum uncertainty in the light curves. The setup of the 3D GRHMD simulation is equatorially symmetric, so on average, the light curves from the North and South should be the same. However, in our calculations the South pole is substantially brighter than the North when there is no r-process heating in the axisymmetrized GRHD run. This is likely due to a numerical artefact on the North pole of the 3D GRMHD simulation that disrupts the polar density structure. The South pole light curves are likely somewhat more reliable than those on the North pole. However, the details of the polar structure of the ejecta are, like the high-velocity distribution, sensitive to physics details that we do not explore here. The difference between the hemispheres goes away once r-process heating is included, which suggests that r-process heating may homogenize the wind structure.

The early light curve is unaffected by r-process heating, consistent with the interpretation that it primarily arises from high-velocity material that is too fast to be substantially affected by r-process heating. The primary effect of dynamical r-process heating is to make the light curves brighter but fade more quickly. This is consistent with the scaling between light curve time-scale and ejecta velocity, $t_{\mathrm{peak}} \propto v^{-1/2}$ (Metzger, 2019). On the South pole, the light curves from the models with and without dynamic r-process heating are quite similar from ~ 6–30 h. This may be a coincidence.

Our model with dynamic r-process heating has a bolometric luminosity of $\sim 3 \times 10^{41}$ erg/s at 1 d. We do not directly compare our models with observations of AT 2017gfo since our wind has a mass of only $0.013\,\mathrm{M_\odot}$, differing by at least a factor of 3 from the $> 0.04\,\mathrm{M_\odot}$ of AT 2017gfo (e.g. Villar et al., 2017). The peak luminosity for a kilonova is expected to scale as $L_{\mathrm{peak}} \propto M^{0.35}$, while the peak time (and therefore time-scale) $t_{\mathrm{peak}} \propto M^{1/2}$ (Metzger,
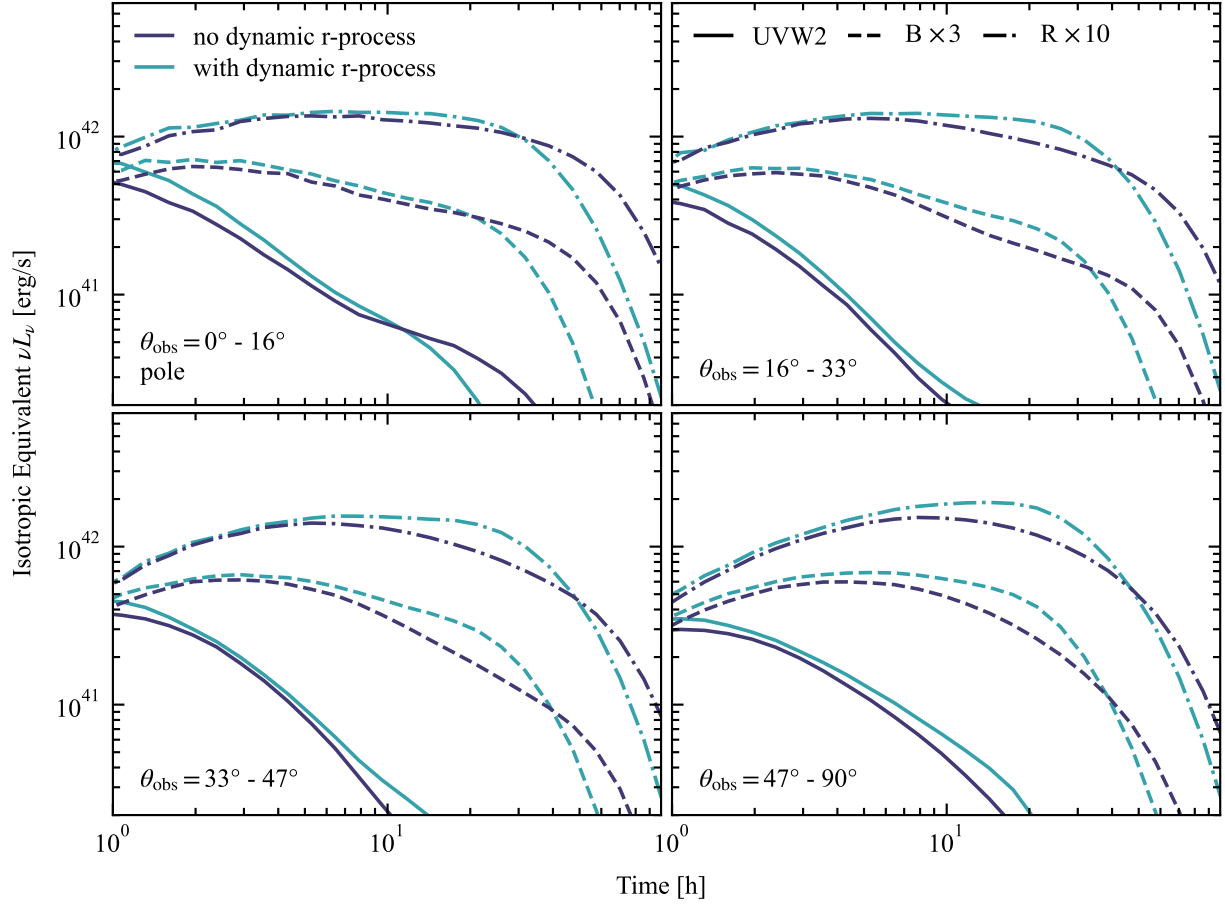
Figure 4.9: *Swift* UVOT W2 (solid), *B* (dashed), and Cousins *R*-band (dashed-dotted) light curves for Southern hemisphere viewing angle ranges (measured from the South pole). Viewing angles > 47° are similar and are therefore shown together. Models without (with) dynamic r-process heating are in purple (teal). As with the bolometric light curves, dynamic r-process only affects the light curves after 8 h. When dynamic r-process is included, the light curves are brighter and fade more quickly. Equatorial *B* and *R*-band light curves peak later than polar viewing angles, which have flatter light curves in the redder bands.

2019). Applying this correction brings our light curves to the rough energy and time-scale of AT 2017gfo (4×10^{41} erg/s at 1.7 d), though the light curves have a different shape before $\lesssim 1$ d. There is somewhat better agreement between the observations and the model with r-process heating. That is likely because the characteristic velocity of our r-process wind is higher, more consistent with the inferred velocity of AT 2017gfo.

The broadband light curves follow similar trends to the bolometric. We show *Swift* UVOT W2 (average wavelength 193 nm), Johnson *B* (442 nm), and Cousins *R*-band (635 nm) light
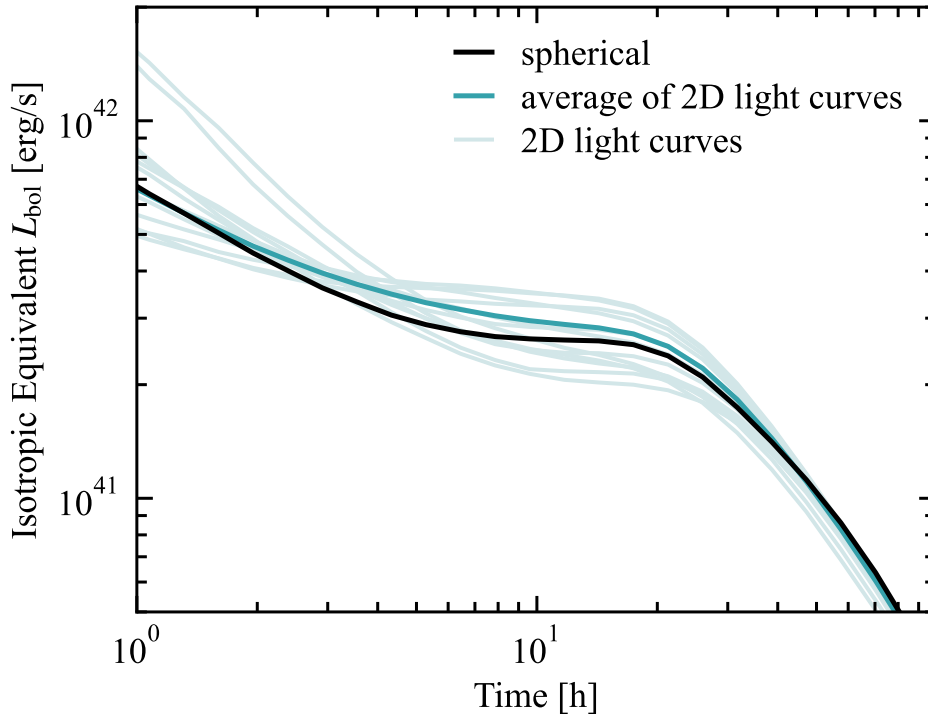
Figure 4.10: Comparison of spherically averaged light curves for the model with dynamic r-process heating. The black line shows the light curve obtained by spherically averaging the SEDONA model before the radiation transport. At a few hours, this is 20 per cent dimmer than the average of the 2D light curves shown in Fig. 4.7. Light curves from individual viewing angles are shown in light teal, while the average is in dark teal.

curves in Fig. 4.9. Across all bands, the polar angles are brighter than equatorial ones in the first few hours. Subsequently, the equatorial light curves brighten, while the emission near the pole remains constant or starts to fade. On time-scales of a day, equatorial emission is both brighter and bluer than the poles. Similar to the bolometric light curves, the dynamical effects of r-process heating are only apparent at later times when the photosphere reaches the slower (and therefore more affected) velocities. The difference in the evolution time-scales between models with and without r-process heating is more apparent in the band light curves, especially near the pole. The colors predicted here could also change noticeably with the inclusion of more realistic r-process opacities. Line blanketing, particularly in the blue and ultraviolet, is expected. This would make the UVOT W2 and B-band light curves dimmer, while causing more re-processed emission in the red and infrared. However, opacities at early ($\lesssim 1\,\mathrm{d}$) phases of kilonova evolution may be lower than they are later (Banerjee et al., 2020). The effects of realistic opacities on the colors of these models will be an interesting direction for future work.

If we spherically average the ejecta before the Sedona calculation, we find that the light curve obtained is similar to the average of all viewing angles in the 2D calculation. We make the comparison between the spherical, averaged, and individual 2D light curves in Fig. 4.10. For most viewing angles, the light curves have a similar shape, mostly unaffected by the asymmetry of the ejecta. Off of the poles, the light curve is well-modeled by a spherical average. The asymmetry only matters for the polar viewing angles.

4.4 Discussion & Conclusion

The kinematic effects of r-process heating on a kilonova disc wind can be substantial, accelerating the ejecta from a mass-weighted median velocity of $0.06\,c$ to $0.12\,c$ (Fig. 4.5). The faster wind leads to a transient that evolves more quickly and is brighter and bluer on ∼day time-scales (Figs 4.7 and 4.9). A factor-of-two increase in ejecta velocity will lead to a transient that evolves 1.4 times as quickly. Quantitative predictions of kilonova light curves from disc wind simulations should therefore account for the kinematic effects of r-process heating. The wind velocities from these simulations are sensitive to the initial magnetic field configuration. A toroidal configuration and/or a weaker field than the one used in F19 would produce a slower wind, which would experience a larger change in velocity when heated by the r-process.

The early ($\lesssim 4\,\mathrm{h}$) kilonova light curve is primarily produced by the high-velocity ($\gtrsim 0.3\,c$) ejecta, which is concentrated near the poles. The poles are especially bright early on because the photosphere is at high velocity, and the emission is Doppler shifted towards the polar observers (Kasliwal et al., 2017). The high-velocity portion of the ejecta is mostly unaffected by r-process heating, so the early light curves are similar to one another in the models with and without dynamic r-process heating. This portion of the light curve is likely sensitive to the physics in the underlying 3D GRMHD simulation (e.g. neutrino interactions, realistic magnetic field structure, and equation of state). Additionally, the high-velocity tail is likely to be disrupted as it runs into the dynamical ejecta from the merger. Joint modeling of the dynamical and secular ejecta is likely necessary to predict the early light curves.

After the first few hours, equatorial viewing angles become brighter than polar ones due to the prolate structure of the ejecta, which has a larger projected surface area at the equator. As the photosphere recedes into the slower portions of the ejecta, the kinematic effects of r-process heating become apparent. The light curves from the model with dynamical r-process heating are flatter at ~ 0.5–$1.5\,\mathrm{d}$, and also fade earlier. The faster evolution is particularly apparent near the poles.

The difference we observe in the low-velocity distributions with and without r-process heating contrasts with the results of Kawaguchi et al. (2021). They find that the effects of including r-process heating during the hydrodynamic phase of the calculation are very small. As they discuss, this is likely because they do not account for heating that occurs before the wind enters their simulation grid. In their simulations, most material enters at $\gtrsim 2\,\mathrm{s}$, which is longer than the $1\,\mathrm{s}$ time-scale on which most of the r-process heating occurs. In

our calculations, we modify the boundary conditions of the GRHD simulation to account for r-process heating that occurred before injection, thus capturing in an approximate way the dominant dynamical effects of r-process heating.

Our formalism for pre-injection heating roughly captures the conversion of r-process heating to kinetic energy. However, we are unable to include all of the effects of the early heating. In particular, we do not account for the meridional expansion of the gas, which may be significant, particularly at the poles. The propagation of the jet leaves a polar cavity that may be filled by hot material from the wind. In our simulation with r-process heating, the evacuated region survives. It is possible that the self-consistent inclusion of r-process heating at all times would cause the gas to expand more and fill in the polar region. There is also the competing effect of magnetic fields, which resist polar expansion of the ejecta. A 3D GRMHD accretion disc simulation that includes r-process heating is necessary to understand the interaction between magnetic fields and hot gas in the polar region.

The low-velocity distribution in the 2D GRHD simulation with r-process heating is truncated at $0.07\,c$, which may not be physical. In the simulation of F19, there is material that is marginally bound and reaches a maximum radius $< r_b = 2 \times 10^4\,$km. With r-process heating, some of this material would have become unbound and crossed r_b. Because we apply a cut at r_b, this material is excluded, even though it would have formed part of the wind. This may artificially truncate the velocity distribution, making the late-time light curve fall off more quickly than it would otherwise.

The degree to which the low-velocity distribution is truncated will also have implications for kilonova spectra. Kasen et al. (2015) found that slowly-moving ($\sim 0.03\,c$) ejecta have resolved absorption lines that could possibly be used to identify elements in kilonova ejecta. Faster (0.1–$0.3\,c$) ejecta have broadened lines that make such identification more difficult (Kasen et al., 2013).

This underscores the importance of self-consistently including r-process heating within the accretion disc simulations themselves. Accurately capturing the first few seconds of heating and their kinematic effects is critical for predicting both the radial velocity and angular distributions of the wind.

In our radiation transport calculations, we adopt a grey opacity. Kilonova ejecta are composed of dozens of elements with large, highly frequency-dependent opacities. Our qualitative light curve results are likely robust to changes to the opacity (e.g. faster ejecta leading to faster light curve evolution). The effects of more realistic opacities are none the less an important direction for future investigations. The majority of r-process products have particularly high opacities in the ultraviolet and blue portions of the spectrum, so our grey prescription may lead us to overestimate the UV and blue emission. However, at very early times ($< 0.5\,$d), the opacity may be suppressed due to the high temperatures and ionization fractions (Banerjee et al., 2020; Klion et al., 2021b).

The composition of disc winds is expected to vary, depending on a number of factors including the life time of a central (hyper-massive) neutron star (Lippuner et al., 2017). A single event can produce material with a wide range of Y_e, and the distribution can be both radially and meridionally-stratified (e.g. Just et al., 2015; Siegel & Metzger, 2017; Fernández

et al., 2019). The overlay of distinct ejecta components with different opacities may have interesting consequences for viewing angle-dependent kilonova light curves (Korobkin et al., 2021).

Chapter 5

Conclusions

GW170817 was a watershed event that confirmed much of the basic theory surrounding neutron star mergers and kilonovae. The next several kilonovae will be important for understanding the intrinsic variation in NS-NS and NS-BH systems. Theoretical modeling will be necessary to connect the measured observational properties with physical models.

Time-dependent radiation transport simulations will be an important tool in these analyses. Sedona is a C++ code that adopts a Monte Carlo technique for solving transport problems. Since photon absorption and scattering are probabilistic processes, Monte Carlo methods allow us to explicitly sample photon trajectories. A large number of photon packets are required to reduce the statistical noise in estimates of the radiation field and emission. The problem therefore lends itself well to parallelization. In Sedona's current approach, each parallel process (MPI rank) essentially runs an independent copy of the radiation transport step, and then the results are averaged. The transport step itself parallelizes very well, showing near-ideal behavior in both strong and weak scaling tests (Section 2.2).

The calculation also requires updating the opacity and temperature of the gas. That work is divided between the ranks and then communicated. This substantially increases the costs on large spatio-frequency grids. There is a point of diminishing returns where additional nodes do not improve performance (Fig. 2.3). In the case of the 2D toy Type Ia problem used in Chapter 2, that occurs around 8 nodes (16 ranks). In problems with smaller grids, this point is likely at a larger number of ranks, allowing for the problem to efficiently parallelize on larger pools of resources.

For large grids, the parallelization strategy taken by Sedona may not be ideal. There is a version of Sedona that is built on top of the adaptive mesh refinement framework AMReX (SedonaEx) (Zhang et al., 2019). In SedonaEx, each rank is only responsible for a portion of the grid, and particles are communicated between ranks. The main limitation of SedonaEx is that it does not have all of the physics implemented in Sedona. Porting over the physics from Sedona or, even better, sharing the physics code base between the two codes, would make SedonaEx more accessible.

Sedona is now better-equipped to run on the latest supercomputers. I have implemented checkpointing, which will allow for runs to be stopped and restarted. This is useful for

debugging purposes, as well as for ensuring that computing time is not wasted if a job crashes or times out.

Particle-based output is very useful for understanding the spectral and light curve features of 2D and 3D models. It can be used to map between the ejecta and portions of the emission. Currently, this is done through a C++ program built on top of the SEDONA code base. This feature would be more usable if a user could interface with it via a Jupyter notebook.

Given the increasing prevalence of GPUs, exploring algorithmic changes to SEDONA could be important. SEDONA's current propagation algorithm is likely ill-suited to porting to GPUs, and it may be advantageous to switch to an event-based algorithm that has less branching.

I have used the improvements described in Chapter 2 to study effects of asymmetry in neutron star mergers on kilonovae. Chapter 3 discusses the effects of jet-ejecta interaction on kilonovae. Shock-heating of the ejecta by the jet has a minimal effect on the light curves after 1 h. Particularly if there is no delay between the launch of the ejecta and the jet engine turn-on, it is difficult to thermalize enough energy to outshine r-process heating on longer time-scales. The jet, however, can significantly affect the density structure of the radioactive material, leading to viewing-angle dependence in the light curve. The emission from the pole is brighter because the jet evacuates a cavity in the ejecta. The photosphere is at a smaller radius and a hotter temperature, which will make the polar light curves brighter and bluer.

The extent to which the jet will evacuate the polar region is uncertain. The axisymmetric hydrodynamic simulations that underlie the light curve calculations may have artificially underdense polar regions. A 3D GRMHD simulation of jet-ejecta interaction that included r-process heating would be necessary to understand the actual polar density structure. 3D instabilities can disrupt jet propagation or lead to mixing between the jet and the surrounding cocoon. R-process heating may also cause hot gas to expand into the polar region.

In Chapter 4, r-process heating does not affect the structure of the polar region. Regardless of whether dynamical r-process heating is considered, the polar region is under-dense, though it is not as evacuated as in the models of Chapter 3. We cannot account for meridional expansion that occurs interior to the inner boundary of the 2D GRHD calculation (and therefore before injection onto our domain). The inclusion of r-process heating in the 3D GRMHD phase of the calculation may cause the polar region to fill in more.

The early ($\lesssim 4$ h) light curves are brighter on the pole than on the equator. This appears to be due to the blueshifted emission from the high-velocity tail of the wind at the poles, rather than from the ability of an observer to see "deeper" into the ejecta (as in Chapter 3). In either case, the early light curves are likely to be brighter and bluer to an observer near the pole than to one near the equator. Due to the overall prolate structure of the disc wind in Chapter 4, the equator is brighter than the pole after the high-velocity material becomes optically thin (after ~ 4 h). The first hours of the light curve may be particularly interesting, especially if there is an early break in the light curves, as seen in e.g. Fig. 4.7. This underscores the utility of efficient gravitational wave trigger follow-up and rapid kilonova discovery. Wide-field ultraviolet instruments like ULTRASAT would be a particularly exciting addition to the arsenal of high-cadence transient searches like the Zwicky Transient

Facility and the forthcoming Rubin Observatory.

Bibliography

Abbott B. P., et al., 2009, Reports on Progress in Physics, 72, 076901

Abbott B. P., et al., 2017a, Phys. Rev. Lett., 119, 161101

Abbott B. P., et al., 2017b, ApJ, 848, L12

Abbott B. P., et al., 2017c, ApJ, 848, L13

Abbott B. P., et al., 2019, Physical Review X, 9, 031040

Abbott B. P., et al., 2020, ApJ, 892, L3

Abbott R., et al., 2021, Physical Review X, 11, 021053

Aizenman M., Smeyers P., Weigert A., 1977, A&A, 58, 41

Arcavi I., et al., 2017, Nature, 551, 64

Banerjee S., Tanaka M., Kawaguchi K., Kato D., Gaigalas G., 2020, ApJ, 901, 29

Barnes J., Kasen D., 2013, ApJ, 775, 18

Barnes J., Kasen D., Wu M.-R., Martínez-Pinedo G., 2016, ApJ, 829, 110

Bauswein A., Goriely S., Janka H. T., 2013, ApJ, 773, 78

Beck P. G., et al., 2012, Nature, 481, 55

Berger E., 2014, ARA&A, 52, 43

Bloom J. S., Kulkarni S. R., Djorgovski S. G., 2002, AJ, 123, 1111

Bloom J. S., et al., 2006, ApJ, 638, 354

Bovard L., Martin D., Guercilena F., Arcones A., Rezzolla L., Korobkin O., 2017, Phys. Rev. D, 96, 124005

Bromberg O., Singh C. B., Davelaar J., Philippov A. A., 2019, ApJ, 884, 39

Brown T. M., Christensen-Dalsgaard J., Dziembowski W. A., Goode P., Gough D. O., Morrow C. A., 1989, ApJ, 343, 526

Brun A. S., Palacios A., 2009, ApJ, 702, 1078

Cantiello M., Mankovich C., Bildsten L., Christensen-Dalsgaard J., Paxton B., 2014, ApJ, 788, 93

Chornock R., et al., 2017, ApJ, 848, L19

Christensen-Dalsgaard J., 2014, Lecture Notes on Stellar Oscillations, `http://astro.phys.au.dk/~jcd/oscilnotes/`

Christie I. M., Lalakos A., Tchekhovskoy A., Fernández R., Foucart F., Quataert E., Kasen D., 2019, MNRAS, 490, 4811

Coulter D. A., et al., 2017, Science, 358, 1556

Cowperthwaite P. S., et al., 2017, ApJ, 848, L17

Darbha S., Kasen D., 2020, ApJ, 897, 150

Darbha S., Kasen D., Foucart F., Price D. J., 2021, ApJ, 915, 69

Deheuvels S., et al., 2014, A&A, 564, A27

Dere K. P., Del Zanna G., Young P. R., Landi E., Sutherland R. S., 2019, ApJS, 241, 22

Di Mauro M. P., et al., 2016, ApJ, 817, 65

Dietrich T., Ujevic M., Tichy W., Bernuzzi S., Brügmann B., 2017, Phys. Rev. D, 95, 024029

Drout M. R., et al., 2017, Science, 358, 1570

Duffell P. C., MacFadyen A. I., 2011, ApJS, 197, 15

Duffell P. C., MacFadyen A. I., 2013, ApJ, 775, 87

Duffell P. C., Quataert E., Kasen D., Klion H., 2018, ApJ, 866, 3

van Eerten H. J., MacFadyen A. I., 2011, ApJ, 733, L37

van Eerten H., Zhang W., MacFadyen A., 2010, ApJ, 722, 235

Eggenberger P., Meynet G., Maeder A., Hirschi R., Charbonnel C., Talon S., Ekström S., 2008, Ap&SS, 316, 43

Eggenberger P., Montalbán J., Miglio A., 2012, A&A, 544, L4

Eichler D., Livio M., Piran T., Schramm D. N., 1989, Nature, 340, 126

Fernández R., Metzger B. D., 2013, MNRAS, 435, 502

Fernández R., Quataert E., Schwab J., Kasen D., Rosswog S., 2015, MNRAS, 449, 390

Fernández R., Tchekhovskoy A., Quataert E., Foucart F., Kasen D., 2019, MNRAS, 482, 3373

Fong W., Berger E., 2013, ApJ, 776, 18

Fong W., Berger E., Margutti R., Zauderer B. A., 2015, ApJ, 815, 102

Fontes C. J., Fryer C. L., Hungerford A. L., Wollaeger R. T., Korobkin O., 2020, MNRAS, 493, 4143

Fujibayashi S., Kiuchi K., Nishimura N., Sekiguchi Y., Shibata M., 2018, ApJ, 860, 64

Gammie C. F., McKinney J. C., Tóth G., 2003, ApJ, 589, 444

Goldstein A., et al., 2017, ApJ, 848, L14

Gompertz B. P., et al., 2018, ApJ, 860, 62

Gottlieb O., Nakar E., Piran T., 2018, MNRAS, 473, 576

Gottlieb O., Nakar E., Bromberg O., 2021, MNRAS, 500, 3511

Goupil M. J., Mosser B., Marques J. P., Ouazzani R. M., Belkacem K., Lebreton Y., Samadi R., 2013, A&A, 549, A75

Granot J., De Colle F., Ramirez-Ruiz E., 2018, MNRAS, 481, 2711

Grossman D., Korobkin O., Rosswog S., Piran T., 2014, MNRAS, 439, 757

HDF Group 1997-2021, Hierarchical Data Format (version 5), `https://www.hdfgroup.org/HDF5/`

Hamilton S. P., Evans T. M., 2019, Annals of Nuclear Energy, 128, 236

Hillier D. J., Lanz T., 2001, in Ferland G., Savin D. W., eds, Astronomical Society of the Pacific Conference Series Vol. 247, Spectroscopic Challenges of Photoionized Plasmas. p. 343

Hotokezaka K., Kiuchi K., Kyutoku K., Okawa H., Sekiguchi Y.-i., Shibata M., Taniguchi K., 2013a, Phys. Rev. D, 87, 024001

Hotokezaka K., Kiuchi K., Kyutoku K., Muranushi T., Sekiguchi Y.-i., Shibata M., Taniguchi K., 2013b, Phys. Rev. D, 88, 044026

Huber D., et al., 2013, Science, 342, 331

Just O., Bauswein A., Ardevol Pulpillo R., Goriely S., Janka H. T., 2015, MNRAS, 448, 541

Kallinger T., et al., 2010, A&A, 522, A1

Karp A. H., Lasher G., Chan K. L., Salpeter E. E., 1977, ApJ, 214, 161

Kasen D., Thomas R. C., Nugent P., 2006, ApJ, 651, 366

Kasen D., Badnell N. R., Barnes J., 2013, ApJ, 774, 25

Kasen D., Fernández R., Metzger B. D., 2015, MNRAS, 450, 1777

Kasen D., Metzger B., Barnes J., Quataert E., Ramirez-Ruiz E., 2017, Nature, 551, 80

Kasliwal M. M., et al., 2017, Science, 358, 1559

Kawaguchi K., Shibata M., Tanaka M., 2020, ApJ, 889, 171

Kawaguchi K., Fujibayashi S., Shibata M., Tanaka M., Wanajo S., 2021, ApJ, 913, 100

Kissin Y., Thompson C., 2015, ApJ, 808, 35

Klion H., Quataert E., 2017, MNRAS, 464, L16

Klion H., Tchekhovskoy A., Kasen D., Kathirgamaraju A., Quataert E., Fernández R., 2021a, arXiv e-prints, p. arXiv:2108.04251

Klion H., Duffell P. C., Kasen D., Quataert E., 2021b, MNRAS, 502, 865

Korobkin O., et al., 2021, ApJ, 910, 116

Kramida A., Yu. Ralchenko Reader J., and NIST ASD Team 2019, NIST Atomic Spectra Database (ver. 5.7.1), [Online]. National Institute of Standards and Technology, Gaithersburg, MD. Available at: `https://physics.nist.gov/asd` [2020, August 12].

Lattimer J. M., Mackie F., Ravenhall D. G., Schramm D. N., 1977, ApJ, 213, 225

Lehner L., Liebling S. L., Palenzuela C., Caballero O. L., O'Connor E., Anderson M., Neilsen D., 2016, Classical and Quantum Gravity, 33, 184002

Leibler C. N., Berger E., 2010, ApJ, 725, 1202

Li L.-X., Paczyński B., 1998, ApJ, 507, L59

Lippuner J., Roberts L. F., 2015, ApJ, 815, 82

Lippuner J., Fernández R., Roberts L. F., Foucart F., Kasen D., Metzger B. D., Ott C. D., 2017, MNRAS, 472, 904

Margutti R., et al., 2018, ApJ, 856, L18

Metzger B. D., 2019, Living Reviews in Relativity, 23, 1

Metzger B. D., Berger E., 2012, ApJ, 746, 48

Metzger B. D., Piro A. L., Quataert E., 2009, MNRAS, 396, 304

Metzger B. D., et al., 2010, MNRAS, 406, 2650

Miller J. M., et al., 2019, Phys. Rev. D, 100, 023008

Mooley K. P., et al., 2018a, Nature, 561, 355

Mooley K. P., et al., 2018b, ApJ, 868, L11

Mosser B., et al., 2010, A&A, 517, A22

Mosser B., et al., 2012, A&A, 548, A10

Nagakura H., Hotokezaka K., Sekiguchi Y., Shibata M., Ioka K., 2014, ApJ, 784, L28

Narayan R., Paczynski B., Piran T., 1992, ApJ, 395, L83

Narayan R., Piran T., Kumar P., 2001, ApJ, 557, 949

Nativi L., Bulla M., Rosswog S., Lundman C., Kowal G., Gizzi D., Lamb G. P., Perego A., 2021, MNRAS, 500, 1772

Noble S. C., Gammie C. F., McKinney J. C., Del Zanna L., 2006, ApJ, 641, 626

Oechslin R., Janka H. T., Marek A., 2007, A&A, 467, 395

Paczynski B., 1986, ApJ, 308, L43

Paxton B., et al., 2013, ApJS, 208, 4

Perego A., Radice D., Bernuzzi S., 2017, ApJ, 850, L37

Piro A. L., Kollmeier J. A., 2018, ApJ, 855, 103

Popham R., Woosley S. E., Fryer C., 1999, ApJ, 518, 356

Radice D., Perego A., Hotokezaka K., Fromm S. A., Bernuzzi S., Roberts L. F., 2018, ApJ, 869, 130

Rosswog S., Korobkin O., Arcones A., Thielemann F. K., Piran T., 2014, MNRAS, 439, 744

Roth N., Kasen D., 2015, ApJS, 217, 9

Sari R., Piran T., Narayan R., 1998, ApJ, 497, L17

Savchenko V., et al., 2017, ApJ, 848, L15

Shibata M., Fujibayashi S., Hotokezaka K., Kiuchi K., Kyutoku K., Sekiguchi Y., Tanaka M., 2017, Phys. Rev. D, 96, 123012

Siegel D. M., Metzger B. D., 2017, Phys. Rev. Lett., 119, 231102

Siegel D. M., Metzger B. D., 2018, ApJ, 858, 52

Soares-Santos M., et al., 2017, ApJ, 848, L16

Tanaka M., Kato D., Gaigalas G., Kawaguchi K., 2020, MNRAS, 496, 1369

Tchekhovskoy A., 2019, HARMPI: 3D massively parallel general relativictic MHD code (ascl:1912.014)

Townsend R. H. D., Teitler S. A., 2013, MNRAS, 435, 3406

Villar V. A., et al., 2017, ApJ, 851, L21

Wollaeger R. T., et al., 2018, MNRAS, 478, 3298

Wu Y., MacFadyen A., 2019, ApJ, 880, L23

Zhang W., et al., 2019, Journal of Open Source Software, 4, 1370

Appendix A

Example scripts

A.1 Example Run Script for Scaling Tests on Cori

This is an example Slurm script for running timing tests on Cori's Haswell partition. In this example, SEDONA will be run on 64 nodes, with 2 MPI ranks (tasks) per node, one allocated to each CPU on the node. Each rank has 16 OpenMP threads, which are each bound to physical cores. The output will include thread affinity information (generated by `check-hybrid.intel.cori`), and a total wallclock time for the `srun` command. This is a useful check that the timers used by SEDONA are returning values that make sense.

It assumes that the module `PrgEnv-intel` is loaded, and that the binary has been compiled with the Intel compilers. If using the GNU compiler, different OpenMP options will be required to achieve the same thread layout. The `check-hybrid.gnu.cori` program can be used to check thread and rank affinity when using the GNU environment.

```
#!/bin/bash

#SBATCH --job-name=sedona
#SBATCH --account=m2864
#SBATCH --nodes=64
#SBATCH --qos=debug
#SBATCH --time=00:30:00
#SBATCH --tasks-per-node=2
#SBATCH --cpus-per-task=32
#SBATCH --constraint=haswell
#SBATCH --license=SCRATCH
#SBATCH --output="n64_r02_t16_%A.out"
#SBATCH --error="n64_r02_t16_%A.out"

export HDF5_USE_FILE_LOCKING=FALSE
export OMP_PROC_BIND=true
```

```
export OMP_PLACES=threads
export OMP_NUM_THREADS=16

module load cray-hdf5
module load gsl

cd $SLURM_SUBMIT_DIR

srun check-hybrid.intel.cori

SECONDS=0
srun ./gomc_intel

echo "Ran in $SECONDS seconds"
```

Appendix B

A Diagnostic for Localizing Red Giant Differential Rotation

This chapter is modified from Klion & Quataert (2017) with permission from the co-author.

I present a simple diagnostic that can be used to constrain the location of the differential rotation in red giants with measured mixed mode rotational splittings. Specifically, in red giants with radii $\sim 4\,\mathrm{R_\odot}$, the splittings of p-dominated modes (sound wave dominated) relative to those of g-dominated modes (internal gravity wave dominated) are sensitive to how much of the differential rotation resides in the outer convection zone versus the radiative interior of the red giant. An independently measured surface rotation rate significantly aids breaking degeneracies in interpreting the measured splittings. I apply these results to existing observations of red giants, particularly those of Kepler-56, and find that most of the differential rotation resides in the radiative region rather than in the convection zone. This conclusion is consistent with results in the literature from rotational inversions, but my results are insensitive to some of the uncertainties in the inversion process and can be readily applied to large samples of red giants with even a modest number of measured rotational splittings. I argue that differential rotation in the radiative interior strongly suggests that angular momentum transport in red giants is dominated by local fluid instabilities rather than large-scale magnetic stresses.

B.1   Background

Asteroseismology with the *Kepler* satellite has allowed measurements of red giant differential rotation (Beck et al., 2012). In red giants, the frequencies of core gravity (g-) modes and envelope pressure (p-) modes overlap, producing what are known as mixed modes (e.g. Aizenman et al., 1977). Broadly, these can be classified as either gravity- or pressure-dominated (hereafter g-m and p-m, respectively), depending on where in the star they have greater amplitude. Differences between the splittings of p-m and g-m modes show that

red giant cores rotate faster than their envelopes (Beck et al., 2012; Mosser et al., 2012). However, inversions of mixed mode splittings have difficulty constraining the exact location of the differential rotation (Deheuvels et al., 2014; Di Mauro et al., 2016).

The measured ratios between the core and envelope angular velocities in red giants are generally smaller than predicted by stellar evolution calculations that employ standard models of angular momentum transport (Cantiello et al., 2014). Angular momentum transport is thought to arise from fluid instabilities and/or torques due to the large-scale magnetic fields. Since these effects are intrinsically multi-dimensional, and the time-scale of a star's evolution is much longer than its hydrodynamic time-scale, stellar evolution codes must parametrize the (magneto-)hydrodynamic effects of rotation (e.g. Eggenberger et al., 2008; Paxton et al., 2013).

Angular momentum can also be transported by non-rotational instabilities such as convection; the rapid convective mixing time-scales suggest efficient angular momentum transport in convection zones. This is typically taken to imply nearly solid body rotation since convection is assumed to act like a viscosity and eliminate angular velocity gradients. This conclusion is, however, at odds with three-dimensional simulations of rotating convection and analytic arguments, which find significant differential rotation in red giant convection zones (Brun & Palacios, 2009; Kissin & Thompson, 2015). From helioseismology, it is also known that the sun's convection zone is not rotating as a solid body, but rather has differential rotation in latitude as well as in radius (Brown et al., 1989).

In red giants, g-m mode splittings are predominantly sensitive to core rotation, while p-m mode splittings can be affected by the rotation profile of the entire star (e.g. Beck et al., 2012). Our goal in this chapter is to assess how this difference can be used to constrain the location of the differential rotation. Specifically, we compare rotational splittings for two classes of rotation profiles: one where the differential rotation is concentrated just outside of the hydrogen burning shell (e.g. Eggenberger et al., 2012), and another where the differential rotation resides in the convective envelope (e.g. Kissin & Thompson, 2015). We provide simple diagnostics that can be used to distinguish between these different models, complementing more detailed studies based on full inversions.

B.2 Methods

Stellar Evolution and Asteroseismology Calculations

The distribution of red giant masses with observed asteroseismic modes is centered near $1.3\,M_\odot$ (Mosser et al., 2010; Kallinger et al., 2010). We therefore take as our representative model a star with solar metallicity and a zero-age main sequence (ZAMS) mass of $1.33\,M_\odot$. We use version 8118 of the 1D stellar evolution code MESA to evolve it from the pre-main sequence until it has a radius of $10\,R_\odot$ as a red giant.

Due to the uncertainties in stellar angular momentum transport, a self-consistently evolved rotational profile would not necessarily be accurate. We therefore evolve our repre-

sentative model without rotation. This also gives more flexibility in the rotational profiles we can study. The majority of the stellar models and rotation profiles we consider reach at most a few per cent of the surface break-up angular velocity. Modifications to the structure caused by centrifugal forces and rotationally-induced mixing are therefore small.

We calculate the frequencies and eigenfunctions of the dipole mixed modes using the adiabatic version of the pulsation code GYRE, version 4.1 (Townsend & Teitler, 2013).

The frequency of maximum power in the asteroseismic spectrum is ν_{max}. Since the mode amplitudes are peaked near this value, we only consider modes within two radial orders of ν_{max}, that is ones with frequencies in the interval $[\nu_{\mathrm{max}} - 2\Delta\nu,\ \nu_{\mathrm{max}} + 2\Delta\nu]$. This is the typical frequency range of observed modes (e.g. Beck et al., 2012; Di Mauro et al., 2016). $\Delta\nu$ is the large separation, the separation in frequency between p-m modes near ν_{max}. We estimate $\Delta\nu$ and ν_{max} for a given stellar model using the scaling relations calibrated in Mosser et al. (2010).

Model Rotation Profiles

We consider an idealized rotation profile in which there is differential rotation both near the core and in the convection zone. Differential rotation outside the core is natural if angular momentum transport is dominated by local fluid instabilities that must overcome the strongly stabilizing composition gradient at the hydrogen burning shell. By contrast, if transport is dominated by large-scale magnetic fields, solid-body rotation across the composition gradient is plausible and the star's differential rotation may instead be largely contained within the convection zone.

We assume that the inner portion of the star undergoes solid body rotation at the rate Ω_{c}. Some of the differential rotation is concentrated at 1.5 times r_{H}, the outer radius of the hydrogen burning shell. We model this as a step function decrease from Ω_{c} to Ω_{m}, the angular velocity of the middle of the star. Models with a continuous transition between Ω_{c} and Ω_{m} at $1.5r_{\mathrm{H}}$ (e.g. a hyperbolic tangent) yield essentially identical results. The remainder of the star's differential rotation is located in the convection zone, where it follows a power law profile. The full profile can therefore be written as

$$\Omega(r) = \begin{cases} \Omega_{\mathrm{c}} & r \le 1.5r_{\mathrm{H}} \\ \Omega_{\mathrm{m}} & 1.5r_{\mathrm{H}} < r \le r_{\mathrm{rcb}} \ , \\ \Omega_{\mathrm{e}} \left(\frac{R}{r}\right)^{\alpha} & r > r_{\mathrm{rcb}} \end{cases} \tag{B.1}$$

where

$$\alpha = \frac{\log(\Omega_{\mathrm{m}}/\Omega_{\mathrm{e}})}{\log(R/r_{\mathrm{rcb}})} \tag{B.2}$$

is chosen so that $\Omega(r_{\mathrm{rcb}}) = \Omega_{\mathrm{m}}$ and $\Omega(R) = \Omega_{\mathrm{e}}$. The surface angular velocity is Ω_{e}, and r_{rcb} is the radius of the radiative-convective boundary.

We initially focus on two limiting cases. We call the first "convection power law." All of the differential rotation in this model is contained in the convection zone ($\Omega_{\mathrm{m}} = \Omega_{\mathrm{c}}$ in

equation B.1). We will be most interested in the $4\,R_\odot$ model, for which $\Omega_c/\Omega_e \sim 5$–30 implies $\alpha \sim 1.2$–2.5. By comparison, Kissin & Thompson (2015) argue on theoretical grounds that $\alpha \sim 1$–3/2. In our alternative "core step" model, all of the differential rotation is concentrated at $1.5r_H$. In this case, $\Omega_m = \Omega_e$ and $\alpha = 0$. Examples of these profiles for the $4\,R_\odot$ and $10\,R_\odot$ models are shown in Fig. B.1. We also show a profile with $\alpha = 1$ for the $4\,R_\odot$ model.

Calculating Mode Frequency Splittings

In a non-rotating star, asteroseismic modes with the same spherical harmonic ℓ but different m are degenerate in frequency. Rotation introduces a perturbation to the frequency spectrum that causes the modes to split into $(2\ell + 1)$-plets at frequencies

$$\nu_{n\ell m} = \nu_{n\ell} + \delta\nu_{n\ell m} = \nu_{n\ell} + m\frac{\beta_{n\ell}}{2\pi}\int_0^R K_{n\ell}(r)\Omega(r)\,\mathrm{d}r \tag{B.3}$$

where $\nu_{n\ell}$ is the frequency of the unsplit mode. $\beta_{n\ell}$ and $K_{n\ell}(r)$ depend on the mode eigenfunctions (Christensen-Dalsgaard, 2014). $K_{n\ell}$, the rotational kernel, which is normalized to 1 integrated over the star, is greater where the mode has greater amplitude. $\beta_{n\ell} \approx 1/2$ for g-m $\ell = 1$ modes, and approaches 1 for p-m modes.

To maximize the splitting of a g-m mode, the mode should have nearly all of its amplitude within the core such that its rotational kernel integrated in the core is nearly 1. These modes also have $\beta_{n\ell} \approx 1/2$. Therefore the theoretical maximum g-m mode splitting is $\max(\delta\nu_g) = \Omega_c/(4\pi)$ for dipole modes, and we can estimate

$$\Omega_c \approx 4\pi\max(\delta\nu_g). \tag{B.4}$$

Empirically, the mode with the largest rotational splitting is a g-m mode, so the maximal splitting $\max(\delta\nu) = \max(\delta\nu_g)$.

The integrated rotational kernels for the $4\,R_\odot$ and $10\,R_\odot$ models (Fig. B.1) are dominated by contributions from the helium core, hydrogen burning shell, and the outer convection zone, with little contribution from the non-burning radiative hydrogen region. The splittings are insensitive to changes in the location of the differential rotation within the radiative hydrogen region. Our profile with differential rotation concentrated near the hydrogen burning shell is therefore representative of all rotation profiles whose differential rotation is contained within the radiative hydrogen region.

B.3 Results

Fig. B.2 shows frequency splittings for red giants with $R = 4$–$10\,R_\odot$ for modes within two radial orders of ν_{max}. These results are for rotation profiles with $\Omega_c/\Omega_e = 7.89$. The splittings are normalized to the maximum splitting for each radius (i.e. $\delta\nu/\max(\delta\nu)$). As the star
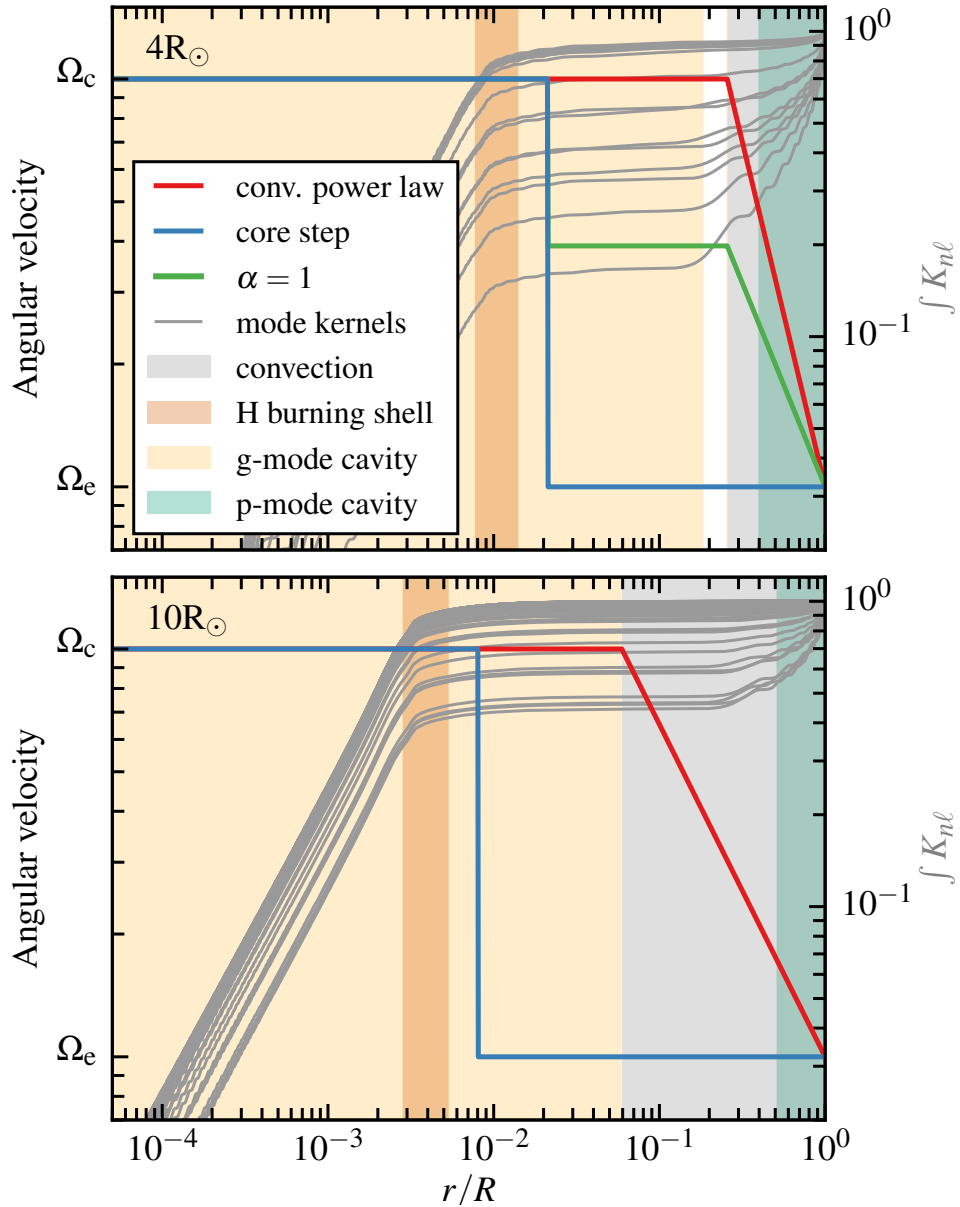
Figure B.1: Integrated rotational kernels (grey lines, right y-axes, see equation B.3) for asteroseismic modes within two radial orders of $\nu_{\max}$ calculated using GYRE for $1.33\,\mathrm{M}_\odot$, $4\,\mathrm{R}_\odot$ (top) and $10\,\mathrm{R}_\odot$ (bottom) red giant models evolved using MESA. The x-axes show the normalized radial coordinate of the stars. The left y-axes correspond to the colored lines, which show three of our model rotational profiles: differential rotation in the convection zone (red), near the hydrogen burning shell (blue), or both (green). The hydrogen burning region, g-mode, and p-mode cavities are shaded in light orange, dark orange, and green, respectively. Convection zones are shaded in grey.
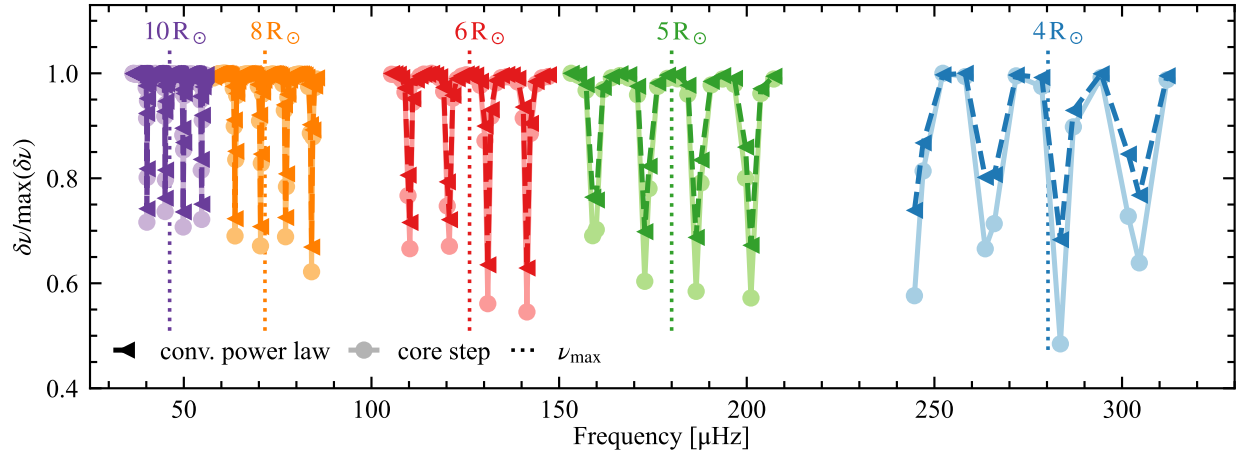
Figure B.2: Rotational mode splittings for models with differential rotation just outside of the hydrogen burning shell (core step, lighter solid lines) and in the convection zone (convection power law, darker dashed lines) at several points along the evolution of a $1.33\,M_\odot$ ZAMS red giant. At each radius, we take $\Omega_c/\Omega_e = 7.89$, and normalize the splittings to the largest splitting at that radius. Each color denotes a different radius. The dotted lines show $\nu_{\max}$ for each model. Only splittings for modes within two radial orders of $\nu_{\max}$ are shown. More evolved stars have smaller values of both $\nu_{\max}$ and $\Delta\nu$. In the less evolved models $(4\,R_\odot)$ the most pressure-dominated p-m modes have splittings that are ~ 50–60 per cent of the maximum splitting, while more evolved models have $\min(\delta\nu/\max(\delta\nu)) \approx 70$ per cent. The different internal rotation profiles also become less distinguishable at larger radius, i.e. later in the star's evolution.

expands, its effective temperature decreases and its luminosity increases, causing $\nu_{\max}$ and $\Delta\nu$ to decrease.

The p-m splittings are smaller than the g-m splittings in all models because the surface rotation rate is significantly less than the core rotation rate in these rotation profiles. The ratio between the p-m and g-m splittings increases from ~ 50–60 per cent in the $4\,R_\odot$ model to ~ 70 per cent in the $10\,R_\odot$ model, indicating that the p-m modes have greater core amplitudes in models with larger radii. In general, higher frequency modes are more p-like. Since $\nu_{\max}$ is smaller in larger models, modes near $\nu_{\max}$ will become less p-like as the star evolves, leading to the smaller difference between the p-m and g-m splittings seen in Fig. B.2 for larger red giants.

For all models shown in Fig. B.2, normalized p-m mode splittings are lower when the radiative hydrogen zone ("core step"), instead of the convective envelope ("conv. power law"), is differentially rotating. Concentrating the differential rotation outside of the hydrogen shell produces a smaller p-m/g-m splitting ratio than nearly any other possible rotation profile. Since there is solid body rotation at Ω_c and Ω_e in the core and envelope, the g-m and p-m

mode splittings are extremized, and their ratio is minimized.

Fig. B.2 shows that the effect of different rotation profiles on the relative p-m and g-m splittings becomes significantly weaker as the star evolves. Since the differential rotation in both profiles shown in Fig. B.2 is outside of the core, the core contribution to the splitting is the same in both profiles. The difference in the splittings arises from the difference in the rotation profile in the convective envelope. In red giants with larger radii, the rotation in the convective envelope contributes less to the overall splittings. As a result, the signature of the location of differential rotation is stronger in less evolved red giants, as noted in Goupil et al. (2013) and Deheuvels et al. (2014).

Fig. B.3 shows the minimum normalized splitting for both of our model rotation profiles for a range of core to envelope rotation ratios Ω_c/Ω_e. The minimum normalized splitting is the smallest value of $\delta\nu/\max(\delta\nu)$ for the range of modes considered (near ν_{max}, see Fig. B.2). The curves in Fig. B.3 all have a two-component structure. Below $\Omega_c/\Omega_e \sim 2$–3, increasing core rotation relative to the envelope causes the minimum normalized splitting to increase, whereas for faster core rotation the opposite occurs. This is because below $\Omega_c/\Omega_e \sim 2$–3, p-m splittings exceed g-m splittings due to the p-m modes' greater values of $\beta_{n\ell}$ (see equation B.3). The vast majority of published red giant asteroseismic observations find g-m splittings that exceed p-m splittings, so we focus on profiles with $\Omega_c/\Omega_e \gtrsim 2$.

As expected, Fig. B.3 shows that the model with differential rotation outside of the hydrogen burning shell has lower minimum normalized splittings than the model with differential rotation in the convection zone. Consistent with Fig. B.2, the $4\,R_\odot$ model has lower minimum normalized splittings and shows a larger difference between the two types of rotation profiles than the $10\,R_\odot$ model. Fig. B.3 also shows a typical observed value of $\min(\delta\nu/\max(\delta\nu))$, calculated from measured splittings of the red giants KIC 5356201 (0.38), KIC 8366239 (0.54), KIC 12008916 (0.35) (Beck et al., 2012), Kepler-56 (0.39) (Huber et al., 2013), and KIC 4448777 (0.41) (Di Mauro et al., 2016). These stars have masses between 1.02–1.49 $M_\odot$ and radii in the range 3.79–5.13 $R_\odot$.

Fig. B.3 highlights an important degeneracy between the location of the differential rotation and Ω_c/Ω_e. For a given measured minimum normalized splitting, there are in general two possible solutions: a smaller value of Ω_c/Ω_e with differential rotation in the core (blue line, core step) or a larger value of Ω_c/Ω_e with differential rotation in the convection zone (red line, convection power law). For instance, Fig. B.3 shows that core differential rotation is generally preferred if $\Omega_c/\Omega_e \sim 7$–20 (similar to the conclusion of Di Mauro et al. 2016) but convection zone differential rotation with a more slowly rotating envelope could also be consistent with the observed splittings of some red giants. Note, however, that $\min(\delta\nu/\max(\delta\nu))$ tends to a constant value as $\Omega_c/\Omega_e \to \infty$. In some cases, this asymptotic value for the profile with differential rotation in the convection zone is so large that this profile is strongly disfavored (Fig. B.3).

The degeneracy between high $\Omega_c/\Omega_e \sim 50$ and lower $\Omega_c/\Omega_e \sim 10$ can be broken if there are enough measured rotational splittings to perform a full inversion. In practice, however, this has been difficult given the finite numbers of available modes (Deheuvels et al., 2014; Di Mauro et al., 2016). Consequently, a measurement of Ω_c/Ω_e is very useful for estimating
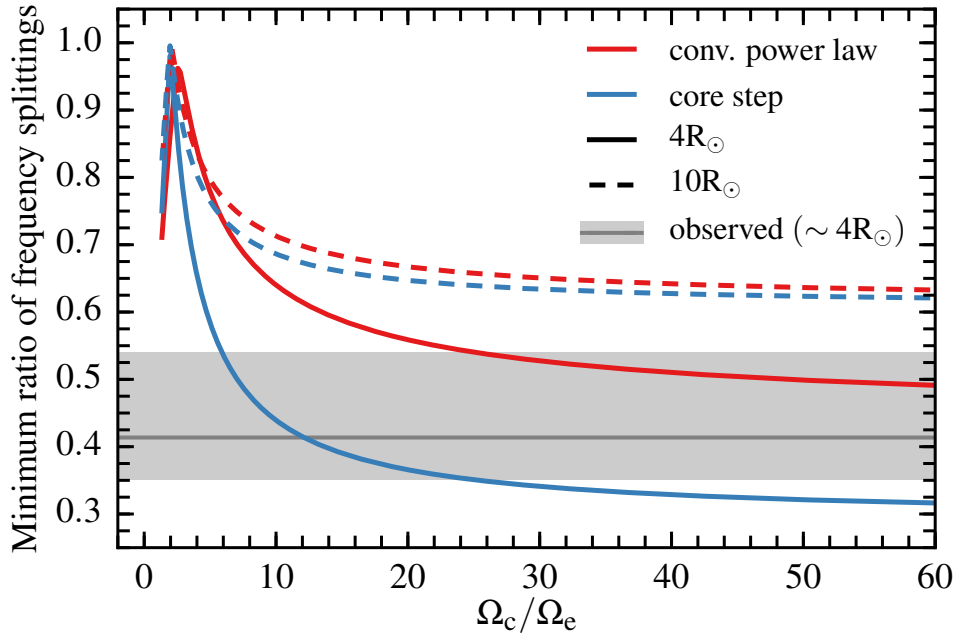
Figure B.3: Comparison of minimum ratios of rotational splittings near ν_{max} for models with differential rotation in the convection zone (red, conv. power law) and outside of the hydrogen burning shell (blue, core step) for a range of ratios of core to surface rotation rates, $\Omega_{\mathrm{c}}/\Omega_{\mathrm{e}}$. For $\Omega_{\mathrm{c}}/\Omega_{\mathrm{e}} \gtrsim 2\text{--}3$ the y-axis represents the smallest value of the ratio of p-m to g-m rotational splittings ($\min(\delta\nu/\max(\delta\nu))$, as in, e.g. Fig. B.2). Values for the $4\,\mathrm{R}_\odot$ ($10\,\mathrm{R}_\odot$) model are shown as solid (dashed) lines. The grey line (shaded region) marks a typical observed value (full range) of $\min(\delta\nu/\max(\delta\nu))$, calculated for five observed $1.02\text{--}1.49\,\mathrm{M}_\odot$ red giants with radii in the range $3.79\text{--}5.13\,\mathrm{R}_\odot$ (see text). Differential rotation in the convection zone is possible for red giants with larger measured $\min(\delta\nu/\max(\delta\nu))$, but splittings for the majority of $\sim 4R_\odot$ red giants are consistent with differential rotation concentrated outside of the core. In general, there is a degeneracy between the location of the differential rotation and $\Omega_{\mathrm{c}}/\Omega_{\mathrm{e}}$, such that the location of the differential rotation can be more robustly constrained if there is an independent measurement of the surface rotation rate, Ω_{e}.

the radial location of the differential rotation, particularly given uncertainties in the stellar parameters and structure which further complicate performing full inversions. Ω_c can be estimated from the splittings of the most core-dominated g-m modes using equation B.4, but an independent measurement of the surface rotation rate, Ω_e, can significantly help pinpoint the interior location of the differential rotation.

B.4   Kepler-56

Huber et al. (2013) measured the dipole mixed modes of Kepler-56, a red giant with a radius of $4.23 \pm 0.15\,R_\odot$ and a mass of $1.32 \pm 0.13\,M_\odot$. They also found a rotational period of $74 \pm 3\,d$, measured from flux variations assumed to be due to starspots. This is nominally more precise than the rotational period of 60–230 d inferred from rotational broadening, $v \sin i = 1.7 \pm 1.0\,\mathrm{km s^{-1}}$, where v is the surface rotational velocity, and $i = 47 \pm 6°$ is the inclination of the star's rotation axis relative to the line of sight, determined from asteroseismology.

Fig. B.4 shows the minimum ratios of rotational splittings for a model with radius $4.23\,R_\odot$. Huber et al. (2013) only report rotational splittings for select modes, so we estimate the remaining splittings as the average splitting of the $m = \pm 1$ modes. For the rotation profile with differential rotation near the core, we also show results for models with radii of $4.08\,R_\odot$ and $4.38\,R_\odot$ corresponding to the observed radius plus or minus the uncertainty. In addition to our two default rotation profiles, we consider two additional rotation profiles with differential rotation both near the core and in the convection zone. For these profiles, we fix α (see equations B.1 and B.2) at 1 and 3/2, bracketing the range of α considered in Kissin & Thompson (2015). We only consider profiles with $\Omega_c \geq \Omega_m \geq \Omega_e$, which means that the models with fixed α require $\Omega_c / \Omega_e \geq 4.4$ and 9.1 for $\alpha = 1$ and 3/2, respectively.

In Kepler-56, the minimum observed ratio between g-m and p-m splittings is

$$\frac{0.482\,\mu\mathrm{Hz}}{0.198\,\mu\mathrm{Hz}} = 0.388 \tag{B.5}$$

and is marked as a grey horizontal line in Fig. B.4. We estimate Ω_c using equation B.4 and compute Ω_c / Ω_e and its uncertainty using both measurements of Ω_e (see Fig. B.4). Fig. B.4 strongly suggests that differential rotation predominantly in the convection zone ("conv. power law") is disfavored. Fig. B.4 also shows that none of the rotation profiles we consider is consistent with the starspot-inferred Ω_c / Ω_e and observed $\min(\delta\nu / \max(\delta\nu))$. The splittings are, however, consistent with the value of Ω_c / Ω_e inferred from $v \sin i$ if the differential rotation is primarily near the core ("core step"). The $v \sin i$ measurement and observed normalized splittings are also consistent with $\alpha \sim 1$, which corresponds to a factor of ~ 4 change in angular velocity across the convection zone.

The minimum ratio between the g-m and p-m splittings is lowest when the differential rotation is in the non-burning hydrogen region, where the rotational kernels are very small. No other rotation profile will have substantially lower minimum normalized splittings than the core step profile (dark blue line in Fig. B.4). The calculated splittings are also largely
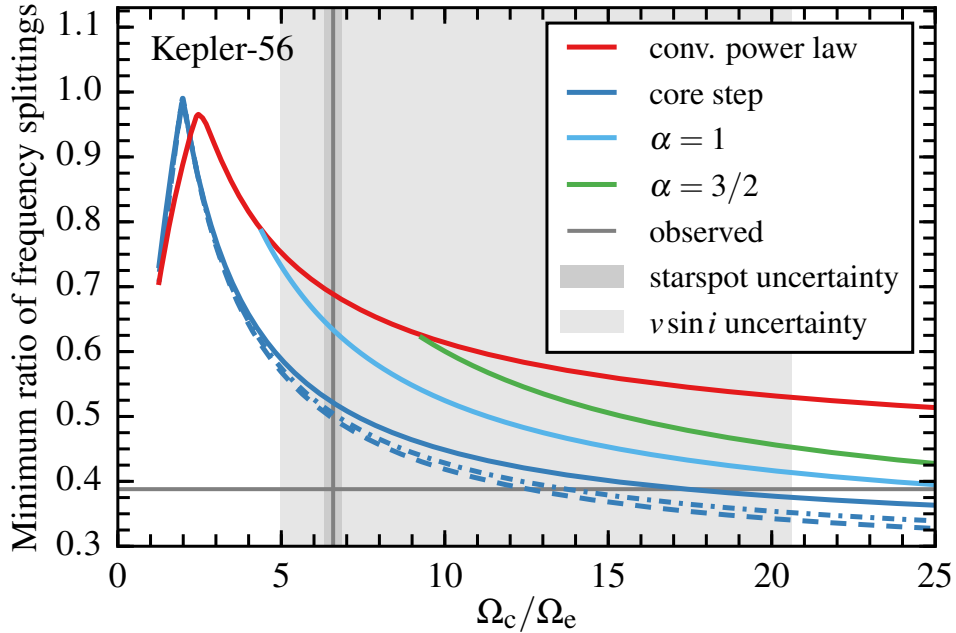
Figure B.4: Minimum ratios of rotational splittings (as in Fig. B.3) for models of the red giant Kepler-56. Results are shown for a range of Ω_c/Ω_e, the ratio of the core to surface rotation rates and four rotation profiles with: differential rotation in the convection zone (conv. power law, red), just outside of the hydrogen burning shell (core step, blue), and in both the convection and radiative zones ($\alpha = 1$ and $\alpha = 3/2$, light blue and green, respectively). In these latter models, $\Omega \propto r^{-\alpha}$ in the convection zone (see equation B.1). The observed minimum normalized splitting for Kepler-56 is marked with a grey horizontal line. The darker (lighter) shaded region shows the uncertainty in the measurement of Ω_c/Ω_e from flux variations due to starspots (rotational line broadening). The value of Ω_c/Ω_e implied by the minimum normalized splitting is inconsistent with the starspot measurement, but consistent with spectroscopic estimates of the surface rotation rate. The convection power law and $\alpha = 3/2$ models (red and green lines) are inconsistent with the data implying that most of the differential rotation cannot be in the convection zone, but must instead primarily reside in the core. 4.08 (dashed), 4.23 (solid), and $4.38\,R_\odot$ (dashed-dotted line) models bracket the measured radii and have little effect on our conclusions.

insensitive to details of the stellar model. Fig. B.4 shows that the model radius has only a small effect on the normalized splittings. We have also varied the mixing length parameter, the stellar mass, and the metallicity. None of these changes significantly affects the minimum splitting ratios shown in Fig. B.4. In all cases, the minimum splitting ratio in Kepler-56 implies that Ω_c/Ω_e is 2–3 times the value given by the starspot measurements. Since the core rotation rate is well-known from the maximum g-m splittings, this suggests that the starspot measurement of the surface rotation rate is too large by a factor of ~ 2.

B.5   Discussion

Measurements of the rotational splittings of mixed modes in red giants show that the core is rotating much faster than the envelope but have not definitively determined where in the interior most of the differential rotation is localized (Beck et al., 2012; Deheuvels et al., 2014; Di Mauro et al., 2016). Constraining the location of the differential rotation within the interior would strongly constrain the dominant mechanisms of angular momentum transport in red giants. We have shown that the ratio between the minimum p-m rotational splitting and the maximum g-m splitting is a simple yet effective diagnostic for localizing the differential rotation in the interiors of red giants (Figs. B.2 and B.3). It has the most discriminating power for modestly evolved red giants with radii $\sim 4\,R_\odot$. A non-asteroseismic measurement of the surface rotation is particularly valuable in applying this diagnostic because it anchors the contribution of the surface rotation to the splittings of the p-m modes (Fig. B.4). The technique considered here allows for the rapid characterization of differential rotation in large numbers of red giants without the need for many measured frequency splittings or detailed rotational inversions, which are uncertain because of the small numbers of modes available in red giants.

We find that for a typical ~ 1–$1.5\,M_\odot$ red giant with a radius of ~ 4–$5\,R_\odot$, observed frequency splittings are best explained by differential rotation just outside of the hydrogen burning shell and a core that rotates $\sim 5 - 20$ times faster than the envelope (Fig. B.3). This is consistent with prior studies that have calculated red giant rotation profiles by inverting the frequency splittings (e.g. Deheuvels et al., 2014; Di Mauro et al., 2016). A much more slowly rotating envelope with substantial differential rotation in the convection zone is disfavored in general but can produce measured p-m to g-m splitting ratios at the upper end of the measured values (Fig. B.3). Surface rotation measurements would break the degeneracy between Ω_c/Ω_e and the location of the differential rotation.

Kepler-56 has estimated surface rotation rates from both rotational line broadening and spots. As Fig. B.4 shows, the measured $v \sin i$ rules out most of the differential rotational residing in the convection zone; differential rotation near the H burning shell with $\Omega_c/\Omega_e \sim$ 13–17 is preferred. Alternatively, differential rotation can exist both near the core and in the convection zone, but the net change in rotation rate across the convection zone is constrained to be less than a factor of ~ 4 (corresponding to $\alpha \lesssim 1$ in Fig. B.4). Our inferred surface rotation period for Kepler-56 suggests that the starspot measurement of the surface rotation

rate from Huber et al. (2013) is too large by a factor of ~ 2. This is plausible because long-time-scale flux variations (including any due to rotation) are removed in calibrating and de-trending *Kepler* light curves.

Kissin & Thompson (2015) proposed that the Kepler core rotation measurements of red giants could be explained by differential rotation residing primarily in the convection zone rather than in the radiative interior. They showed that the evolution of red giant core rotation rates with increasing radius is broadly consistent with theoretical models in which the rotation rate varies as $\Omega \sim r^{-\alpha}$ in convection zones, with $\alpha \sim 1\text{--}3/2$. Significant differential rotation in convection zones is also consistent with numerical simulations (Brun & Palacios, 2009) and theoretical arguments based on geostrophic balance and/or conservation of angular momentum by convective eddies (Kissin & Thompson, 2015). Our results strongly disfavor Kissin & Thompson (2015)'s proposal that *most* of the differential rotation in red giants is in the convection zone (Fig. B.3 and B.4). However, models with some differential rotation in the convection zone ($\alpha \sim 1$) are marginally consistent with the measured rotational splittings and the surface rotation rate of Kepler-56. Constraining better the degree of differential rotation in red giant convection zones would be very valuable because this is one of the significant uncertainties in models of rotating stellar evolution. Doing so would require a larger sample of accurate surface rotation rates for $R \sim 4\,\mathrm{R_\odot}$ red giants with seismology, either from spectroscopy or spot measurements.

The presence of strong differential rotation in the non-burning hydrogen region of Kepler-56 (and other red giants) strongly suggests that angular momentum transport in the interior is dominated by small-scale instabilities that cannot overcome the strongly stabilizing composition gradient at the edge of the helium core. By contrast, large scale magnetic fields are disfavored because they would be expected to maintain solid body rotation across this stabilizing composition gradient.