Theoretical Concepts for Multimessenger Astrophysics

Daniel M. Siegel

Department of Physics and Columbia Astrophysics Laboratory,
Columbia University, New York, NY 10027, USA

December 9, 2018

Abstract

The recent discovery of gravitational waves from a binary neutron star merger and its accompanying electromagnetic counterparts was a watershed moment in the history of physics and astrophysics and marked the beginning of multimessenger astronomy including gravitational waves. In this series of lectures, I will discuss the theoretical concepts we use to describe such electromagnetic counterparts and to infer source properties. I will discuss both non-thermal and thermal transients, including theory of short gamma-ray bursts and their afterglows as well as theory of kilonovae and related aspects of r-process nucleosynthesis. Although the field is expected to evolve rapidly, thanks to new observational discoveries and theoretical advances, the methods discussed here will still form the foundation of our understanding for many years to come.

Contents

1 Kilonovae

1.1 Introduction

Kilonovae are thermal transients powered by the radioactive decay of heavy neutron-rich nuclei, which have been synthesized by the rapid neutron capture process (r-process; Burbidge et al. 1957; Cameron 1957) in neutron-rich ejecta from neutron star mergers (Li & Paczyński 1998; Metzger et al. 2010). These transients typically peak on a timescale of days to a week and represent a robust electromagnetic counterpart to the gravitational-wave signal of binary neutron star mergers; this is because the ejection of neutron rich material from binary neutron star mergers is essentially inevitable. A fraction of black hole–neutron star mergers will also be accompanied by kilonovae, provided the neutron star does not directly plunge into the black hole, which depends on the angular momentum of the black hole, the mass ratio, and the equation of state of high-density nuclear matter (Foucart 2012). Furthermore, kilonovae are intrinsically interesting because they provide a direct probe of the r-process and thus provide insight into the astrophysical origin of the heavy elements, which has been a long-standing mystery for more 70 years. For a discussion of the historical background of kilonovae and r-process nucleosynthesis from neutron star mergers, I refer to Metzger (2017).

We will start our discussion of kilonovae with a brief overview of the basic ingredients for various types of thermal transients (Sec. 1.2), which highlights some fundamental similarities to supernovae. We will then discuss the specific ingredients for kilonovae (Sec. 1.3) and related microphysics.

1.2 Thermal transients

In this section, we will discuss a basic model that applies to various types of thermal transients, in particular Type I supernovae, which are powered by the radioactive decay of ^{56}Ni into ^{56}Co and ^{56}Fe, and superluminous supernovae, which require additional energy injection, e.g., from the spin-down of a magnetar. Kilonovae are conceptually similar in the sense that the power source is replaced by radioactive decay from heavy nuclei synthesized by the r-process. However, as we shall discuss, there are also important differences, e.g., in terms of the microphysics related to composition, opacity, thermalization etc.

Let us assume that the material from which the thermal emission arises is to first order spherically symmetric and in homologous expansion away from the explosion site. This is motivated by the fact that the ejecta has to expand over many orders of magnitude in radius from the ejection site before radiation can escape from the initially optically thick ejecta and the kilonova or supernova emission peaks. As we shall show below, kilonova ejecta is typically required to expand to radii $r_{\rm peak} \sim 10^{15}$ cm from the initial scales $r \sim 10\,{\rm km} = 10^6$ cm of the merger site. Outflows will typically show a spread in ejection velocity and thus eventually approach homologous expansion. Therefore, one can assume the ejecta material to have a power law density structure

$$\rho(r,t) = \frac{\beta}{4\pi}\frac{M_{\rm ej}}{R_{\rm in}(t)^3}\left(\frac{r}{R_{\rm in}(t)}\right)^{-(\beta+3)}, \quad r \geq R_{\rm in}(t), \tag{1}$$

where $R_{\rm in}(t) = v_0 t$ denotes the innermost radius of the ejecta and v_0 its speed. Since typically the density profile is rather steep, most of the ejecta mass resides at radii close to $R_{\rm in}(t)$ and thus v_0 is a typical velocity for the bulk of the ejecta matter. For homologous expansion,

$$r(v) = r_v = vt, \tag{2}$$

where v is the expansion velocity (constant in the local comoving fluid frame), and Eq. (1) can be integrated to give

$$M(v) = M_v = M_{\rm ej}\left(\frac{v}{v_0}\right)^{-\beta}, \quad v \geq v_0, \tag{3}$$

thus eliminating the radial and time dependence by introducing the velocity coordinate; M_v represents the total mass outside the velocity v. In practice, realistic ejecta profiles require multiple power-law segments; however, the following analysis can easily be generalized to accommodate such additional segments. For the time being we are agnostic about the actual ejection mechanism (i.e., the physics of the supernova or kilonova 'explosion' that gave rise to mass ejection), but rather assume some generic ejecta of the type (1), (3).

The internal energy E_v of a spherical shell of volume V_v at radius $r(v)$ is determined by the first law of thermodynamics,

$$\frac{\partial E_v}{\partial t} = -p_v\frac{\partial V_v}{\partial t} + \dot{\epsilon}_v(t) - L_v(t). \tag{4}$$

Here, the first term on the right hand side,

$$-p_v\frac{\partial V_v}{\partial t} = -\frac{E_v}{r_v}\frac{{\rm d}r_v}{{\rm d}t}, \tag{5}$$

represents the work done by the shell due to adiabatic homologous expansion; we assume the material to be radiation pressure dominated, such that $p_v = E_v/(3V_v)$. In general, the heating rate supplied to the fluid, $\dot{\epsilon}_v$, can be a combination of various energy sources,

$$\dot{\epsilon}_v(t) = \dot{\epsilon}_{{\rm nuc},v}(t) + \dot{\epsilon}_{\rm NS} + \ldots, \tag{6}$$

where $\dot{\epsilon}_{{\rm nuc},v}(t)$ denotes the nuclear power supply from radioactive decays, $\dot{\epsilon}_{\rm NS}$ is the power supplied by a central magnetar (Kasen & Bildsten 2010; Metzger 2017; Siegel & Ciolfi 2016a), which becomes important for superluminous supernovae, and further terms may include, e.g., fall-back accretion (Metzger 2017). In the following (see Sec. 1.3.1), we will concentrate on radioactively supplied energy. The third term on the right hand side of Eq. (4) represents the radiated luminosity L_v.

Radiation emitted by a shell at velocity v has to diffuse out of the outer layers $>v$ for it to become observable. The diffusion time to the outer radius $R_{\rm out}$ of the ejecta is given by

$$\begin{aligned}
\bar{t}_{{\rm d},v} &= \frac{r_v - R_{\rm out}(t)}{c}[\tau(r_v,t)+1] && (7)\\
&\approx \frac{r_v}{c}[\tau(r_v,t)+1] && (8)\\
&\equiv t_{{\rm d},v} + r_v/c, && (9)
\end{aligned}$$

where $r_v - R_{\rm out}(t) \approx r_v$ for homologous expansion. The light travel time has been added in the previous equations to floor the diffusion time once the optical depth

$$\tau(r_v,t) = \int_{r_v}^{\infty}\kappa(r)\rho(r,t){\rm d}r \tag{10}$$

drops below unity and radiation can leak out directly; κ denotes the opacity of the ejecta material (cross section per unit mass).

Exercise 1.1 *Assuming a constant opacity $\kappa(r) = \kappa = $ const., show that*

$$t_{{\rm d},v} = \frac{\beta}{\beta+2}\frac{M_v\kappa}{4\pi r_v c} = \frac{\beta}{\beta+2}\frac{M_v^{(\beta+1)/\beta}\kappa}{4\pi c v_0 M_{\rm ej}^{1/\beta}t}. \tag{11}$$

In the optically thick regime $\tau(r_v, t) \gg 1$, we can use the diffusion equation, $\mathbf{F} = \frac{c}{3\kappa\rho}\nabla u$, where $\mathbf{F}$ is the radiative flux and u the radiation density, to write the radiative luminosity as

$$L_v = 4\pi r_v^2 \frac{c}{3\kappa_v \rho_v} \frac{\partial (E_v/V_v)}{\partial r} \approx \frac{4\pi c}{3\kappa_v} \frac{E_v r_v}{M_v} \approx \frac{E_v}{t_{\mathrm{d},v}}, \tag{12}$$

where we have used $M_v \approx \rho_v V_v$. More explicitly,

$$L_v(t) \approx \frac{E_v t}{(\beta \kappa_v M_v^{(\beta+1)/\beta})/(4(\beta+2)\pi c v_0 M_{\mathrm{ej}}^{1/\beta})}. \tag{13}$$

When evolving the equations to late times, we shall use the modified expression

$$L_v(t) = \frac{E_v(t)}{t_{\mathrm{d},v}(t)}, \tag{14}$$

in order to transition to the optically thin regime (cf. Eq. (7)).

General transient evolution. Dividing Eq. (4) by E_v we can write the evolution of the internal energy in terms of timescales,

$$\frac{\dot{E}_v}{E_v} = -\frac{1}{t_{\mathrm{exp},v}} + \frac{1}{t_{\mathrm{heat},v}} - \frac{1}{t_{\mathrm{d},v}}, \tag{15}$$

where we have defined the *expansion timescale* of the ejecta,

$$t_{\mathrm{exp},v} = r_v/v = t, \tag{16}$$

and the *heating timescale*,

$$t_{\mathrm{heat},v} = E_v/\dot{\epsilon}_v. \tag{17}$$

Equation (16) shows that energy (in particular any initial energy from the 'explosion' itself) injected on a timescale t will be lost on the same timescale due to adiabatic expansion. At very early times the ejecta material is very dense and has a very high optical depth (Eq. (10)) such that radiation cannot leak out from the interior, $L_v \approx 0$, and the ejecta material thus rapidly looses any initial energy from the explosion due to adiabatic expansion, i.e., the initial conditions are 'hidden' and 'washed out'. Therefore, the ejecta material only generates an observable transient if the ejecta material is reheated at late times, i.e., if heat is continuously supplied by $\dot{\epsilon}_v$ on timescales required for the bulk of the ejecta material to become transparent, such that the internal energy can leak out as radiation (note that $\tau \propto t^{-1}$). As one application of this argument, it appears very unlikely that the bright early blue emission from the recent kilonova associated with GW170817 on timescales of days was powered by initial energy deposited in the ejecta by the interaction of the GRB jet (the so-called cocoon heating), as proposed by several cocoon scenarios (e.g., Piro & Kollmeier 2017; Kasliwal et al. 2017). Even the first kilonova data points on a timescale of $\sim$11 h seem to be unlikely due to cocoon heating unless (unphysically) high energies were injected, as the jet interaction timescale (heating/injection timescale) is on the order of the GRB duration, i.e., seconds.

The presence of a transient requires heating to dominate adiabatic losses overall. At early times until the radiation peaks, $L_v \approx 0$ is a good approximation due to the high optical depth. Therefore, the present internal energy is roughly given by the integrated heating rate (cf. Eq. (4)), and we can approximate the heating timescale by

$$t_{\mathrm{heat},v}(t) \approx \left| \frac{\int_{t_{\mathrm{min}}}^{t} \dot{\epsilon}_v(t')\mathrm{d}t'}{\dot{\epsilon}_v(t)} \right|, \tag{18}$$

where t_{min} is some effective minimum time at which heating sets in.

Exercise 1.2 (*i*) *Show that for* Type I supernovae,

$$t_{\mathrm{heat},v}(t) \approx t. \tag{19}$$

Assume that the heating rate for ^{56}Ni *decay is of the form* $\dot{\epsilon}_v(t) = \epsilon_0 \exp(-t/\tau_{^{56}Ni})$ *($t \geq 0$), where* $\tau_{^{56}Ni} \approx 8.8\,\mathrm{d}$ *is the decay timescale of* ^{56}Ni.

(*ii*) *Show that for* kilonovae,

$$t_{\mathrm{heat},v}(t) \approx t/(\alpha - 1) \sim t. \tag{20}$$

Assume that the heating rate from the radioactive elements produced by the r-process is of the form $\dot{\epsilon}_v(t) = \epsilon_0 t^{-\alpha}$ *for* $t \geq t_{\mathrm{min}} \sim 1\,\mathrm{s}$, *where* $\alpha \approx 1.3$ *(Metzger et al. 2010; Roberts et al. 2011; see Sec. (1.3.1)).*

These expressions (Eqs. (19) and (20)) are only accurate for the bulk of the ejecta material until peak light, the conditions of which we will discuss now.

Arnett's law. Note that using Eq. (5), one can rewrite Eq. (4) as

$$\frac{1}{t}\frac{\partial(E_v t)}{\partial t} = \dot{\epsilon}_v(t) - L_v(t). \tag{21}$$

From Eq. (13), it follows that the luminosity peaks when $\partial(E_v t)/\partial t = 0$. Equation (21) then implies that at peak light, the radiated luminosity equals the heating rate,

$$L_{v,\mathrm{peak}} = \dot{\epsilon}_v(t_{\mathrm{peak},v}), \tag{22}$$

which is a general version of Arnett's law (Arnett 1979, 1982). Combining Eqs. (15) and (21), we find that the peak time $t_{\mathrm{peak},v}$ is determined by the condition

$$t_{\mathrm{heat},v} = t_{\mathrm{d},v}, \tag{23}$$

which making use of Eqs. (19), (20) translates into the more familiar approximate condition

$$t_{\mathrm{d},v}(t_{\mathrm{peak},v}) \approx t_{\mathrm{peak},v}. \tag{24}$$

Employing Eq. (11) we find

$$t_{\mathrm{peak},v} = \left(\frac{\beta}{\beta+2}\frac{M_v \kappa}{4\pi v c}\right)^{1/2}. \tag{25}$$

The overall peak time and peak luminosity of the transient is determined by the innermost mass layer at v_0, where the bulk of the ejecta mass and thus radioactive power resides. For typical parameters of *kilonovae* (and setting $\beta = 3$), we thus find

$$t_{\mathrm{peak}} \approx 1.2\,\mathrm{d}\left(\frac{M_{\mathrm{ej}}}{10^{-2}M_\odot}\right)^{1/2}\left(\frac{v_0}{0.1c}\right)^{-1/2}\left(\frac{\kappa}{1\,\mathrm{cm^2 g^{-1}}}\right)^{1/2}. \tag{26}$$

Varying $\kappa = 1 - 100\,\mathrm{cm^2 g^{-1}}$, representative of lanthanide-free to lanthanide-rich matter (see Sec. 1.3.2), one obtains characteristic peak timescales of kilonovae from $\sim 1\,\mathrm{day}$ to $\sim 1\,\mathrm{week}$, which is in remarkably good agreement with the observed peak timescales of the blue (lanthanide-poor) and red (lanthanide-rich) kilonova component of GW170817 (e.g., Villar et al. 2017). From Eqs. (22) and (25), (26) it also immediately follows that *higher opacities* imply *longer peak timescales* and thus *dimmer transients*, as the heating rates are typically a decreasing function of time (see above and Sec. 1.3.1).

Thermal emission and emergent spectrum. In Eq. (4) we have assumed that the power supplied by $\dot{\epsilon}_v(t)$ is completely thermalized into internal energy. The radiated luminosity must therefore be thermal or quasi-thermal black-body emission. The local gas temperature is (assuming, as usual, a radiation dominated gas $E/V = aT^4$)

$$T_v = \left(\frac{3E_v}{4\pi a r_v^3}\right)^{1/4}, \tag{27}$$

where a is the radiation constant. Note that the opacity (Eq. (12); Sec. 1.3.2) depends on temperature and thus on v. The radius $r_{\rm ph}(t)$ of the photosphere inside the ejecta is set by the condition that the optical depth (cf. Eq. (10)) becomes unity,

$$\tau(r_{\rm ph}(t), t) = \int_{r_{\rm ph}(t)}^{\infty} \kappa(r)\rho(r,t){\rm d}r = 1. \tag{28}$$

The effective temperature of the emission is determined by the Stefan-Boltzmann law,

$$T_{\rm eff}(t) = \left(\frac{L_{\rm tot}(t)}{4\pi\sigma r_{\rm ph}(t)^2}\right)^{1/4}, \tag{29}$$

where

$$L_{\rm tot}(t) = \int_{R_0}^{\infty} L_v(t){\rm d}r_v \tag{30}$$

is the total luminosity of the ejecta material at time t and σ is the Stefan-Boltzmann constant. Finally, the emergent flux spectrum of the transient is given by (assuming perfect black-body emission),

$$F_\nu(t) = \frac{2\pi h\nu^3}{c^2}\frac{1}{\exp[h\nu/kT_{\rm eff}(t)] - 1}\frac{r_{\rm ph}(t)^2}{D^2}, \tag{31}$$

where ν denotes frequency, h is Planck's constant, and D is the distance to the source.

In conclusion, we note that the above discussion provides a very simple one-dimensional framework to model thermal transients. Despite its simplicity, it is powerful enough to obtain a qualitative understanding of various types of thermal transients and even allows for quantitative predictions that are reasonably accurate within a factor of a few. In practice, accurate modeling requires detailed microphysics in combination with multi-dimensional hydrodynamic and radiative transfer simulations.

Exercise 1.3 *Using a programming language of your choice (e.g., python), implement the model discussed here, together with the expressions for heating rates and opacities provided in Sec. 1.3, in order to explore kilonova emission in different wavelength bands (U,V,R,I,J,K) for various ejecta types (cf. Sec. 1.3.3) and compositions.*

1.3 Kilonova microphysics

From the previous discussion on the general evolution of a thermal transient, it emerged that in order to describe any particular transient, one needs to understand and specify the heating source $\dot{\epsilon}(t)$, and the characteristics of the ejecta, including the opacity κ (which depends on the composition), the total ejecta mass $M_{\rm ej}$, its typical velocity v_0, and the ejecta profile characterized by the power law exponent β. We shall briefly summarize these characteristic for kilonovae below.

1.3.1 Heating rates

In kilonovae, the ejecta material is heated by the radioactive decay of heavy nuclei synthesized by the r-process. For the time being, we shall neglect other possible energy sources, such as continuous heating from a magnetar via its pulsar wind nebula interior to the kilonova ejecta. In general, the nuclear heating rate for kilonovae can then be expressed as

$$\dot{\epsilon}_{{\rm nuc},v}(t) = \dot{e}_{\rm r}(t)\eta_{{\rm th},v}(t)X_{{\rm r},v}{\rm d}m_v \tag{32}$$

where $\dot{e}_{\rm r}(t)$ is the specific heating rate (energy per unit time per unit mass), $\eta_{{\rm th},v}$ is the thermalization efficiency with which the supplied energy is thermalized in the ejecta, ${\rm d}m_v$ is the mass of a spherical shell at velocity v, and $X_{{\rm r},v}$ is the mass fraction of r-process elements within this shell.

Specific heating. In general, the outcome of the r-process depends on (see, e.g., Lippuner & Roberts 2015) the proton fraction[1]

$$Y_{\rm e} = \frac{n_{\rm p}}{n_{\rm p} + n_{\rm n}}, \tag{33}$$

where $n_{\rm p}$ and $n_{\rm n}$ denote the number densities of free protons and neutrons, respectively, the specific entropy s and the expansion timescale τ of the matter at the onset of the r-process (when the initial temperature of the explosion material drops below $\approx 5\,{\rm GK}$, nuclear statistical equilibrium breaks down, and neutron captures set in). The heating rate is mostly controlled by the electron fraction (e.g., Lippuner & Roberts 2015). For sufficiently neutron-rich (lanthanide-rich) ejecta with mean proton fraction $Y_{\rm e} < 0.25$, the specific heating rate from the r-process can be reasonably approximated by (Korobkin et al. 2012; see also Fig. 1)

$$\dot{e}_{\rm r}(t) = \dot{e}_{\rm r,0} \left(\frac{1}{2} - \frac{\arctan[(t-t_0)/\sigma]}{\pi} \right)^{\alpha}, \tag{34}$$

where $\dot{e}_{\rm r,0} \approx 4 \times 10^{18}\,{\rm erg\,s^{-1}\,g^{-1}}$, $t_0 \approx 1.3\,{\rm s}$, $\sigma \approx 0.11\,{\rm s}$, and $\alpha \approx 1.3$. This analytic fit to r-process heating reflects the fact that during the actual r-process itself, i.e., as long as free neutrons are captured onto seed nuclei (which is typically a matter of $t_0 \sim {\rm s}$), the heating rate stays roughly constant. Thereafter, the heavy neutron-rich nuclei synthesized by the r-process release nuclear binding energy by α-decay, β-decay, and fission, as they gradually decay into stable isotopes at the center of the valley of β-stability. This being a statistical ensemble of nuclei, heating from the various individual isotopes with their own characteristic decay timescale superimpose to a power law heating rate (Metzger et al. 2010; Roberts et al. 2011; Fig. 1)

$$\dot{e}_{\rm r}(t) \propto t^{-\alpha}, \quad t > t_0. \tag{35}$$

Due to this statistical effect, the heating rate is largely insensitive to uncertainties in the nuclear mass models, cross sections, and fission fragment distribution (although, e.g., over-production of actinides with some mass models relative to others could lead to significantly enhanced heating at late times from their α-decays that are thermalized efficiently, Barnes et al. 2016; Rosswog et al. 2017). However, the final detailed r-process abundance pattern will depend on these details (e.g., Eichler et al. 2015; Mumpower et al. 2016).

For higher mean electron fractions $Y_{\rm e} \gtrsim 0.25$ (or for $Y < 0.25$ in certain combinations with entropy and expansion timescale), the ejecta is lanthanide-poor and individual bumps and wiggles become apparent in the heating rate due to individual nuclides dominating the heating (Lippuner & Roberts 2015). Due to the dominance of individual isotopes, the heating rate is also generally reduced with respect to the lanthanide-rich case.

On longer timescales of weeks to months, even in the lanthanide-rich case the decay energy input can be dominated by a discrete number of α-decays; resulting 'bumps' in the heating that translate into the bolometric lightcurve could serve as observational fingerprints for the production of specific elements (Wu et al. 2018).

Thermalization efficiency. The amount of nuclear binding energy absorbed by the ejecta material depends on the efficiency with which the decay products from α- and β-decay and fission deposit their energy in the ejecta (see Metzger et al. 2010; Barnes et al. 2016; Hotokezaka et al. 2016 for more details). In summary, neutrinos are lost entirely without depositing energy, gamma-rays ($\sim$MeV) are only trapped at early times ($\lesssim 1\,{\rm d}$) and

[1] For historic reasons called electron fraction, and referred to as "$Y_{\rm e}$".

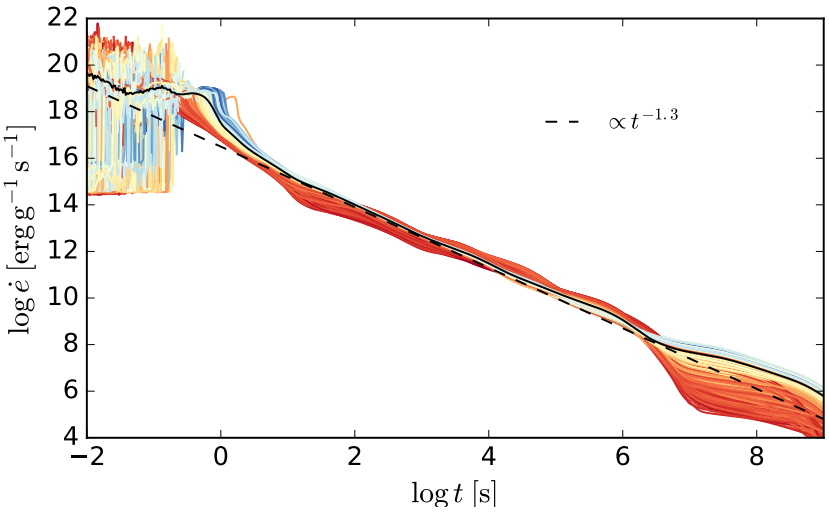

Figure 1: Heating rates following r-process nucleosynthesis in trajectories of post-merger accretion disk winds (lanthaninde-rich; from Siegel & Metzger 2017, 2018), indicating the expected $\propto t^{-1.3}$ behavior after the r-process concluded on a timescale $\sim$1 s.

then leak out due to the very small electron scattering cross section at high energies (Klein-Nishina regime). On the contrary, α-particles and electrons as well as fission fragments deposit their kinetic energy efficiently through Coulomb collision and ionization. In general, fission fragments thermalize most efficiently, followed by α-particles, followed by β-particles, and, finally, γ-rays. All processes combined, the overall thermalization efficiency can be approximately parametrized by an analytic fit (Barnes et al. 2016, their Eq.(34))

$$\eta_{\rm th,}v(t) = 0.36 \left[\exp(-a_v t_{\rm 1d}) + \frac{\ln(1+2b_v t_{\rm 1d}^{d_v})}{2b_v t_{\rm 1d}^{d_v}} \right], \tag{36}$$

where t is in units of days, and a_v, b_v, and d_v are fitting constants that, in general, depend on the mass and velocity of the mass shell under consideration. As a rough representative fit for the kilonova as a whole, one may employ this fit evaluated for the innermost (typical) ejecta layer with $v = v_0$, setting $M_v = M_{\rm ej}$. For typical kilonova values, $v_0 = 0.1c$, $M_{\rm ej} = 10^{-2}\,M_\odot$, one obtains $a_v = 0.56$, $b_v = 0.17$, $c_v = 0.74$ (Table 1 of Barnes et al. 2016).

1.3.2 Opacities

1.3.3 Origin of neutron-rich ejecta

2 Gamma-ray burst afterglows

2.1 Overview

The aim of this section is to derive a model for gamma-ray burst (GRB) afterglow emission, arising as synchrotron emission from electrons accelerated at the shock front of a relativistic blast wave that decelerates from the explosion site into the interstellar medium (ISM). We will show that the synchrotron afterglow emission of gamma-ray bursts can be described as broken power-law segments in frequency ν and time t of the type $F_\nu \propto \nu^\beta t^\alpha$. We will first derive the underlying hydrodynamic solution that describes the relativistic outflow of the GRB (Sec. 2.2) and then compute the synchrotron emission arising from this solution (Sec. 2.3), following original papers (Blandford & McKee 1976; Sari et al. 1998; Granot & Sari 2002).

2.2 Relativistic blast waves (Blandford & McKee)

We will describe the GRB as an impulsive spherical explosion, an adiabatic spherical relativistic blast wave, that decelerates into a cold uniform circumburst medium (the ISM) with constant particle density n. We assume an adiabatic outflow, i.e., we assume that losses due to synchrotron radiation are negligible, such that they do not affect the hydrodynamic solution itself. Such a blast wave arises as a self-similar solution to the hydrodynamic equations in spherical symmetry and represents the relativistic analogue of the Newtonian Sedov-Taylor blast wave solution (Sedov 1959; Taylor 1950). We will construct the blast wave solution in three steps: characterizing the local properties of a relativistic blast wave (relativistic shock; Sec. 2.2.1), formulating the relevant evolution equations in spherical symmetry (Sec. 2.2.2), and constructing a self-similar solution imposing the local properties (Sec. 2.2.3). In this section we will set the speed of light $c = 1$ for convenience.

2.2.1 Relativistic shocks

A shock is a discontinuity in pressure, temperature, and density that propagates supersonically through a fluid. It is characterized by so-called jump conditions (or Rankine-Hugoniot conditions) that relate the fluid quantities across such a discontinuity. The special-relativistic Rankine-Hugoniot conditions (e.g., Taub 1948; Marti et al. 1994) were obtained by Taub (1948). For it to be a solution of the hydrodynamics equations, such a shock wave must satisfy continuity of the mass and energy-momentum flux across the discontinuity:

$$[\rho u^\mu] n_\mu = 0, \tag{37}$$

$$[T^{\mu\nu}] n_\mu = 0. \tag{38}$$

Here, $[f] = f_2 - f_1$ relates the values of a fluid quantity f on one side (2), which we shall assume to be the shocked fluid, to the other (unshocked) side (1) of the discontinuity surface with normal vector n^μ. For convenience, let us choose a local Cartesian coordinate frame that is comoving with the unshocked fluid (1), setting $n^\mu = (0, 1, 0, 0)$. Let us denote the Lorentz factor of the shock itself as measured in this frame by Γ. Furthermore, we assume an ideal fluid with rest-mass density ρ, particle density n, pressure p, specific internal energy ϵ, energy density $e = \rho(1 + \epsilon)$, enthalpy $w = e + p$, four-velocity $u^\mu = (\gamma, \gamma u, 0, 0)$, where $\gamma = (1 - u^2)^{-1/2}$, and energy-momentum tensor $T^{\mu\nu} = \rho(1 + \epsilon + p/\rho)u^\mu u^\nu + p\eta^{\mu\nu}$, where $\eta^{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$ is the Minkowski metric and where we have set $c = 1$. We assume an ideal gas equation of state, $p = (\Gamma_{\rm ad} - 1)\rho\epsilon$, where Γ denotes the adiabatic index. We are interested in the case of a strong shock, in which $p_2/n_2 \gg p_1/n_1$. Under these assumptions, the Rankine-Hugoniot conditions (Eqs. (37) and (38)) can be written as

$$\frac{e_2}{n_2} = \gamma_2 \frac{w_1}{n_1}, \tag{39}$$

$$\frac{n_2}{n_1} = \frac{\Gamma_{\rm ad,2}\gamma_2 + 1}{\Gamma_{\rm ad,2} - 1}, \tag{40}$$

$$\Gamma^2 = \frac{(\gamma_2 + 1)[\Gamma_{\rm ad,2}(\gamma_2 - 1) + 1]^2}{\Gamma_{\rm ad,2}(2 - \Gamma_{\rm ad,2})(\gamma_2 - 1) + 2}. \tag{41}$$

Exercise 2.1 *Derive Eqs.* (39)–(41).

Furthermore, we are interested in the ultra-relativistic case, in which $\Gamma \gg 1$ and $\Gamma_{\rm ad} = 4/3$, i.e., $p = e/3$. Retaining only terms up to order $\mathcal{O}(\Gamma^{-1})$, one obtains the surprisingly simple form of the jump conditions:

$$p_2 = \frac{2}{3}\Gamma^2 w_1, \tag{42}$$

$$n_2' = 2\Gamma^2 n_1, \tag{43}$$

$$\gamma_2^2 = \frac{1}{2}\Gamma^2. \tag{44}$$

Here, $n_2' = \gamma_2 n_2$ denotes the density measured in the frame of the unshocked gas.

Exercise 2.2 *Derive the ultra-relativistic limit Eqs.* (42)–(44).

2.2.2 Hydrodynamics in spherical symmetry

The blast wave solution to be constructed must satisfy the equations of baryon (particle) number conservation and energy and momentum conservation,

$$\nabla_\mu (n u^\mu) = 0, \tag{45}$$
$$\nabla_\mu T^{\mu\nu} = 0, \tag{46}$$

which in spherical symmetry can be written as

$$\frac{\partial n'}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 n' u) = 0, \tag{47}$$
$$\frac{\partial}{\partial t}[\gamma^2(e+u^2 p)] + \frac{1}{r^2}\frac{\partial}{\partial r}[r^2\gamma^2 u(e+p)] = 0, \tag{48}$$
$$\frac{\partial}{\partial t}[\gamma^2 u(e+p)] + \frac{1}{r^2}\frac{\partial}{\partial r}[r^2\gamma^2 u^2(e+p)] = 0, \tag{49}$$

where, as usual, $n' = \gamma n$ denotes the particle density in the fixed lab frame with coordinates (r, t). In the following, we will focus on the ultra-relativistic equation of state

$$p = \frac{1}{3}e, \tag{50}$$

which is justified as long as the internal energy is dominated by radiation or relativistic particles, such that their rest masses can be neglected. This is appropriate for gamma-ray bursts, which we think of as an essentially baryon-free '*fireball*' of high entropy material (radiation and ultra-relativistic particles). Using this relation and noting that the total time derivative of a function $y(r, t)$ is given by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\partial y}{\partial t} + \frac{\partial r}{\partial t}\frac{\partial y}{\partial r} = \frac{\partial y}{\partial t} + u\frac{\partial y}{\partial r}, \tag{51}$$

the so-called convective or material derivative, one can rewrite Eqs. (47)–(49) in the more compact form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{p}{n^{4/3}}\right) = 0, \tag{52}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma^4 p) = \gamma^2\frac{\partial p}{\partial t}, \tag{53}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\ln(\gamma^4 p^3) = -\frac{4}{r^2}\frac{\partial}{\partial r}(r^2 u). \tag{54}$$

For later reference, using the total energy density $T^{00} = (e+p)(u^0)^2 - p = (4\gamma^2 - 1)p \simeq 4\gamma^2 p$ in the ultra-relativistic limit ($\gamma \gg 1$), we also note that the instantaneous total energy contained in a volume enclosed by radii $r_{\rm in} < r_{\rm out}$ is given by

$$E(t; r_{\rm in}, r_{\rm out}) = \int_{r_{\rm in}}^{r_{\rm out}} 16\pi\gamma^2 p r^2 \mathrm{d}r. \tag{55}$$

2.2.3 Constructing a self-similar blast wave solution

In this section, we will construct a self-similar ultra-relativistic blast wave solution to the spherical hydrodynamic equations (Eqs. (52)–(54)) knowing that at the shock front the solution must satisfy the jump conditions Eqs. (42)–(44). We will assume an adiabatic blast wave, i.e., that there are no significant energy losses through radiation leaking out of the fluid.

We will look for solutions for which the shock front propagates as a power-law in time,

$$\Gamma^2 \propto t^{-m}, \quad m > -1. \tag{56}$$

Essentially, the motivation for this is that $\Gamma^2 t^3 = \text{const.}$ implies that the total energy of the blast wave is conserved (as shown below, Eq. (70)). Deviating from $m = 3$ means that we are also allowing solutions in which energy is supplied to the blast wave as a power-law in time. The instantaneous radius of the shock front is then given by

$$R(t) = \int_0^t \sqrt{1 - \frac{1}{\Gamma(\hat{t})^2}}\mathrm{d}\hat{t} = t\left[1 - \frac{1}{2(m+1)\Gamma^2}\right] \tag{57}$$

where we have only retained terms up to $\mathcal{O}(\Gamma^{-2})$. It is reasonable to assume that a spherical shock wave represents a self-similar solution to the hydrodynamics equations, i.e., that the solution at each time and shock radius (i.e., on all scales) only depends on the shock properties and the relative distance to the shock front $R - r$. This motivates a dimensionless similarity variable

$$\chi = 1 + 2(m+1)\Gamma^2\,(1 - r/R) = 1 + 2(m+1)\Gamma^2\,(1 - r/t)\,, \tag{58}$$

where we have again only retained terms up to $\mathcal{O}(\Gamma^{-2})$. We note that $\chi \geq 1$, with $\chi = 1$ at the shock front, $r = R$, and $\chi = 1 + 2(m+1)\Gamma^2 \simeq 2(m+1)\Gamma^2$ at the origin, $r = 0$. With this definition, we have $\chi \simeq (R/t)(1 - r/R) = (R - r)/t$ for $r \neq R$. Following this line of thought, we make the ansatz that all quantities in the shocked fluid ($r \leq R$) only depend on the shock properties and the similarity variable χ,

$$p(r,t) = \frac{2}{3} w_1 \Gamma^2 f(\chi), \tag{59}$$

$$\gamma^2(r,t) = \frac{1}{2}\Gamma^2 g(\chi), \tag{60}$$

$$n'(r,t) = 2n_1\Gamma^2 h(\chi), \tag{61}$$

and impose

$$f(1) = g(1) = h(1) = 1 \tag{62}$$

to satisfy the jump conditions (42)–(44) at the shock front. In the following, we will drop the subscript 2 as it is implicitly understood that we are only interested in the hydrodynamic solution interior to the shock front (with the exterior solution being trivial). Changing independent variables $(r,t) \to (\Gamma^2, \chi)$, the partial differential equations (59)–(61) translate into a set of ordinary differential equations in the similarity variable χ,

$$\frac{1}{g}\frac{\mathrm{d}\ln f}{\mathrm{d}\chi} = \frac{8(m-1) - (m-4)g\chi}{(m+1)(4 - 8g\chi + g^2\chi^2)}, \tag{63}$$

$$\frac{1}{g}\frac{\mathrm{d}\ln g}{\mathrm{d}\chi} = \frac{(7m-4) - (m+2)g\chi}{(m+1)(4 - 8g\chi + g^2\chi^2)}, \tag{64}$$

$$\frac{1}{g}\frac{\mathrm{d}\ln h}{\mathrm{d}\chi} = \frac{2(9m-8) - 2(5m-6)g\chi + (m-2)g^2\chi^2}{(m+1)(4 - 8g\chi + g^2\chi^2)(2 - g\chi)}, \tag{65}$$

These equations have a simple solution for a blast wave with constant total energy E, i.e., in the case $m = 3$:

$$f = \chi^{-17/12} \tag{66}$$

$$g = \chi^{-1} \tag{67}$$

$$h = \chi^{-7/4}. \tag{68}$$

This corresponds to our description of a GRB as an *impulsive* injection of a fixed amount of energy E at the origin at $r = t = R = 0$ into an otherwise uniform cold medium. For other solutions of blast waves with energy supply ($m < 3$) and or blast waves running into an external medium with density gradient $n_1 \propto r^{-k}$, see Blandford & McKee (1976).

In order to show that Eqs. (66)–(68) indeed correspond to an impulsive injection, we compute the total instantaneous energy of the blast wave using Eq. (55),

$$E = \int_0^R 16\pi\gamma^2 p r^2 \mathrm{d}r \tag{69}$$

$$= \frac{8\pi}{17} w_1 \Gamma^2 t^3, \tag{70}$$

which shows that $\Gamma^2 t^3 = \text{const.}$ (cf. Eq. (56)) indeed implies $E = \text{const.}$.

Exercise 2.3 *Derive the result Eq. (70) by changing the integration variable $r \to \chi$ and only retaining terms up to order $\mathcal{O}(\Gamma^{-2})$, keeping in mind that we are interested in the instantaneous energy at given time t (given Γ^2). Note that the transformation of the differential is given by $\mathrm{d}r = (\partial r/\partial\Gamma^2)_\chi \mathrm{d}\Gamma^2 + (\partial r/\partial\chi)_{\Gamma^2}\mathrm{d}\chi$.*

2.3 Synchrotron emission from blast waves

In this section we shall compute the broadband synchrotron emission of an impulsive relativistic blast wave (Sec. 2.2.3) as seen by a distant observer. To this end, we first discuss the hydrodynamic blast wave as seen by a distant observer (Sec. 2.3.1), then elaborate on how electrons accelerated at the shock front give rise to synchrotron emission (Sec. 2.3.2), and finally construct lightcurves and spectra as seen by a distant observer (Sec. 2.3.3).

2.3.1 Blast wave as seen by a distant observer

A photon emitted from the shock front at radius R_1 at $t = t_0$ is received by a distant observer at $r = d \to \infty$ at time $t_1 = t_0 + (d - R_1)/c$, where t refers to the coordinate time of the blast wave solution (rest frame of the unshocked fluid (ISM); Sec. 2.2.3). A photon emitted later at $R_2 > R_1$ and $t_0 + (R_2 - R_1)/v$, where v denotes the velocity of the shock front, is received by the distant observer at time $t_2 = t_0 + (R_2 - R_1)/v + R_2/c$. The distant observer notices a time difference

$$\Delta t_{12} = t_2 - t_1 = \frac{R_2 - R_1}{v} - \frac{R_2 - R_1}{c} = \frac{R_2 - R_1}{c}\left(\frac{1-\beta}{\beta}\right), \tag{71}$$

where $\beta = v/c$. In the ultra-relativistic case ($\Gamma >> 1$, $\beta \simeq 1$),

$$\frac{1}{\Gamma^2} = (1-\beta)(1+\beta) \simeq 2(1-\beta), \tag{72}$$

such that

$$\Delta t_{12} = \frac{R_2 - R_1}{2c\Gamma^2}, \tag{73}$$

Taking $R_1 \to 0$, one can associate the location of the shock front $R(t)$ with Lorentz factor Γ with an *observer time*

$$t_{\rm obs}(R) = \int_0^R \frac{\mathrm{d}R}{2c\Gamma^2(R)}. \tag{74}$$

For simplicity (the integration will only change the result by order unity), let us keep the Lorentz factor Γ constant and use an approximate observer time

$$t_{\rm obs} = \frac{R}{4c\gamma^2}, \tag{75}$$

where we have already substituted the Lorentz factor of the shocked fluid at the location of the shock front (Eq. (60)).

The evolution of the blast wave radius as seen by the distant observer can be obtained from Eq. (70),

$$E = \frac{16\pi}{17}\rho_1 c^2 \gamma^2 R^3, \tag{76}$$

where we have assumed that the ISM (unshocked fluid) is cold and pressure-less ($\epsilon_1 = p_1 = 0$) and that approximately $R \simeq ct$ for $\Gamma \gg 1$ (cf. Eq. (57)). Either replacing γ or R in Eq. (76) with the help of Eq. (75), we obtain the scalings

$$R(t_{\rm obs}) = \left(\frac{17E}{4\pi m_{\rm p} nc}\right)^{1/4} t_{\rm obs}^{1/4}, \tag{77}$$

$$\gamma(t_{\rm obs}) = \left(\frac{17E}{1024\pi m_{\rm p} nc^5}\right)^{1/8} t_{\rm obs}^{-3/8}, \tag{78}$$

where we have used the particle density n and particle (proton) mass $m_{\rm p}$ of the ISM. In the following, we will only use results scaled to the observer time and drop the subscript "obs", assuming that it is implicitly understood that all expressions are given in observer time.

2.3.2 Synchrotron emission

Shock energetics. Assuming a cold and pressure-less ISM with particle density n ($\epsilon_1 = p_1 = 0$, i.e., $w_1 = \rho_1 = nm_{\rm p}c^2$), the shock compresses the flow and heats up the shocked material to thermal energies $e_2 = 4\gamma^2 nm_{\rm p}c^2$ (Eqs. (42) and (44)). In order to compute the synchrotron emission from electrons in the shocked fluid, one needs to know the detailed energy distribution of electrons and the magnetic field, which is determined by complicated microphysics at the shock front and difficult to estimate from first principles. For simplicity, we will therefore parametrize these unknowns and assume that a fraction ϵ_e of the total internal energy goes into random kinetic energy of the electrons, and another fraction ϵ_B into (small-scale) magnetic field energy,

$$e_e = \epsilon_e e_2 = 4\epsilon_e \gamma^2 nm_{\rm p}c^2, \tag{79}$$

$$e_B = \frac{B^2}{8\pi} = \epsilon_B e_2 = 4\epsilon_B \gamma^2 nm_{\rm p}c^2, \tag{80}$$

$$\tag{81}$$

which implies a typical magnetic field strength

$$B = (32\pi\epsilon_B m_{\rm p} n)^{1/2}\gamma c. \tag{82}$$

Focusing only on synchrotron losses implies the constraint $\epsilon_B < \epsilon_e$, as inverse Compton scattering would otherwise be significant. Furthermore, we assume that the electrons are accelerated into a power-law distribution in the Lorentz factor γ_e above some minimum $\gamma_{e,\rm m}$,

$$\mathcal{N}(\gamma_e) = C\gamma_e^{-p}, \quad \gamma_e \geq \gamma_{e,\rm m}, \tag{83}$$

with electron number density

$$n_e = \int_{\gamma_{e,\rm m}}^{\infty} \mathcal{N}(\gamma_e){\rm d}\gamma_e = \frac{C\gamma_{e,\rm m}^{-(p-1)}}{p-1}, \quad p > 1 \tag{84}$$

and energy density in the ultra-relativistic case

$$e_e = \int_{\gamma_{e,\rm m}}^{\infty} (\gamma_e - 1)m_e c^2 \mathcal{N}(\gamma_e){\rm d}\gamma_e = \frac{m_e c^2 C\gamma_{e,\rm m}^{-(p-2)}}{p-2}, \quad p > 2. \tag{85}$$

Combining Eqs. (84) and (85), we find the minimum electron Lorentz factor

$$\gamma_{e,\rm m} = \frac{p-2}{p-1}\frac{e_e}{n_e m_e c^2} = 4\frac{p-2}{p-1}\frac{m_{\rm p}}{m_{\rm e}}\epsilon_e\gamma, \tag{86}$$

where $\langle\gamma_e\rangle = e_e/n_e m_e c^2$ is the typical electron Lorentz factor, and where we have assumed ionized hydrogen gas, i.e., $n_e = n/\gamma$. From Eq. (85) one can also see that the electron distribution must be restricted to $p > 2$, in order to keep the electron energy finite, although $2 > p > 1$ is also allowed with an appropriate high-energy cut-off.

Synchrotron emission from a single electron. The spontaneously emitted spectral power at frequency ν of a single electron of charge e moving at Lorentz factor γ_e as seen by a distant observer at rest with the ISM is given by (Rybicki & Lightman 2004; Crusius & Schlickeiser 1986)

$$P_\nu(\theta) = \gamma \frac{\sqrt{3} e^3 B}{m_e c^2} \sin\theta \frac{\gamma\nu}{\tilde{\nu}_{\rm char}} \int_{\gamma\nu/\tilde{\nu}_{\rm char}}^{\infty} K_{\frac{5}{3}}(z)\,{\rm d}z, \tag{87}$$

where θ is the angle between the magnetic field and the direction of particle motion, $K_{\frac{5}{3}}$ denotes the modified Bessel function of order 5/3, and $\tilde{\nu}_{\rm char} = \nu_{\rm char}\sin\theta$, with

$$\nu_{\rm char} = \frac{3}{4\pi}\frac{eB\gamma_e^2}{m_e c}\gamma. \tag{88}$$

Here and in the following, factors of γ represent the transformation from the rest frame of the shocked fluid to the rest frame of the ISM (the distant observer). Assuming an isotropic distribution of particle velocities, we can average over all possible angles for a given Lorentz factor γ_e to obtain the average energy loss rate for a particle (cf. Crusius & Schlickeiser 1986):

$$P_\nu = \frac{\gamma}{4\pi}\int_0^{2\pi}{\rm d}\phi \int_0^{\pi}{\rm d}\theta\,\sin\theta\, P_\nu(\theta) = \frac{\sqrt{3}e^3 B}{m_e c^2}\gamma\mathcal{R}(\nu/\nu_{\rm char}), \tag{89}$$

where

$$\mathcal{R}(z) = \frac{\pi}{2} z[W_{0,\frac{4}{3}}(z)W_{0,\frac{1}{3}}(z) - W_{\frac{1}{2},\frac{5}{6}}(z)W_{-\frac{1}{2},\frac{5}{6}}(z)]. \tag{90}$$

Here, $W_{\lambda,\mu}(z) = {\rm e}^{-\frac{1}{2}z} z^{\frac{1}{2}+\mu} U(0.5+\mu-\lambda, 1+2\mu, z)$ is the Whittaker function and U denotes the confluent hypergeometric function of second kind (Abramowitz & Stegun 1972, Section 13.1). Employing asymptotic expansions of the Whittaker functions one can show that (Crusius & Schlickeiser 1986)

$$\mathcal{R}(x) \simeq \begin{cases} \text{const.} \times x^{1/3}, & x \ll 1 \\ \frac{\pi}{2}\exp(-x)\left(1 - \frac{99}{162}x^{-1}\right), & x \gg 1 \end{cases}, \tag{91}$$

which represents an excellent approximation except for $x \sim 1$, i.e., near the global maximum of $\mathcal{R}$. This means that the spectral power increases as $P_\nu \propto \nu^{1/3}$ at small frequencies, essentially up to a maximum around

$$\nu_{\rm max} \sim \frac{2}{3}\nu_{\rm char} = \frac{eB\gamma_e^2}{2\pi m_e c}\gamma, \tag{92}$$

and then exponentially decreases, $P_\nu \propto \exp(-\nu)$. The maximum spectral power is roughly given by

$$P_{\nu,\rm max} \approx \frac{P}{\nu_{\rm max}} = \frac{\sigma_{\rm T} m_e c^2}{3e} B\gamma, \tag{93}$$

where $\sigma_{\rm T} = (8\pi/3)(e^4/m_e^2 c^4)$ is the Thomson cross section and

$$P = \frac{4}{3}\sigma_{\rm T} c \gamma_e^2 \frac{B^2}{8\pi}\gamma^2 \tag{94}$$

is the total radiation power (Rybicki & Lightman 2004). Note that $P_{\nu,\rm max}$ is independent of γ_e and only depends on the magnetic field strength; the location of the peak (Eq. (92)), however, does depend on γ_e.

Electrons radiate at the expense of their kinetic energy, ${\rm d}(\gamma\gamma_e m_e c^2)/{\rm d}t = -P$, i.e., the electrons cools at a rate

$$\dot{\gamma_e} = -\gamma\frac{\sigma_{\rm T} c B^2 \gamma_e^2}{6\pi m_e c^2}. \tag{95}$$

From Eq. (94) one can see that the total radiated power increases with the electron Lorentz factor as $\propto \gamma_e^2$, such that for very high γ_e the instantaneously radiated power becomes

comparable to the kinetic energy $\gamma\gamma_e m_e c^2$ of the electron and the expression for the spectral power (89) is no longer accurate (as it assumes a time-independent (fixed) γ_e). Above a critical $\gamma_e > \gamma_{e,\rm c}$ the electron will thus cool fast and loose most of its kinetic energy, where $\gamma_{e,\rm c}$ is determined by equating the kinetic energy with the radiated synchrotron energy, $\gamma\gamma_{e,\rm c} m_e c^2 = P(\gamma_{e,\rm c})t$, i.e., using Eq. (82) one finds

$$\gamma_{e,\rm c} = \frac{6\pi m_e c}{\sigma_{\rm T} B^2 \gamma t} = \frac{3 m_e}{16 \sigma_{\rm T} m_{\rm p} c}\frac{1}{\epsilon_B n \gamma^3 t}. \tag{96}$$

Electrons with initial Lorentz factors in this fast cooling regime, $\gamma_e > \gamma_{e,\rm c}$, rapidly cool down to $\gamma_{e,\rm c}$, emitting their instantaneous kinetic energy at the characteristic frequency Eq. (92), such that in a time-integrated sense, the spectral power scales as

$$P_\nu = \frac{\gamma_e m_e c^2}{\dot\gamma_e} \propto \gamma_e^{-1} \propto \nu^{-1/2}. \tag{97}$$

Synchrotron emission from a distribution of electrons. The collective synchrotron emission from a distribution of electrons $N(\gamma_e)$ above some minimum $\gamma_e > \gamma_{e,\rm m}$ is given by

$$P_{\nu,\rm tot} = \int_{\gamma_{e,\rm m}}^{\infty} N(\gamma_e) P_\nu {\rm d}\gamma_e, \tag{98}$$

where P_ν is the spectral power of a single electron (Eq. (89)). In general, the electron distribution is determined by the continuity equation (conservation of the number of electrons),

$$\frac{\partial N}{\partial t} + \frac{\partial}{\partial \gamma_e}(\dot\gamma_e N) = \mathcal{N}(\gamma_e). \tag{99}$$

The second term on the left-hand side represents the cooling term due to synchrotron emission and the source term on the right-hand side represents the injection of electrons as given by Eq. (83). We will only focus on stationary particle distributions and therefore neglect the first term on the left-hand side ($\partial/\partial t = 0$).

Suppose that $N(\gamma_e) \propto \gamma_e^{-q}$. From Eq. (98) and (89),

$$P_{\nu,\rm tot} \propto \int_{\gamma_{e,\rm m}}^{\infty} \mathcal{R}(\nu/\nu_{\rm char}) \gamma_e^{-q} {\rm d}\gamma_e. \tag{100}$$

Noting that $\nu_{\rm char} \propto \gamma_e^2$ (Eq. (88)) and changing the integration variable to $x = \nu/\nu_{\rm char} \propto \nu/\gamma_e^2$, with ${\rm d}\gamma \propto \nu^{-1/2} x^{3/2}$, we find that

$$P_{\nu,\rm tot} \propto \nu^{-(q-1)/2} \int_0^1 \mathcal{R}(x) x^{(q-3)/2} {\rm d}x, \quad \nu > \nu_{\rm m}. \tag{101}$$

Here, we have assumed that we are only interested in frequencies larger than

$$\nu_{\rm m} = \nu_{\rm max}(\gamma_{e,\rm m}), \tag{102}$$

such that the integration bounds become independent of ν, as $\mathcal{R}(x)$ essentially vanishes above $x = 1$ (cf. Eq. (91)). For frequencies $\nu \leq \nu_{\rm m}$ (and assuming cooling is negligible) the total spectrum is simply a superposition of the individual electron spectra at $\gamma_e = \gamma_{e,\rm m}$ (Eqs. (89), (91)),

$$P_{\nu,\rm tot} \propto \nu^{1/3}, \quad \nu < \min(\nu_{\rm m}, \nu_{\rm c}), \tag{103}$$

as there are no electrons below $\gamma_{e,\rm m}$ in the absence of cooling.

For the global shape of $P_{\nu,\rm tot}$ there are, in general, two cases referred to as *fast cooling*, defined by $\gamma_{e,\rm m} > \gamma_{e,\rm c}$, i.e., all injected electrons first need to undergo fast cooling down to $\gamma_{e,\rm c}$, and *slow cooling*, defined by $\gamma_{e,\rm m} < \gamma_{e,\rm c}$, i.e., only the high-energy tail of the electron distribution undergoes initial fast cooling.

Let us first concentrate on the former case of *fast cooling*. At low frequencies $\nu < \nu_{\rm c}$, where

$$\nu_{\rm c} = \nu_{\rm max}(\gamma_{e,{\rm c}}), \tag{104}$$

the spectrum is given by Eq. (103), as all electrons that underwent fast cooling will 'pile up' at $\gamma_e = \gamma_{e,{\rm c}}$. At intermediate frequencies $\nu_{\rm c} < \nu < \nu_{\rm m}$ ($\gamma_{e,{\rm c}} < \gamma_e < \gamma_{e,{\rm m}}$) there is no injection of particles ($\mathcal{N} = 0$) and we obtain from Eq. (99): $N(\gamma_e) \propto \dot{\gamma}_e^{-1} \propto \gamma_e^{-2}$. Using $q = 2$ in Eq. (101) we thus find

$$P_{\nu,{\rm tot}} \propto \nu^{-1/2}, \quad \nu_{\rm c} < \nu < \nu_{\rm m}. \tag{105}$$

At high frequencies, $\nu > \nu_{\rm m}$ ($\gamma_e > \gamma_{e,{\rm m}}$), Eq. (99) yields ${\rm d}N/{\rm d}\gamma_e \propto \gamma_e^{-p-2}$, and thus $N(\gamma_e) \propto \gamma_e^{-p-1}$. Setting $q = p + 1$ in Eq. (101), we find

$$P_{\nu,{\rm tot}} \propto \nu^{-p/2}, \quad \nu > \max(\nu_{\rm m}, \nu_{\rm c}). \tag{106}$$

For *slow cooling*, the low-frequency part is again given by Eq. (103). At intermediate frequencies $\nu_{\rm m} < \nu < \nu_{\rm c}$ ($\gamma_{e,{\rm m}} < \gamma_e < \gamma_{e,{\rm c}}$), $N(\gamma_e) = \mathcal{N}(\gamma_e)$, and setting $q = p$ we obtain from Eq. (101):

$$P_{\nu,{\rm tot}} \propto \nu^{-(p-1)/2}, \quad \nu_{\rm m} < \nu < \nu_{\rm c}. \tag{107}$$

At high frequencies $\nu > \nu_{\rm c}$ ($\gamma_e > \gamma_{e,{\rm c}}$), electrons cool fast and particles are injected according to $\mathcal{N} \propto \gamma_e^{-p}$, such that the spectral power is given by Eq. (106).

In summary, and converting to observed total flux $F_{\nu,{\rm tot}} = P_{\nu,{\rm tot}}/4\pi D$, where $D = d - R$ is the distance to the observer (Sec. (2.3.1)), we find for *fast cooling* ($\gamma_{e,{\rm m}} > \gamma_{e,{\rm c}}$)

$$F_{\nu,{\rm tot}} = \begin{cases} (\nu/\nu_{\rm c})^{1/3} F_{\nu,{\rm max}}, & \nu < \nu_{\rm c} \\ (\nu/\nu_{\rm c})^{-1/2} F_{\nu,{\rm max}}, & \nu_{\rm c} < \nu < \nu_{\rm m} \\ (\nu/\nu_{\rm m})^{-p/2} (\nu_{\rm m}/\nu_{\rm c})^{-1/2} F_{\nu,{\rm max}}, & \nu > \nu_{\rm m} \end{cases} \tag{108}$$

and for *slow cooling* ($\gamma_{e,{\rm m}} < \gamma_{e,{\rm c}}$)

$$F_{\nu,{\rm tot}} = \begin{cases} (\nu/\nu_{\rm m})^{1/3} F_{\nu,{\rm max}}, & \nu < \nu_{\rm m} \\ (\nu/\nu_{\rm m})^{-(p-1)/2} F_{\nu,{\rm max}}, & \nu_{\rm m} < \nu < \nu_{\rm c} \\ (\nu/\nu_{\rm c})^{-p/2} (\nu_{\rm c}/\nu_{\rm m})^{-(p-1)/2} F_{\nu,{\rm max}}, & \nu > \nu_{\rm c} \end{cases} . \tag{109}$$

The global maximum of the spectra is determined by the electrons at $\gamma_e = \min(\gamma_{\rm m}, \nu_{\rm c})$, where most of the electron population resides. Making use of Eqs. (93) and (82) the maximum flux is then roughly given by

$$F_{\nu,{\rm max}} = \frac{1}{4\pi D^2} N_e P_{\nu,{\rm max}} = \frac{m_e c^3 \sigma_{\rm T}}{9e} (32\pi m_{\rm p})^{1/2} \epsilon_B^{1/2} n^{3/2} \gamma^2 R^3 D^{-2}, \tag{110}$$

which sets the overall normalization of the spectra. Here, $N_e = 3\pi R^3 n/4\pi$ is the total number of electrons swept up by the blast wave.

2.3.3 Constructing GRB afterglow lightcurves & spectra

In this section, we compute the total synchrotron emission of an ultrarelativistic blast wave by combining the hydrodynamic evolution with the synchrotron spectrum that we computed for a generic shock front in the previous section. Using the hydrodynamic blast wave solution for γ and R as seen by the distant observer (Eqs. (77) and Eq. (78)) in Eqs. (104), (102), and (110), we find that the synchrotron spectrum of the blast wave evolves according to

$$\nu_{\rm c} = \frac{9}{2\sqrt{34}} \frac{m_e e c^{1/2}}{m_{\rm p} \sigma_{\rm T}^2} \epsilon_B^{-3/2} n^{-1} E^{-1/2} t^{-1/2}, \tag{111}$$

$$\nu_{\rm m} = \frac{\sqrt{34}}{\pi} \frac{m_{\rm p}^2 e}{m_e^3 c^{5/2}} \left(\frac{p-2}{p-1}\right)^2 \epsilon_e^2 \epsilon_B^{1/2} E^{1/2} t^{-3/2}, \tag{112}$$

$$F_{\nu,{\rm max}} = \frac{17}{18\sqrt{2\pi}} \frac{m_e c \sigma_{\rm T}}{m_{\rm p}^{1/2} e} \epsilon_B^{1/2} n^{1/2} E D^{-2}, \tag{113}$$

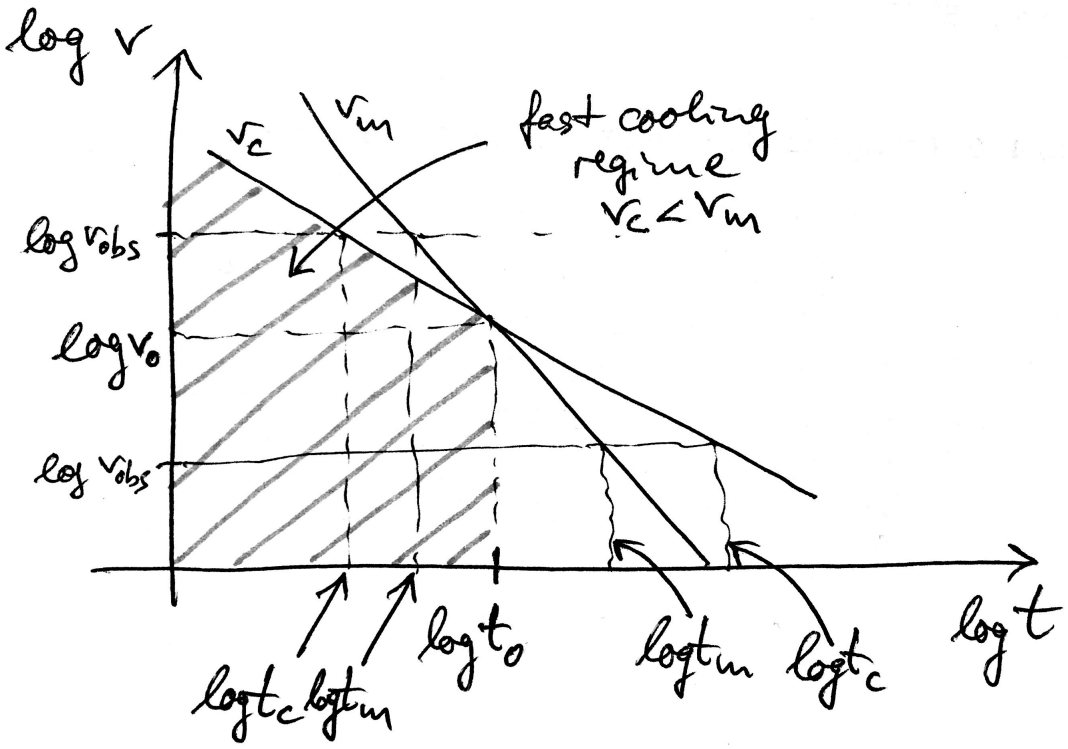

Figure 2: Illustration of the temporal behavior of the characteristic break frequencies $\nu_{\rm c}$ and $\nu_{\rm m}$ as a function of time for an adiabatic ultra-relativistic blast wave, indicating a *high-frequency lighcurve* scencario ($\nu_{\rm obs} > \nu_0$) and a *low-frequency lighcurve* scencario ($\nu_{\rm obs} < \nu_0$), as well as the fast-cooling regime for this parameter space.

where the individual power-law segments of the spectrum are given by Eqs. (108) and (109). For fiducial parameters typical of short GRBs ($E \sim E_{\gamma,\rm iso} \sim 10^{51}$erg, $n \sim 0.1$, $\epsilon_e \sim 0.1$, $\epsilon_B \sim 0.01$, $p \approx 2.4$; see, e.g., Berger 2014), these scalings read

$$\nu_{\rm c} = 8.5\times10^{16}\,{\rm Hz}\,\epsilon_{B,0.01}^{-3/2} n_{0.1}^{-1} E_{51}^{-1/2} t_{\rm d}^{-1/2}, \quad (114)$$

$$\nu_{\rm m} = 2.2\times10^{12}\,{\rm Hz}\left(\frac{p-2}{p-1}\right)^2 \epsilon_{e,0.1}^{2}\epsilon_{B,0.01}^{1/2} E_{51}^{1/2} t_{\rm d}^{-3/2}, \quad (115)$$

$$F_{\nu,\rm max} = 0.35\,{\rm mJy}\,\epsilon_{B,0.01}^{1/2} n_{0.1}^{1/2} E_{51} D_{28}^{-2}. \quad (116)$$

At sufficiently early times, $\nu_{\rm m} > \nu_{\rm c}$ (cf. Eqs. (111) and (112)), i.e., the spectrum is in the fast cooling regime. A transition to the slow cooling spectrum occurs at a time t_0 when $\nu_{\rm m} = \nu_{\rm c}$,

$$t_0 = \frac{68}{9\pi}\left(\frac{p-2}{p-1}\right)^2 \frac{m_{\rm p}^3\sigma_T^2}{m_e^4 c^3}\epsilon_B^2\epsilon_e^2 En \quad (117)$$

$$= 27\,{\rm s}\,\epsilon_{B,0.01}^2\epsilon_{e,0.1}^2 E_{51} n_{0.1}. \quad (118)$$

Let us turn to the temporal behavior of the synchrotron lightcurve. Suppose that the observer detects the blast wave synchrotron radiation in a frequency band around $\nu_{\rm obs}$. The observed lightcurve is then roughly given by $L \simeq \nu_{\rm obs} F_{\nu,\rm tot}(\nu_{\rm obs})$ and will show breaks around characteristic times when the spectral break frequencies cross the observing band, i.e., when

$\nu_{\rm c} = \nu_{\rm obs}$ at

$$t_{\rm c} = \frac{81}{136}\frac{m_e^2 e^2 c}{m_{\rm p}^2\sigma_{\rm T}^4}\epsilon_B^{-3} n^{-2} E^{-1}\nu_{\rm obs}^{-2} \tag{119}$$

$$= 20\,{\rm yr}\,\epsilon_{B,0.01}^{-3} n_{0.1}^{-2} E_{51}^{-1}\nu_{\rm obs,15}^{-2}, \tag{120}$$

and when $\nu_{\rm m} = \nu_{\rm obs}$ at

$$t_{\rm m} = \left(\frac{34}{\pi^2}\right)^{1/3}\frac{m_{\rm p}^{4/3}e^{2/3}}{m_e^2 c^{5/3}}\left(\frac{p-2}{p-1}\right)^{4/3}\epsilon_e^{4/3}\epsilon_B^{1/3}E^{1/3}\nu_{\rm obs}^{-2/3} \tag{121}$$

$$= 2.1\,{\rm min}\left(\frac{p-2}{p-1}\right)^{4/3}\epsilon_{e,0.1}^{4/3}\epsilon_{B,0.01}^{1/3}E_{51}^{1/3}\nu_{\rm obs,15}^{-2/3}. \tag{122}$$

Furthermore, the transition between the fast and slow cooling regime at t_0 gives rise to a critical frequency $\nu_0 = \nu_{\rm c}(t_0) = \nu_{\rm m}(t_0)$,

$$\nu_0 = \left(\frac{9}{2}\right)^{3/2}\frac{\sqrt{\pi}}{34}\frac{m_{\rm e}^3 e c^2}{m_{\rm p}^{5/2}\sigma_{\rm T}^3}\left(\frac{p-2}{p-1}\right)^{-1}\epsilon_e^{-1}\epsilon_B^{-5/2}E^{-1}n^{-3/2} \tag{123}$$

$$= 4.8\times10^{18}\,{\rm Hz}\left(\frac{p-2}{p-1}\right)^{-1}\epsilon_{e,0.1}^{-1}\epsilon_{B,0.01}^{-5/2}E_{51}^{-1}n_{0.1}^{-3/2}. \tag{124}$$

This critical frequency defines two observing scenarios, referred to as the *high-frequency* and *low-frequency lightcurve*. If $\nu_{\rm obs} > \nu_0$, then the only possible sequence of events for this high-frequency lightcurve is $t_{\rm c} < t_{\rm m} < t_0$ (cf. Fig. 2). In contrast, when $\nu_{\rm obs} < \nu_0$ the only possible sequence for the low-frequency lightcurve is $t_0 < t_{\rm m} < t_{\rm c}$. The corresponding power-law segments for $F_{\nu,{\rm tot}}(\nu_{\rm obs})$ from Eqs. (108) and (109) can best be identified visually from Fig. 2. In detail we find for the *high-frequency lightcurve*

$$L \simeq \nu_{\rm obs}F_{\nu,{\rm tot}}(\nu_{\rm obs}) \propto \begin{cases} \nu_{\rm c}^{-1/3}, & \propto t^{1/6}, & t < t_{\rm c} \\ \nu_{\rm c}^{1/2}, & \propto t^{-1/4}, & t_{\rm c} < t < t_{\rm m} \\ \nu_{\rm c}^{1/2}\nu_{\rm m}^{(p-1)/2}, & \propto t^{-(3p-2)/4}, & t > t_{\rm m} \end{cases}, \quad \nu_{\rm obs} > \nu_0, \tag{125}$$

where we have used the scalings of $\nu_{\rm c}$ and $\nu_{\rm m}$ with time (cf. Eqs. (111) and (112)). Analogously, we obtain for the *low-frequency lightcurve*

$$L \simeq \nu_{\rm obs}F_{\nu,{\rm tot}}(\nu_{\rm obs}) \propto \begin{cases} \nu_{\rm c}^{-1/3}, & \propto t^{1/6}, & t < t_0 \\ \nu_{\rm m}^{-1/3}, & \propto t^{1/2}, & t_0 < t < t_{\rm m} \\ \nu_{\rm m}^{(p-1)/2}, & \propto t^{-3(p-1)/4}, & t_{\rm m} < t < t_{\rm c} \\ \nu_{\rm m}^{(p-1)/2}\nu_{\rm c}^{1/2}, & \propto t^{-(3p-2)/4}, & t > t_{\rm c} \end{cases}, \quad \nu_{\rm obs} < \nu_0. \tag{126}$$

For a given wavelength band $\nu_{\rm obs}$, Eqs. (125) and (126) represent the expected temporal evolution of the afterglow lightcurve.

Fitting lightcurves (and/or spectra) as given in Eqs. (125), (126), (108), and (109) together with the scalings of break frequencies and times as in Eqs. (117), (119), (121), (111)–(113) to data of GRB afterglows allows us to infer both macroscopic and microphysical parameters of the GRB. In essence, the free parameters of the model are E, n, ϵ_e, ϵ_B, and p. One can thus obtain macroscopic quantities of the GRB, such as the (kinetic, isotropic-equivalent) blast wave energy, i.e., the energy of the explosion, and the density of the circumburst ISM, but also microphysical quantities, such as the power-law index p of the injected electron spectrum at the shock or the shock energetics in terms of ϵ_e and ϵ_B. We note that the assumption of a spherical blast wave made here is adequate although the early outflow is thought to be jetted, i.e., collimated within some jet opening angle $\theta_{\rm j}$. This is because for an ultra-relativistic blast wave, $\Gamma > 1/\theta_{\rm j}$ at sufficiently early times, and radiation emitted locally is beamed into a cone of opening angle $1/\Gamma$, causally disconnected from other regions

within or exterior to the outflow. This means that the outflow within the jet does not 'know' about its boundary and thus locally behaves like a spherical outflow. Therefore, inferring the properties of the explosion from the afterglow within the formalism described here is accurate.

The typical values for the break frequencies (Eqs. (111) and (112)) show that multi-wavelength observations across the entire electromagnetic spectrum from radio to X-ray frequencies are required in order to cover and probe the afterglow spectrum. While this model has been successfully applied to various long GRBs (e.g., Panaitescu & Kumar 2002; Yost et al. 2003), there is much less data on short GRB afterglows. Only very few examples with comprehensive multi-wavelength afterglows of short GRBs exist, including GRB050724, the first burst with detected radio afterglow (Berger et al. 2005), or GRB130603B (Fong et al. 2014), which was also the first candidate for a kilonova (Tanvir et al. 2013; Berger et al. 2013). Afterglow data (X-ray, in particular) are sometimes not well fit by the model discussed here at early times ($\lesssim$ hours, and modifications are required. This is due to the initial steep decline in X-ray flux caused by the prompt-emission, X-ray flares, and plateaus. Interested readers are referred to, e.g., Berger (2014), Kumar & Zhang (2015), and references therein for more details on these features.

3 Gamma-ray burst central engine

3.1 Overview

When deriving the GRB afterglow model in the previous section, we were agnostic about the central engine, i.e., about what actually generated the explosion energy E of the blast wave in the first place. In this section, I turn to a discussion of mechanisms that can generate such a (initially jetted) relativistic outflow and give rise to a GRB and its afterglow as described by the blast wave solution.

Early on GRBs (both long and short) as described by the blast wave solution were thought of as a relativistic 'fireball', originating from baryon-poor high-entropy material that has been heated up by annihilation of electron and anti-electron neutrinos in the course of a neutron star merger or collapse of a rapidly rotating massive star (e.g., Paczynski 1986; Eichler et al. 1989; Narayan et al. 1992; Woosley 1993; MacFadyen & Woosley 1999). However, more detailed theoretical calculations and simulations showed that the total energy available via annihilation of neutrinos from an accretion disk surrounding the final black hole remnant in these systems is not sufficient to explain the full range of observed (gamma-ray) energies of GRBs (). Instead, electromagnetic extraction of rotational energy from a black hole via the Blandford-Znajek mechanism (Blandford & Znajek 1977) is now widely believed to be one of the most promising mechanisms for powering relativistic outflows from black holes not just in GRBs, but also in various other astrophysical systems such as active galactic nuclei and X-ray binaries. Although GRBs could also be generated by neutron stars (Usov 1992; Thompson 1994; Thompson et al. 2004; Metzger et al. 2011), for the purpose of these lectures we shall focus here on a black hole as the central engine of the GRB and discuss the energy generation via the Blandford-Znajek process. In the future and time provided, I may add a section on neutron stars as central engines; for now I refer the reader to, e.g., Metzger et al. (2011). Observationally, neutron star central engines have been proposed for a subclass of GRBs based on the discovery of long-lasting X-ray plateaus by NASA's Swift mission in a number of long and short GRBs (e.g., Troja et al. 2007; Lyons et al. 2010; Rowlinson et al. 2010, 2013). However, at least in the case of short GRBs, such plateaus could also result from a black hole central engine in the 'time-reversal' scenario (Ciolfi & Siegel 2015; Siegel & Ciolfi 2016a,b).

An excellent and succinct discussion of black hole electrodynamics and the Blandford-Znajek process in a modern 3+1 split approach (but with a neat comparison to the original work by Blandford & Znajek (1977)) can be found in Komissarov (2004, 2009). In order to illustrate how rotational energy can be extracted from a black hole and converted into a flux of electromagnetic energy that can potentially power a magnetically dominated relativistic

outflow, we shall employ a relatively simple argument along the lines of Komissarov (2004, 2009).

3.2 Black hole electrodynamics

3.2.1 Kerr black hole and 3+1 split

The electrodynamics of black holes and their magnetospheres are most conveniently discussed in the framework of a '3+1 split' of spacetime (see Appendix A for a very brief overview). This approach allows us to work with three-dimensional vector fields in very close analogy to electrodynamics in flat spacetime, and another benefit of this approach is that results can easily be carried over to the analysis of numerical simulations, since they are also formulated in a 3+1 decomposition of spacetime.

In the 3+1 split, spacetime is foliated into three-dimensional spatial hypersurfaces that are parametrized by a global time function (or coordinate) t. The general spacetime metric in this approach reads (cf. Eq. (224))

$$\mathrm{d}s^2 = -(\alpha^2 - \beta^2)\mathrm{d}t^2 + 2\beta_i \mathrm{d}x^i \mathrm{d}t + \gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{127}$$

where α is the lapse function, $\boldsymbol{\beta} = \beta^i\partial_i$ is the shift vector, and γ_{ij} is the three-dimensional metric of the spatial hypersurfaces. In this foliation of spacetime there is a special observer, namely the observer who is at rest with respect to the spatial hypersurfaces; this observer referred to as the Eulerian observer moves with 4-velocity n^ν (cf. Eqs. (217) and (218)) perpendicular to the spatial hypersurfaces of constant coordinate time t.

In the following, we shall describe the spacetime of a rotating black hole with mass M and angular momentum J in Boyer-Lindquist coordinates $\{(t, r, \phi, \theta)\}$,

$$\mathrm{d}s^2 = g_{tt}\mathrm{d}t^2 + 2g_{t\phi}\mathrm{d}t\mathrm{d}\phi + \gamma_{rr}\mathrm{d}r^2\gamma_{\phi\phi}\mathrm{d}\phi^2 + \gamma_{\theta\theta}\mathrm{d}\theta^2, \tag{128}$$

where (Boyer & Lindquist 1967)

$$\begin{aligned}
g_{tt} &= (2Mr/\Sigma) - 1, && (129)\\
g_{t\phi} &= -2aMr\sin^2\theta/\Sigma, && (130)\\
\gamma_{rr} &= \Sigma/\Delta, && (131)\\
\gamma_{\phi\phi} &= A\sin^2\theta/\Sigma, && (132)\\
\gamma_{\theta\theta} &= \Sigma, && (133)
\end{aligned}$$

and

$$\begin{aligned}
\Sigma &= r^2 + a^2\cos^2\theta, && (134)\\
\Delta &= r^2 - 2Mr + a^2, && (135)\\
A &= (r^2 + a^2)^2 - a^2\Delta\sin^2\theta, && (136)\\
a &= J/M, \quad -1 < a/M < +1. && (137)
\end{aligned}$$

We note that the horizon of the black hole is located at r_+, the largest root of $\Delta = 0$,

$$r_\pm = M \pm \sqrt{M^2 - a^2}. \tag{138}$$

One can easily show (see Exercise below) that the components of the inverse metric are given by

$$\begin{aligned}
g^{tt} &= -A/(\Delta\Sigma), && (139)\\
g^{t\phi} &= -\frac{2aMr}{\Sigma\Delta}, && (140)\\
\gamma^{rr} &= \Delta/\Sigma, && (141)\\
\gamma^{\phi\phi} &= \frac{\Delta - a^2\sin^2\theta}{\Sigma\Delta\sin^2\theta}, && (142)\\
\gamma^{\theta\theta} &= \Sigma^{-1}. && (143)
\end{aligned}$$

Exercise 3.1 *Derive Eqs. (139)–(143). Hint: Make use of the fact that only the t-ϕ block of $g_{\mu\nu}$ needs to be inverted, as the inversion of the r-θ part is trivial. Proceed by first showing that the determinant of the t-ϕ part is*

$$\tilde{g} = -\Delta \sin^2\theta \tag{144}$$

and then use Cramer's rule. Conclude that

$$g = \det(g_{\mu\nu}) = -\Sigma^2 \sin^2\theta. \tag{145}$$

One can now compare Eqs. (139)–(143) to Eq. (224) and identify the lapse and shift of the Boyer-Lindquist foliation of spacetime:

$$\alpha^2 = -1/g^{tt} = \Delta\Sigma/A, \tag{146}$$
$$\beta^2 = \alpha^2 + g_{tt} = 4a^2r^2\sin^2\theta/A, \tag{147}$$
$$\beta^\phi = \alpha^2 g^{t\phi} = -2aMr/A, \quad \beta^r = \beta^\theta = 0. \tag{148}$$

Frame dragging. For an observer with timelike four-velocity u^μ, the angular velocity is given by

$$\Omega \equiv \frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\frac{\mathrm{d}\phi}{\mathrm{d}\tau}}{\frac{\mathrm{d}t}{\mathrm{d}\tau}} = \frac{u^\phi}{u^t}. \tag{149}$$

This shows that even zero-angular momentum observers ($u_\phi = 0$), such as the Eulerian observer ($u_\phi = n_\phi = 0$, see Eq. (218)), are dragged into co-rotation by the black hole with angular velocity $\Omega \neq 0$.

Exercise 3.2 *Show that for the Eulerian observer (zero-angular momentum observer),*

$$\Omega_{\mathrm{E}} = -\beta^\phi, \quad \textit{i.e.,} \quad \boldsymbol{\beta} = -\Omega_{\mathrm{E}}\partial_\phi. \tag{150}$$

Note that this result is more general than it might look; it holds for any metric of the type (128). In particular, it is also valid for Kerr-Schild coordinates.

Ergosphere. Another interesting feature of the Kerr solution is the existence of an ergosphere. When g_{tt} becomes positive at radii smaller than the largest root r_{S^+} of $g_{tt} = -\alpha^2 + \beta^2 = 0$, where

$$r_{S^\pm} = M \pm \sqrt{M^2 - a^2\cos^2\theta}, \tag{151}$$

the otherwise timelike Killing vector $t^\mu = \partial_t$ is not timelike anymore,

$$g_{\mu\nu}t^\mu t^\nu = g(t^\mu, t^\nu) = g_{tt} > 0. \tag{152}$$

This means that there cannot be a static observer inside this region, referred to as the ergosphere. Any timelike observer (following a timelike world line) and even light is dragged into rotation within the ergosphere.

3.2.2 Electrodynamics in stationary spacetimes

In this section, we briefly sketch the derivation of the foundational equations of electrodynamics in stationary spacetimes, i.e., we shall assume that

$$\partial_t g_{\mu\nu} = 0. \tag{153}$$

While the equations to be derived here will generally apply to any stationary spacetime, they are, in particular, valid for the case of a rotating black hole in Boyer-Lindquist coordinates we are interested in (see previous section). If we define the electric

Electrodynamics is governed by Maxwell's equations, which, in general relativity, can be expressed as

$$\nabla_\nu F^{*\mu\nu} = 0, \tag{154}$$
$$\nabla_\nu F^{\mu\nu} = I^\mu. \tag{155}$$

Here, $F^{\mu\nu}$ is the Maxwell tensor, $F^{*\mu\nu}$ is the Faraday tensor, and I^μ is the electric 4-current (see below). Let us define the magnetic and electric field as seen by the Eulerian observer (projection of the Maxwell and Faraday tensors onto the four-velocity of the observer),

$$D^\mu = -F^{\mu\nu} n_\nu, \tag{156}$$
$$B^\mu = -F^{*\mu\nu} n_\nu, \tag{157}$$

as well as the auxiliary fields

$$E^\mu = -\frac{1}{2}\gamma^{\mu\nu}\eta_{\nu\alpha\beta\gamma}k^\alpha F^{*\beta\gamma}, \tag{158}$$
$$H^\mu = -\frac{1}{2}\gamma^{\mu\nu}\eta_{\nu\alpha\beta\gamma}k^\alpha F^{\beta\gamma}, \tag{159}$$
$$J^\mu = 2I^{[\nu}k^{\mu]}n_\nu, \tag{160}$$
$$\rho_q = -I^\nu n_\nu. \tag{161}$$

Here, $k^\alpha = \partial_t$,

$$\eta^{\mu\nu\lambda\delta} = \frac{1}{\sqrt{-g}}\epsilon^{\mu\nu\lambda\delta} \tag{162}$$

is the Levi-Civita alternating pseudo-tensor of spacetime, and $\epsilon^{\mu\nu\lambda\delta}$ is the four-dimensional Levi-Civita symbol. Note that

$$\eta_{\mu\nu\lambda\delta} = \sqrt{-g}\epsilon_{\mu\nu\lambda\delta}. \tag{163}$$

Exercise 3.3 *Show that the fields* (158)–(160) *are purely spatial, i.e., that they only live in the spatial hypersurfaces*

$$X^\mu n_\mu = 0, \quad X^\mu = D^\mu, B^\mu, E^\mu, H^\mu, J^\mu. \tag{164}$$

Due to the property (164), we can think of $D^\mu, B^\mu, E^\mu, H^\mu, J^\mu$ simply as three-dimensional vectors on the spatial hypersurfaces, $\boldsymbol{D}$, $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{H}$, $\boldsymbol{J}$ whose indices can be raised and lowered with the spatial metric γ_{ij}. Using the definitions (156)–(161), we obtain from the time and spatial components of Eq. (154)

$$\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0 \tag{165}$$

and

$$\partial_t\boldsymbol{B} + \boldsymbol{\nabla}\times\boldsymbol{E} = 0, \tag{166}$$

respectively. Similarly, Eq. (155) yields

$$\boldsymbol{\nabla}\cdot\boldsymbol{D} = \rho_q \tag{167}$$

and

$$-\partial_t\boldsymbol{D} + \boldsymbol{\nabla}\times\boldsymbol{H} = \boldsymbol{J}. \tag{168}$$

Exercise 3.4 *Derive Eqs.* (165)–(168).

In vacuum or in highly ionised plasma the electric and magnetic susceptibilities vanish and the Faraday tensor is dual to the Maxwell tensor,

$$F^{*\mu\nu} = \frac{1}{2}\eta^{\mu\nu\alpha\beta}F_{\alpha\beta}, \quad F^{\mu\nu} = \frac{1}{2}\eta^{\mu\nu\alpha\beta}F^*_{\alpha\beta}. \tag{169}$$

and we can write Eqs. (158) and (159) as

$$\boldsymbol{E} = \alpha\boldsymbol{D} + \boldsymbol{\beta}\times\boldsymbol{B}, \tag{170}$$
$$\boldsymbol{H} = \alpha\boldsymbol{B} - \boldsymbol{\beta}\times\boldsymbol{D}. \tag{171}$$

Exercise 3.5 *Derive Eqs. (170)–(171). Note that $\eta_{\alpha\beta\gamma} = \sqrt{\gamma}\epsilon_{\alpha\beta\gamma}$ is the spatial Levi-Civita pseudo-tensor, $\eta^{\alpha\beta\gamma} = n_\nu \eta^{\nu\alpha\beta\gamma}$.*

Due to the stationary condition (153), we obtain from the time component of the energy-momentum conservation equation,

$$\nabla_\nu T^\nu{}_\mu = -F_{\mu\gamma} I^\gamma, \tag{172}$$

the continuity equation (conservation equation) for the electromagnetic energy flux

$$\partial_t e + \boldsymbol{\nabla} \cdot \boldsymbol{S} = -\boldsymbol{E} \cdot \boldsymbol{J}. \tag{173}$$

Here,

$$T^{\mu\nu} = \frac{1}{4\pi}\left(F^{\mu\lambda}F^\nu{}_\lambda - \frac{1}{4} g^{\mu\nu} F^{\lambda\delta} F_{\lambda\delta}\right), \tag{174}$$

is the electromagnetic energy-momentum tensor and

$$e = n_\nu T^\nu{}_t = \frac{1}{2}(\boldsymbol{E}\cdot\boldsymbol{D} + \boldsymbol{B}\cdot\boldsymbol{H}) \tag{175}$$

and

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H} \tag{176}$$

are, respectively, the energy density and flux of energy at infinity. As mentioned above, $\boldsymbol{E}$ and $\boldsymbol{H}$ are auxiliary fields, as are e and $\boldsymbol{S}$; they only have physical meaning at infinity, where they represent the electric and magnetic fields as well as the energy density and energy flux. The Eulerian observer is a locally inertial observer (at rest wrt to space) and in this observer frame the laws of special relativity apply due to the equivalence principle. In this frame, the physical electric and magnetic fields are $\boldsymbol{D}$ and $\boldsymbol{B}$ and thus this observer will measure the actual physical energy density of the electromagnetic field as

$$e_{\rm E} = \frac{1}{2}(D^2 + B^2) \tag{177}$$

and the associated energy flux as

$$\boldsymbol{S}_{\rm E} = \boldsymbol{D} \times \boldsymbol{B}. \tag{178}$$

3.3 The Blandford-Znajek mechanism

In this section, we shall show that in a

(*i*) stationary,

(*ii*) axisymmetric,

(*iii*) 'force-free'

magnetosphere around a rotating black hole energy (and angular momentum) is extracted along the poloidal magnetic field. This is known as the Blandford-Znajek mechanism. While we shall implicitly assume Boyer-Lindquist coordinates here (Sec. (3.2.1)), the results are also valid for any black hole metric that can be brought into the form (128) with $\partial_t g_{\mu\nu} = \partial_\phi g_{\mu\nu} = 0$; in particular, the results also apply to Kerr-Schild coordinates.

Condition (*iii*) means that we are interested in magnetospheres of black holes with a highly conductive plasma with negligible inertia, i.e., we assume that that the transfer of energy and momentum from the field to the plasma can be neglected (not because the current is negligible, but because the energy and momentum of the electromagnetic field is much larger than that of the plasma). This limit of electrodynamics is called 'force-free electrodynamics'. Note that the four-force describing this transfer of energy and momentum is given by the right-hand side of Eq. (172); the force-free condition thus reads

$$F_{\mu\nu} I^\nu = 0. \tag{179}$$

Making use of Eq. (169) we find

$$\boldsymbol{E}\cdot\boldsymbol{J} = 0, \tag{180}$$
$$\rho\boldsymbol{E}+\boldsymbol{J}\times\boldsymbol{B} = 0. \tag{181}$$

These two relations then imply

$$\boldsymbol{E}\cdot\boldsymbol{B}=0 \tag{182}$$

and with Eq. (170) we immediately obtain

$$\boldsymbol{D}\cdot\boldsymbol{B}=0. \tag{183}$$

From Eq. (180) we conclude that the energy at infinity Eq. (173) is strictly conserved and, using condition (i), we find

$$\boldsymbol{\nabla}\cdot\boldsymbol{S}=0. \tag{184}$$

From condition (ii), we have $\partial_\phi() = 0$. Expressing the Maxwell tensor in terms of the four-potential $\mathcal{A}_\mu = (-\Phi, A_i)$,

$$F_{\mu\nu} = \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu, \tag{185}$$

we find together with condition (i)

$$E_\phi = \partial_\phi\mathcal{A}_t - \partial_t\mathcal{A}_\phi = -\partial_\phi\Phi - \partial_t A_\phi = 0. \tag{186}$$

For further reference, we also note the useful result that inserting $E_\phi = 0$ in Eq. (181) yields $(\boldsymbol{J}\times\boldsymbol{B})_\phi = 0$, i.e.,

$$\boldsymbol{J}_{\rm p} \parallel \boldsymbol{B}_{\rm p}, \tag{187}$$

where we have defined the poloidal component $\boldsymbol{X}_{\rm p}$ of a vector $\boldsymbol{X}$ by

$$\boldsymbol{X} = \boldsymbol{X}_{\rm p} + X_\phi\partial_\phi, \tag{188}$$

with $\boldsymbol{X}_{\rm a} = X^\phi\partial_\phi$ being the azimuthal component.

From Eqs. (182) and (186) we conclude that there exists a vector $\boldsymbol{\omega} = \Omega\partial_\phi$, such that

$$\boldsymbol{E} = -\boldsymbol{\omega}\times\boldsymbol{B} \tag{189}$$

and Eq. (170) translates into

$$\boldsymbol{D} = -\frac{1}{\alpha}(\boldsymbol{\omega}+\boldsymbol{\beta})\times\boldsymbol{B} \tag{190}$$

Furthermore, employing Eq. (189) in Eq. (166) together with condition (i) yields

$$\boldsymbol{\nabla}\times(\boldsymbol{\omega}\times\boldsymbol{B}) = 0, \tag{191}$$

which implies

$$\boldsymbol{B}\cdot\nabla\Omega = 0. \tag{192}$$

The latter equation shows that Ω is constant along the magnetic field, which is thus called angular velocity of the magnetic field lines.

Exercise 3.6 *Using Eqs.* (189), (190), (176), (178), (171), *and* (150), *show that*

$$\boldsymbol{S}_{\rm p} = -\Omega H_\phi\boldsymbol{B}_{\rm p}, \tag{193}$$
$$\boldsymbol{S}_{\rm E,p} = -\frac{1}{\alpha^2}(\Omega-\Omega_{\rm E})H_\phi\boldsymbol{B}_{\rm p}. \tag{194}$$

While $\boldsymbol{S}_{\rm E,p}$ represents the direction of electromagnetic field flow (also called electromagnetic wind), $\boldsymbol{S}_{\rm p}$ can be interpreted as the direction of energy flow. We thus conclude from Eq. (193) that electromagnetic energy is transported along poloidal magnetic field lines, i.e., poloidal magnetic fields 'extract' rotational energy from the black hole. We shall interpret and discuss this result in more detail below (Sec. 3.4).

For further reference, let us finally point out a peculiarity associated with the Poynting vectors obtained above. They are anti-parallel if

$$0 < \Omega < \Omega_{\rm E}. \tag{195}$$

In this regime of 'energy counter-flow', energy flows in the opposite direction to that of the electromagnetic fields as seen by the Eulerian observer. Furthermore, we note that using Eqs. (189) and (190), the energy density at infinity (158) can be written as

$$e = \frac{1}{2\alpha}[\alpha^2 B^2 + B_{\rm p}^2(\omega^2 - \beta^2)], \tag{196}$$

which is negative if

$$\alpha^2 B^2 + B_{\rm p}^2(\omega^2 - \beta^2) < 0. \tag{197}$$

This condition can be rewritten in two equivalent ways:

$$\Omega^2 < \Omega_{\rm E}^2 - \frac{\alpha^2 B^2}{\gamma_{\phi\phi} B_{\rm p}^2}, \tag{198}$$

$$\alpha^2 - \beta^2 < -\frac{\alpha^2 B_{\rm a}^2 + \omega^2 B_{\rm p}^2}{B_{\rm p}^2}, \tag{199}$$

where $\boldsymbol{B}_{\rm a} = B^\phi \partial_\phi$ is the azimuthal component of $\boldsymbol{B}$. The relation (198) shows that the condition is more restrictive than the energy counter-flow condition (195), i.e., there can be regions with energy counter-flow but positive energy at infinity. Additionally, the relation (199) shows that negative energy at infinity is restricted to $\beta^2 - \alpha^2 = g_{tt} > 0$, i.e., to within the ergosphere (cf. Eq. (152)).

3.4 The Blandford-Znajek mechanism: the electromotive force for energy extraction

Due to the no hair theorem black holes cannot sustain their own magnetic field; any magnetic field penetrating the ergosphere or horizon has to be generated by external currents. In this section, we discuss why and how external currents are generated in the black hole magnetosphere that lead to the extraction of energy and angular momentum from the black hole as found in the previous section (Eq. (193)).

We start by noting that for a steady-state axisymmetric vacuum magnetosphere (cf. Eqs. (166), (168)),

$$\boldsymbol{\nabla} \times \boldsymbol{H} = \boldsymbol{\nabla} \times \boldsymbol{E} = 0. \tag{200}$$

Therefore, $\boldsymbol{H}$ and $\boldsymbol{E}$ can be written as gradients of a scalar function, $\boldsymbol{H} = \nabla\Psi$, $\boldsymbol{H} = -\nabla\Phi$, where Φ is the electric potential (see also Eq. (186)). Axisymmetry then implies

$$H_\phi = E_\phi = 0. \tag{201}$$

From $H_\phi = 0$ it also follows that $B^\phi = 0$ (cf. Eq. (171)), i.e., the magnetic field is purely poloidal. Together with Eq. (193) this shows that in vacuum (absence of charges) stationary axisymmetric electromagnetic fields cannot extract energy from the black hole.

This seems to suggest that the critical additional ingredient in order to extract energy from the black hole is the presence of charges that are able to drive currents in the magnetosphere. However, in order to to so, electric fields are required. Additionally, creating such charges in the first place (if not present initially) may require strong electric and radiation fields, e.g., in order to initiate a e^+-e^- pair cascade (Beskin et al. 1992). Such electric fields are generated by the frame dragging effect, as one can see by combining the general equations (166) and (170) into

$$\boldsymbol{\nabla} \times (\alpha \boldsymbol{D}) = -\boldsymbol{\nabla} \times (\boldsymbol{\beta} \times \boldsymbol{B}), \tag{202}$$

where the shift vector conspires with the external magnetic field to provide a source for the electric field. The Eulerian observer will therefore see both the external magnetic field $\boldsymbol{B}$ and

the induced electric field $\boldsymbol{D}$. This, however, raises the question of whether such an electric field can be sustained in a stationary scenario, i.e., whether or not this electric field would lead to a separation of charges in such a way as to establish a counter field that effectively screens the actual electric field and thus prevents any currents from flowing and thus energy extraction (a 'dead' magnetosphere). Total screening implies that there be no component of the electric field along the magnetic field and that the magnetic field dominates,

$$\boldsymbol{D}\cdot\boldsymbol{B}=0,\quad B^2-D^2>0. \tag{203}$$

We will now show that these conditions cannot be simultaneously satisfied throughout the entire magnetosphere. Assume that $\boldsymbol{D}\cdot\boldsymbol{B}=0$. Inserting this into Eq. (170), we also find $\boldsymbol{E}\cdot\boldsymbol{B}=0$. Then from axisymmetry (Eq. (201)), it follows that Eqs. (189) and (190) hold.

Exercise 3.7 *Show that squaring Eq.* (190) *yields*

$$(B^2-D^2)\alpha^2=B^2 f(\Omega,r,\theta)+(\omega+\beta^\phi)^2\frac{H_\phi^2}{\alpha^2}, \tag{204}$$

where

$$f(\Omega,r,\theta)=\alpha^2-(\boldsymbol{\omega}+\boldsymbol{\beta})^2. \tag{205}$$

For an entirely screened ('dead') magnetosphere, $H_\phi=0$ (cf. Eqs. (193), (201), (200)), and thus Eq. (204) reads

$$(B^2-D^2)\alpha^2=B^2 f(\Omega,r,\theta). \tag{206}$$

From Eqs. (205) and (206) it follows that the sign of B^2-D^2 does not depend on B, it is only controlled by the spacetime and the angular velocity of the magnetic field lines Ω. This is because a stronger magnetic field generates a stronger electric field (Eq. (202)). Furthermore, we note that far away from the black hole $r\to\infty$, $a\to 1$, $\boldsymbol{\beta}\to 0$, and thus with Eq. (132) we obtain

$$f(\Omega,r,\theta)\to 1-\omega^2=1-\Omega^2\gamma_{\phi\phi}=1-\Omega^2 r^2\sin^2\theta. \tag{207}$$

Therefore, in order to satisfy the inequality in Eq. (203), it follows that $\Omega=0$. This, in turn, implies that

$$f(\Omega,r,\theta)=\alpha^2-\beta^2, \tag{208}$$

which is negative inside the ergosphere ($g_{tt}>0$; see Eq. (152)). This shows that because of the existence of the ergosphere, both conditions in Eq. (203) cannot be satisfied simultaneously throughout the entire magnetosphere and thus the magnetosphere cannot be dead, provided sufficient charges are present. We note that this argument makes two implicit assumptions. First, it assumes that a large-scale—purely poloidal as $H_\phi=0$—magnetic field exists and that it penetrates the ergosphere, which is a reasonable assumption in steady state. Second, it assumes that Ω is constant along the magnetic field lines; this follows from Eq. (192), which also applies here.

In conclusion, H_ϕ must not vanish in order for charges to be able to screen the electric field, which is induced by the externally applied magnetic field (cf. Eq. (204)). According to the induction equation (168), in steady-state we have

$$\boldsymbol{\nabla}\times\boldsymbol{H}=\boldsymbol{J}. \tag{209}$$

Therefore, $H^\phi\neq 0$ implies a poloidal current $\boldsymbol{J}_\mathrm{p}$ along the poloidal magnetic field $\boldsymbol{B}_\mathrm{p}$ (cf. Eq. (187)) that is driven by the charges in the attempt to screen the electric field. Since $\boldsymbol{\nabla}\times(\boldsymbol{\beta}\times\boldsymbol{B}_\mathrm{a})=0$ in axisymmetry, the azimuthal magnetic field $\boldsymbol{B}_\mathrm{a}$ induced by this poloidal current does not amplify the electric field (cf. Eq. (202)), i.e., it solely acts to screen it. However, in order to sustain this current, the electric field must not be screened entirely. This shows that there will be a steady-state non-vanishing poloidal current that mediates energy and angular momentum extraction from the black hole to infinity due to $H_\phi\neq 0$ (cf. Eq.(193)). It is precisely this electromotive force that promotes the energy extraction found in the previous section.

3.5 Activation of the Blandford-Znajek mechanism in GRBs

A 3+1 decomposition of spacetime

For many applications including analytic arguments (see, e.g., Sec. 3.2) and performing general-relativistic numerical simulations, it is convenient to foliate spacetime $(\mathcal{M}, g_{\mu\nu})$ into non-intersecting spacelike 3-surfaces Σ_ψ, which arise as the level sets of a scalar function $\psi : p \in \mathcal{M} \to \psi$ on the spacetime $\mathcal{M}$ that can be interpreted as a global time function. This function defines a one-form $\nabla_\mu\psi$, which, in turn, defines a vector field (its dual), ψ^μ, such that $\psi^\mu\nabla_\mu\psi = 1$. We assume that $\nabla^\mu\psi$ is timelike, i.e., $g(\nabla^\mu\psi, \nabla^\mu\psi) = g_{\mu\nu}\nabla^\mu\psi\nabla^\nu\psi < 0$. The timelike unit normal n^μ of the spatial hypersurfaces of $\psi = \text{const}$ can then be expressed as

$$n^\mu = -\frac{g^{\mu\nu}\nabla_\nu\psi}{\sqrt{-g^{\sigma\delta}\nabla_\sigma\psi\nabla_\delta\psi}}, \tag{210}$$

and one can decompose ψ^μ into a component along n^μ and a component in Σ_ψ,

$$\psi^\mu = \alpha n^\mu + \beta^\mu, \qquad n_\mu\beta^\mu = 0. \tag{211}$$

The duality condition $\psi^\mu\nabla_\mu\psi = 1$ shows that

$$\alpha = \frac{1}{\sqrt{-g^{\sigma\delta}\nabla_\sigma\psi\nabla_\delta\psi}}. \tag{212}$$

Here, the function α is called the *lapse function*, with $\alpha\mathrm{d}\psi$ being the proper time elapsed in normal direction (Eulerian observer) between coordinate times ψ and $\psi + \mathrm{d}\psi$, i.e., between the spatial slices Σ_ψ and $\Sigma_{\psi+\mathrm{d}\psi}$. The vector field β^μ is called the *shift vector*, since $\beta^\mu\mathrm{d}\psi$ measures the spatial shift between a path of coordinate length $\mathrm{d}\psi$ along n^μ (Eulerian observer) and a path of coordinate length $\mathrm{d}\psi$ along ψ^μ. Furthermore, the 4-metric $g_{\mu\nu}$ induces a *spatial metric* $\gamma_{\mu\nu}$ on the spatial hypersurfaces Σ_ψ,

$$\gamma_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu, \tag{213}$$

which entirely resides on Σ_ψ,

$$n^\mu\gamma_{\mu\nu} = n^\mu g_{\mu\nu} + n^\mu n_\mu n_\nu = n_\nu - n_\nu = 0. \tag{214}$$

Now let us choose coordinates such that $\psi^\mu = \partial_0$, with components $\delta^\mu_0 = e^\mu_0 = (1, 0, 0, 0)$ in the local map. Then its dual $\nabla_\mu\psi$ has components $\delta^0_\mu = e^0_\mu = (1, 0, 0, 0)$ in this map. One can define the other basis vector fields by completing $\{\partial^\mu_0\}$ to a set of orthonormal basis vectors $\{\partial^\mu_0, \partial^\mu_i|_{\Psi=\Psi_0}\}$ on some initial time slice $\Sigma_{\psi=\psi_0}$, and then Lie-dragging those basis vector fields along ψ^μ,

$$\partial^\mu_i \equiv \mathcal{L}_\psi(\partial^\mu_i|_{\psi=\psi_0}). \tag{215}$$

This method yields a local set of basis vector fields for the spacetime, since the Lie derivative conserves orthogonality.

Since, in this map, the basis vector fields were constructed such that $0 = \nabla_\mu\psi\partial^\mu_i \propto n_\mu\partial^\mu_i$, n_μ has spatial components $n_i = 0$ in this map. From the requirement $n_\mu\beta^\mu = 0$ we deduce that the shift vector has components

$$\beta^\mu = (0, \beta^i) \tag{216}$$

in this map, and it immediately follows from Eq. (211) that the normal vector has components

$$n^\mu = (\alpha^{-1}, -\alpha^{-1}\beta^i). \tag{217}$$

Using $n^\mu n_\mu = -1$, one finds the components of n_μ,

$$n_\mu = (-\alpha, 0, 0, 0), \tag{218}$$

which together with Eq. (213) immediately show that the spatial components of the 3-metric are given by

$$\gamma_{ij} = g_{ij}. \tag{219}$$

Furthermore, we note that

$$\begin{aligned} g_{tt} &= \psi_\mu \psi^\mu = \alpha^2 n_\mu n^\mu + \beta_\mu \beta^\mu = -\alpha^2 + \beta_k \beta^k, && (220)\\ g_{ti} &= \psi_\mu \partial_i^\mu = \beta_\mu \partial_i^\mu = \beta_i && (221)\\ g^{tt} &= \nabla_\mu \psi \nabla^\mu \psi = -\alpha^{-2} && (222)\\ g^{ti} &= -n^0 n^i = \alpha^{-2} \beta^i, && (223) \end{aligned}$$

where the last identity follows from the fact that $\gamma^{\mu 0} = 0$ as implied by Eqs. (213), (214), and (217). In summary, we find the following coordinate representation of the 4-metric and its inverse:

$$g_{\mu\nu} = \begin{pmatrix} -\alpha^2 + \beta_k \beta^k & \beta_i \\ \beta_j & \gamma_{ij} \end{pmatrix}, \quad g^{\mu\nu} = \begin{pmatrix} -\alpha^{-2} & \alpha^{-2}\beta^i \\ \alpha^{-2}\beta^i & \gamma^{ij} - \alpha^{-2}\beta^i\beta^j \end{pmatrix}. \tag{224}$$

Additionally, we find that

$$\sqrt{-g} = \alpha\sqrt{\gamma}, \tag{225}$$

where $g \equiv \det(g_{\mu\nu})$ and $\gamma \equiv \det(\gamma_{ij})$. We note that the lapse α and the three non-vanishing components β^i of the shift vector can be chosen arbitrarily. They represent four degrees of freedom, which reflect the diffeomorphism invariance of general relativity, i.e., the freedom to choose (local) coordinates of the spacetime.

References

Abramowitz, M. & Stegun, I. A. 1972, Handbook of mathematical functions with formulas, graphs and mathematical tables (Washington D.C., USA: National Bureau of Standards)

Arnett, W. D. 1979, ApJL, 230, L37

Arnett, W. D. 1982, ApJ, 253, 785

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

Berger, E. 2014, Annu. Rev. Astron. Astrophys., 52, 43

Berger, E., Fong, W., & Chornock, R. 2013, ApJL, 774, L23

Berger, E., Price, P. A., Cenko, S. B., et al. 2005, Nature, 438, 988

Beskin, V. S., Istomin, Y. N., & Par'ev, V. I. 1992, Soviet Astronomy, 36, 642

Blandford, R. D. & McKee, C. F. 1976, Phys. Fluids, 19, 1130

Blandford, R. D. & Znajek, R. L. 1977, MNRAS, 179, 433

Boyer, R. H. & Lindquist, R. W. 1967, J. Math. Phys., 8, 265

Burbidge, E. M., Burbidge, G. R., Fowler, W. A., & Hoyle, F. 1957, Rev. Mod. Phys., 29, 547

Cameron, A. G. W. 1957, AJ, 62, 9

Ciolfi, R. & Siegel, D. M. 2015, ApJL, 798, L36

Crusius, A. & Schlickeiser, R. 1986, A&A, 164, L16

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

Eichler, M., Arcones, A., Kelic, A., et al. 2015, Astrophys. J., 808, 30

Fong, W., Berger, E., Metzger, B. D., et al. 2014, ApJ, 780, 118

Foucart, F. 2012, Phys. Rev. D, 86, 124007

Granot, J. & Sari, R. 2002, ApJ, 568, 820

Hotokezaka, K., Wanajo, S., Tanaka, M., et al. 2016, Mon. Not. R. Astron. Soc., 459, 35

Kasen, D. & Bildsten, L. 2010, Astrophys. J., 717, 245

Kasliwal, M. M., Nakar, E., Singer, L. P., et al. 2017, Science, 358, 1559

Komissarov, S. S. 2004, Mon. Not. R. Astron. Soc., 350, 427

Komissarov, S. S. 2009, Journal of Korean Physical Society, 54, 2503

Korobkin, O., Rosswog, S., Arcones, A., & Winteler, C. 2012, MNRAS, 426, 1940

Kumar, P. & Zhang, B. 2015, Phys. Rep., 561, 1

Li, L.-X. & Paczyński, B. 1998, ApJL, 507, L59

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

Lyons, N., O'Brien, P. T., Zhang, B., et al. 2010, MNRAS, 402, 705

MacFadyen, A. I. & Woosley, S. E. 1999, Astrop. J., 524, 262

Marti, J. M., Mueller, E., & Ibanez, J. M. 1994, Astron. Astrophys., 281, L9

Metzger, B. D. 2017, Living Reviews in Relativity, 20, 3

Metzger, B. D., Giannios, D., Thompson, T. A., Bucciantini, N., & Quataert, E. 2011, MNRAS, 413, 2031

Metzger, B. D., Martínez-Pinedo, G., Darbha, S., et al. 2010, MNRAS, 406, 2650

Mumpower, M. R., Surman, R., McLaughlin, G. C., & Aprahamian, A. 2016, Prog. Part. Nucl. Phys., 86, 86

Narayan, R., Paczynski, B., & Piran, T. 1992, ApJL, 395, L83

Paczynski, B. 1986, ApJL, 308, L43

Panaitescu, A. & Kumar, P. 2002, ApJ, 571, 779

Piro, A. L. & Kollmeier, J. A. 2017, ArXiv e-prints

Roberts, L. F., Kasen, D., Lee, W. H., & Ramirez-Ruiz, E. 2011, ApJL, 736, L21

Rosswog, S., Feindt, U., Korobkin, O., et al. 2017, Classical and Quantum Gravity, 34, 104001

Rowlinson, A., O'Brien, P. T., Metzger, B. D., Tanvir, N. R., & Levan, A. J. 2013, MNRAS, 430, 1061

Rowlinson, A., O'Brien, P. T., Tanvir, N. R., et al. 2010, MNRAS, 409, 531

Rybicki, G. B. & Lightman, A. P. 2004, Radiative Processes in Astrophysics (Wiley-VCH)

Sari, R., Piran, T., & Narayan, R. 1998, ApJL, 497, L17

Sedov, L. I. 1959, Similarity and Dimensional Methods in Mechanics

Siegel, D. M. & Ciolfi, R. 2016a, ApJ, 819, 14

Siegel, D. M. & Ciolfi, R. 2016b, ApJ, 819, 15

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

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

Tanvir, N. R., Levan, A. J., Fruchter, A. S., et al. 2013, Nature, 500, 547

Taub, A. H. 1948, Phys. Rev., 74, 328

Taylor, G. 1950, Proc. R. Soc. London Ser. A, 201, 159

Thompson, C. 1994, MNRAS, 270, 480

Thompson, T. A., Chang, P., & Quataert, E. 2004, ApJ, 611, 380

Troja, E. et al. 2007, ApJ, 665, 599

Usov, V. V. 1992, Nature, 357, 472

Villar, V. A., Guillochon, J., Berger, E., et al. 2017, ApJL, 851, L21

Woosley, S. E. 1993, ApJ, 405, 273

Wu, M.-R., Barnes, J., Martinez-Pinedo, G., & Metzger, B. D. 2018, arXiv e-prints

Yost, S. A., Harrison, F. A., Sari, R., & Frail, D. A. 2003, ApJ, 597, 459