3 Relativity

GRBs are relativistic phenomena. GRB ejecta are believed to move towards Earth with a Lorentz factor typically greater than 100. Special relativity is at the core of GRB theory. This chapter starts with an overview of the principles of special relativity (§3.1). Next, the basic space-time definitions are clarified within the GRB context (§3.2), and the relevant Doppler transformations in relativistic systems are summarized (§3.3). Section 3.4 discusses how the observed flux depends on the intrinsic emissivity in the comoving frame, for both the continuous emission case and when the emitter stops shining suddenly (the so-called curvature effect). Some other useful formulae to study GRB problems are derived in §3.5. Finally, in §3.6, basic principles of general relativity are reviewed without going into the tedious mathematical details. For a more complete introduction to relativity, see standard textbooks, e.g. *Gravitation*, by Misner, Thorne, and Wheeler (Misner et al., 1973).

3.1 Special Relativity and Lorentz Transformation

3.1.1 Postulates

Special relativity is based on the following two postulates (Einstein, 1905).

- *Special principle of relativity*: If physical laws hold good in their simplest form in one inertial frame K, the same laws also hold good in any other inertial frame K' that moves with a constant velocity with respect to K.
- *Invariance of c*: As measured in any inertial frame of reference, light traveling in empty space always propagates with a definite speed c that is independent of the state of motion of the emitting body and the receiving body.

3.1.2 Lorentz Transformation

Consider an *inertial frame* of reference K, in which an *event* is recorded as a four-dimensional vector (t, x, y, z), where t is the time and (x, y, z) is the three-dimensional spatial coordinate of the event. A second inertial frame K' is moving with respect to K with a speed v. Without losing generality, one can make the positive direction of the x-axis as the direction of motion. The same event is then recorded in the frame K' as

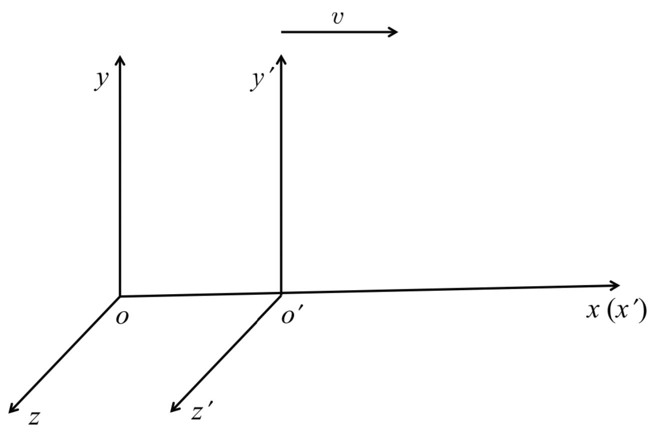

Figure 3.1 Two inertial frames, with K' moving along the x-axis with speed v with respect to K.

(t', x', y', z') (Fig. 3.1). Let's also assume that the origins of the two frames coincide, i.e. $(t', x', y', z') = (0, 0, 0, 0)$ when $(t, x, y, z) = (0, 0, 0, 0)$. The *Lorentz transformation* can be written as

$$\begin{cases} t' &= \gamma(t - \frac{\beta x}{c}) \\ x' &= \gamma(x - \beta ct) \\ y' &= y \\ z' &= z \end{cases} \tag{3.1}$$

or

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \tag{3.2}$$

Here $\beta = v/c$ is the relative *dimensionless speed* between the two frames, and

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{3.3}$$

is the relative *Lorentz factor*. In Eqs. (3.2), we have defined the time element as ct, so that all four elements of the vector have the dimension of length.

Since the "zero points" of both space and time are defined arbitrarily, in general, it is more fundamental to write Eqs. (3.1) and (3.2) in their differential form, i.e.

$$\begin{cases} dt' &= \gamma(dt - \frac{\beta dx}{c}) \\ dx' &= \gamma(dx - \beta cdt) \\ dy' &= dy \\ dz' &= dz \end{cases} \tag{3.4}$$

or

$$\begin{pmatrix} cdt' \\ dx' \\ dy' \\ dz' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} cdt \\ dx \\ dy \\ dz \end{pmatrix}. \tag{3.5}$$

By doing so, no requirement for the zero point coincidence is needed.

The two postulates of special relativity can be absorbed into the requirement of the invariance of the *space-time interval*[1]

$$\begin{aligned} ds^2 &= -c^2dt^2 + dx^2 + dy^2 + dz^2 \\ &= -c^2dt'^2 + dx'^2 + dy'^2 + dz'^2 \end{aligned} \tag{3.6}$$

for the two different inertial frames. This requirement leads to a unique transformation between the two frames through the Lorentz transformation (Exercises 3.1 and 3.2).

In the rest frame K', K is moving with the same speed v but in the opposite direction. The *inverse Lorentz transformation* (in differential form) then reads

$$\begin{cases} dt &= \gamma(dt' + \frac{\beta dx'}{c}) \\ dx &= \gamma(dx' + \beta c dt') \\ dy &= dy' \\ dz &= dz' \end{cases} \tag{3.7}$$

or

$$\begin{pmatrix} cdt \\ dx \\ dy \\ dz \end{pmatrix} = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} cdt' \\ dx' \\ dy' \\ dz' \end{pmatrix}. \tag{3.8}$$

3.1.3 Length Contraction and Time Dilation

Length Contraction

In relativity, the *intrinsic length* of a rod is defined as the distance between the two ends of the rod at a same time in a particular inertial frame. Practically, simultaneity of measurements can be guaranteed by placing many precisely pre-tuned clocks at different locations along the path of the rod and by recording the coordinates of each end of the rod at a particular epoch.

Suppose a rod is moving (with a comoving frame K') along the x-axis with respect to the "lab frame" K. Then, in the lab frame, the rod length is defined as

$$\Delta x = x_2 - x_1, \tag{3.9}$$

[1] Throughout the book, the "$-+++$" convention is adopted. Another convention defines $ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$. The two conventions are mathematically equivalent.

where x_1 and x_2 are the 1-D spatial coordinates of the two ends of the rod. According to the Lorentz transformation and taking $t_1 = t_2$, the comoving length is

$$\begin{aligned}\Delta x' &= x'_2 - x'_1 \\ &= \gamma(x_2 - \beta ct_2) - \gamma(x_1 - \beta ct_1) \\ &= \gamma(x_2 - x_1) \\ &= \gamma \Delta x,\end{aligned} \tag{3.10}$$

so that the intrinsic length in the lab frame contracts by a factor of γ:

$$\Delta x = \frac{\Delta x'}{\gamma}\ . \tag{3.11}$$

This can also be derived more straightforwardly from the second equation of the differential Lorentz transformation (Eq. (3.4)), noting $dt = 0$.

Notice that the *observed length* is defined as the distance between the two ends of the rod at the same observer time. This is discussed in §3.3.3.

Time Dilation

In relativity, a *time interval* between *two* events is also relative. In the comoving frame, the time interval can be defined as the elapsed time at a fixed point ($x'_2 = x'_1$), so that

$$\Delta t' = t'_2 - t'_1\ . \tag{3.12}$$

In the lab frame, the time interval is

$$\begin{aligned}\Delta t &= t_2 - t_1 \\ &= \gamma(t'_2 + \beta x'_2/c) - \gamma(t'_1 + \beta x'_1/c) \\ &= \gamma(t'_2 - t'_1) \\ &= \gamma \Delta t'.\end{aligned} \tag{3.13}$$

So the lab-frame time is dilated by a factor γ. This can be also derived more straightforwardly from the first equation of the differential Lorentz transformation (Eq. (3.7)), noting $dx' = 0$.

3.1.4 Relativistic Velocity Transformation

Let us now consider the motion of an object. Suppose it moves along the x-axis with speed $u' = \frac{dx'}{dt'}$ and $u = \frac{dx}{dt}$ in the comoving frame K' and lab frame K, respectively. With Eq. (3.4), one has

$$u' = \frac{dx'}{dt'} = \frac{\gamma(dx - vdt)}{\gamma(dt - \frac{vdx}{c^2})} = \frac{u - v}{1 - \frac{uv}{c^2}}\ . \tag{3.14}$$

Similarly, one can derive the inverse transformation

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}}\ . \tag{3.15}$$

More generally, the velocity $\mathbf{u}$ of the moving object can have an arbitrary angle with respect to $\mathbf{v}$, say, θ in the lab frame and θ' in the comoving frame. Without losing generality, one can demand that the vector $\mathbf{u}$ (and $\mathbf{u}'$) is in the x–y plane. Noting $dy = dy'$, one has

$$u'_{\parallel} = \frac{dx'}{dt'} = \frac{u_{\parallel} - v}{1 - \frac{u_{\parallel} v}{c^2}}, \tag{3.16}$$

$$u'_{\perp} = \frac{dy'}{dt'} = \frac{u_{\perp}}{\gamma(1 - \frac{u_{\parallel} v}{c^2})}, \tag{3.17}$$

and the inverse transformation:

$$u_{\parallel} = \frac{dx}{dt} = \frac{u'_{\parallel} + v}{1 + \frac{u'_{\parallel} v}{c^2}}, \tag{3.18}$$

$$u_{\perp} = \frac{dy}{dt} = \frac{u'_{\perp}}{\gamma(1 + \frac{u'_{\parallel} v}{c^2})}. \tag{3.19}$$

The comoving-frame angle θ' between $\mathbf{u}'$ and $\mathbf{e_{x'}}$, and the lab-frame angle between $\mathbf{u}$ and $\mathbf{e_x}$ are related to each other through

$$\tan\theta = \frac{u_{\perp}}{u_{\parallel}} = \frac{u'_{\perp}}{\gamma(u'_{\parallel} + v)} = \frac{u' \sin\theta'}{\gamma(u' \cos\theta' + v)}. \tag{3.20}$$

Here $\mathbf{e_{x'}} = \mathbf{e_x}$ are the unit vectors in the x-axis direction in both frames.

When setting $u' = c$, one can derive useful expressions for the *aberration of light*:

$$\tan\theta = \frac{\sin\theta'}{\gamma(\cos\theta' + \beta)}, \tag{3.21}$$

$$\cos\theta = \frac{\cos\theta' + \beta}{1 + \beta\cos\theta'}, \tag{3.22}$$

$$\sin\theta = \frac{\sin\theta'}{\gamma(1 + \beta\cos\theta')}. \tag{3.23}$$

When $\theta' = \pi/2$, one has $\tan\theta = 1/\gamma\beta$, $\cos\theta = \beta$, and $\sin\theta = 1/\gamma$. When $\gamma \gg 1$, one obtains the familiar result $\theta \sim 1/\gamma$.

3.2 The GRB Problem: Rest Frames and Times

3.2.1 Rest Frames

A GRB jet moves towards Earth with a relativistic speed. The detailed reasoning for this conclusion will be summarized in §7.1.

There are three rest frames of reference in a GRB problem:

- Frame I: the rest frame of the central engine (the laboratory frame);
- Frame II: the rest frame of the relativistic ejecta (or flying shells);
- Frame III: the rest frame of the observer.

Even though special relativity states that all inertial frames are equivalent, due to the expansion of the universe, in cosmology a set of inertial frames are special. At a particular redshift z, such a special inertial frame can be defined as the one that is at rest with respect to the cosmic background radiation, i.e. in which the cosmic background radiation is observed as isotropic. Let us call these inertial frames the *cosmic proper frames*.

Since the GRB central engine (a hyper-accreting black hole or a millisecond magnetar) is at rest with respect to the cosmic proper frame at the source redshift z, and since we observers are at rest with the current cosmic proper frame,[2] Frames I and III are related to each other only through a cosmic expansion factor $(1 + z)$. As a result, the time in Frame III is stretched by a factor $(1 + z)$ with respect to that in Frame I, and the photon energy in Frame III is lower by a factor $(1+z)$ than the one in Frame I. Since the $(1+z)$ factor is much smaller than the bulk motion Lorentz factor Γ, for simple derivations, in the following we first neglect cosmic expansion, and regard I and III as the same inertial frame, in which the GRB ejecta is moving with a Lorentz factor Γ. The clocks in this frame can be universally tuned.

As will be evident soon, within this inertial frame, the source and the observer can still define the space-time quantities (e.g. time interval, rod length) differently. This is strictly due to a propagation effect. For example, while the universally tuned clocks measure the time interval between *emitting* two signals by the source, the observer time measures the time interval between *receiving* those two signals. As a result, in the literature, an *observer frame* is still defined to be differentiated from the source frame. Physically the two "frames" are the same inertial rest frame. The reason for different measured times is because one is concerned with the time intervals of two different pairs of *events* (emitting vs. receiving the two signals). In the following, we follow these conventions and define Frame I as the source frame or the *laboratory frame*, and Frame III as the observer frame, keeping in mind that the difference between these two so-called "frames" is attributed purely to the propagation effect.

3.2.2 Times

With the three frames defined, one has four relevant times in the GRB problem:

- the central engine time $t_{\rm eng}$ measured in Frame I;
- the relativistic ejecta *emission* time t_e measured in Frame I;
- the comoving relativistic ejecta *emission* time t'_e measured in Frame II;
- the *observation* time $t_{\rm obs}$ measured in Frame III.

The concept of "time" is valid only when a "time interval" is concerned. In other words, one needs to compare the time span between *two* events. In order to make connections among different frames and different locations, one needs to exchange information. This is realized through emitting and receiving signals with the speed of light (the fastest speed for information transfer).

[2] Strictly speaking, both the central engine and Earth have a proper motion with respect to their respective cosmic proper frames. For example, an Earth observer observes a clear dipole moment in the cosmic microwave background due to the orbital motion of the solar system around the Galactic Center. For the purpose of simplicity, these effects are neglected.

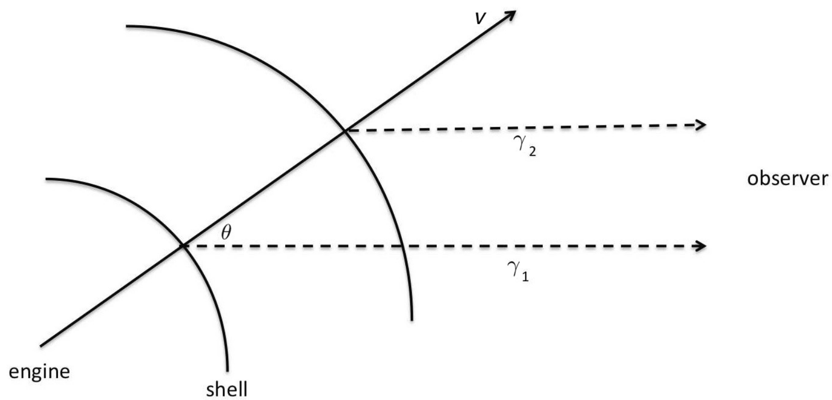

Figure 3.2 The geometry showing the central engine, the moving ejecta, and the observer.

Let us consider a simple problem. The central engine sends two light signals at $t_{\rm eng,1}$ and $t_{\rm eng,2} > t_{\rm eng,1}$ towards the relativistically moving (spherical) ejecta. The ejecta emits two light signals at $t_{e,1}$ and $t_{e,2} > t_{e,1}$ towards an observer immediately when it receives each light signal from the central engine. The two light signals emitted by the ejecta are received by the observer at $t_{\rm obs,1}$ and $t_{\rm obs,2} > t_{\rm obs,1}$. All the times are recorded by the universally tuned clocks in the lab frame. We want to find out the relationship among the *engine time interval* $\Delta t_{\rm eng} = t_{\rm eng,2} - t_{\rm eng,1}$, the *ejecta emission time interval* $\Delta t_e = t_{e,2} - t_{e,1}$, and the *observer time interval* $\Delta t_{\rm obs} = t_{\rm obs,2} - t_{\rm obs,1}$. To keep the discussion general, we consider an ejecta element that is moving with an angle θ with respect to the direction of the observer from the central engine (see Fig. 3.2). We also neglect the cosmological expansion factor $(1+z)$ in the following discussion.

The relationship between $\Delta t_{\rm eng}$ and Δt_e can be derived through a simple geometric effect: the distance Signal 1 travels plus the distance the ejecta travels during the interval between receiving the two signals is equal to the distance Signal 2 travels, i.e.

$$c(t_{e,1} - t_{\rm eng,1}) + v(t_{e,2} - t_{e,1}) = c(t_{e,2} - t_{\rm eng,2}). \tag{3.24}$$

Re-organizing this, one gets

$$c(t_{\rm eng,2} - t_{\rm eng,1}) = (c - v)(t_{e,2} - t_{e,1}), \tag{3.25}$$

or

$$\Delta t_{\rm eng} = (1 - \beta)\Delta t_e. \tag{3.26}$$

The two times are essentially the same in our daily life with $\beta \ll 1$. However, for relativistic shells with $\beta \lesssim 1$, one may approximate

$$1 - \beta = \frac{(1+\beta)(1-\beta)}{1+\beta} \simeq \frac{1-\beta^2}{2} = \frac{1}{2\gamma^2}, \tag{3.27}$$

which means that Δt_e is about $2\gamma^2$ times longer than $\Delta t_{\rm eng}$. The reason is simple: since the ejecta is moving away from the engine with a speed very close to the speed of light, it takes a much longer time for the light signal to catch up with it.

Similarly, $\Delta t_{\rm obs}$ and Δt_e can be connected through a geometric relation (Fig. 3.2):

$$c(t_{\rm obs,1} - t_{e,1}) = v\cos\theta(t_{e,2} - t_{e,1}) + c(t_{\rm obs,2} - t_{e,2}). \tag{3.28}$$

Re-organizing this, one gets

$$c(t_{\rm obs,2} - t_{\rm obs,1}) = (c - v\cos\theta)(t_{e,2} - t_{e,1}), \tag{3.29}$$

or

$$\Delta t_{\rm obs} = (1 - \beta\cos\theta)\Delta t_e. \tag{3.30}$$

Again, when $\beta \ll 1$, the two time intervals are essentially the same. However, when $\beta \lesssim 1$, the two time intervals show significant differences. When $\theta \gg 0$, one has $\Delta t_{\rm obs} \simeq (1 - \cos\theta)\Delta t_e$; while when $\theta \sim 0$, one has $\Delta t_{\rm obs} \simeq (1-\beta)\Delta t_e \simeq \Delta t_e/(2\gamma^2)$. The observed time interval is greatly reduced. This is because the ejecta is moving towards the observer with a speed close to c, so that when the second light signal is emitted, the first light signal does not lead the second one significantly.

Comparing Eqs. (3.26) and (3.30), one gets

$$\Delta t_{\rm obs} = \frac{1 - \beta\cos\theta}{1 - \beta}\Delta t_{\rm eng}. \tag{3.31}$$

When $\theta = 0$ (on axis), one has $\Delta t_{\rm obs} = \Delta t_{\rm eng}$, i.e. the time history an observer records (e.g. in a GRB lightcurve) essentially reflects the time history at the central engine. This is understandable, since the central engine and the observers are at rest with respect to one another. If the central engine directly sends two light signals to the observer, the time interval for the engine to emit the two signals should be the same as the time interval for the observer to receive them.

Equation (3.31) also suggests that the two time scales can be different if $\theta \neq 0$. For example, if the emitter is a "blob" (or narrow jet) that moves in a direction θ with respect to the line of sight, $\Delta t_{\rm obs}$ can be (sometimes much) longer than $\Delta t_{\rm eng}$.

Notice that the three time intervals we have discussed so far, i.e. $\Delta t_{\rm eng}$, Δt_e, and $\Delta t_{\rm obs}$, are all measured in the same rest frame, i.e. the inertial frame where both the central engine and the observer are at rest (Frames I and III). In other words, no special relativity effects such as Lorentz transformation or time dilation are involved. The reason that the three times are different is because they measure three pairs of *different events*: i.e. $\Delta t_{\rm eng}$ measures the time interval between emitting two light signals at the central engine; Δt_e measures the time interval for the ejecta emitting two new signals to the observer upon receiving the two signals emitted from the central engine; and $\Delta t_{\rm obs}$ measures the time interval for the observer receiving those two signals emitted from the ejecta. The relationships among the three time intervals are all a consequence of the *propagation* effect, which also applies (but insignificantly) to our daily life with $\beta \ll 1$.

It is interesting to consider an imaginary problem having a relativistic ejecta falling towards the central engine. The relationship among the three times would be different (Exercise 3.3).

Time dilation comes into effect when one considers the time interval between *the same pair of events* within *two different inertial frames*: e.g. the comoving-frame (Frame II) ejecta emission time interval is related to the lab-frame (Frame I) ejecta emission time interval through

$$\Delta t'_e = \frac{\Delta t_e}{\gamma}. \tag{3.32}$$

Combining Eqs. (3.30) and (3.32), one gets

$$\Delta t'_e = \frac{\Delta t_{\rm obs}}{\gamma(1-\beta\cos\theta)}. \tag{3.33}$$

Putting everything together, one has

$$\Delta t_{\rm eng} : \Delta t_e : \Delta t'_e : \Delta t_{\rm obs} = \frac{1-\beta}{1-\beta\cos\theta} : \frac{1}{1-\beta\cos\theta} : \frac{1}{\gamma(1-\beta\cos\theta)} : 1. \tag{3.34}$$

When $\gamma \gg 1$ and $\theta \simeq 0$, this becomes

$$\Delta t_{\rm eng} : \Delta t_e : \Delta t'_e : \Delta t_{\rm obs} \simeq 1 : 2\gamma^2 : 2\gamma : 1. \tag{3.35}$$

Notice that in the later chapters of the book, the observer time $t_{\rm obs}$ is often denoted as t since it is the time directly related to observations. The lab-frame time t_e, on the other hand, is denoted as $\hat{t}$. The comoving time t'_e is denoted as t' for simplicity.

3.2.3 "Superluminal" Motion

One interesting phenomenon for a relativistic jet moving in a direction close to an observer on Earth (e.g. for GRBs and some AGNs named blazars) is that the apparent transverse motion speed can exceed the speed of light. This is an artifact that does not violate special relativity, and can be derived as follows. By definition, the apparent transverse speed can be written as

$$v_\perp^{\rm app} = \frac{\Delta d_\perp}{\Delta t_{\rm obs}} = \frac{v \cdot \Delta t_e \cdot \sin\theta}{\Delta t_{\rm obs}} = v \cdot \frac{\sin\theta}{1-\beta\cos\theta}, \tag{3.36}$$

or

$$\beta_\perp^{\rm app} = \beta \cdot \frac{\sin\theta}{1-\beta\cos\theta}. \tag{3.37}$$

Setting $d\beta_\perp^{\rm app}/d\theta = 0$, one can derive that the maximum of $\beta_\perp^{\rm app}$ is reached at $\cos\theta = \beta$ (i.e. $\sin\theta = 1/\gamma$):

$$\beta_{\perp,\rm max}^{\rm app} = \beta\gamma = \sqrt{\gamma^2 - 1}, \tag{3.38}$$

which is $\gg 1$ if $\gamma \gg 1$, suggesting that *apparent "superluminal" motion* is possible for relativistic motion.

In order to observe superluminal motion, two conditions should be satisfied. The first one is the general condition $\beta_{\perp,\rm max}^{\rm app} > 1$. This gives

$$\gamma > \sqrt{2}, \tag{3.39}$$

or $1/\sqrt{2} < \beta < 1$. The second condition is that the viewing angle has to be in a range close to the direction of motion. Demanding Eq. (3.37) be greater than unity, i.e.

$$\beta \frac{\sin\theta}{1-\beta\cos\theta} > 1, \tag{3.40}$$

one can solve the inequality

$$\beta^2(1-\cos^2\theta) > 1 - 2\beta\cos\theta + \beta^2\cos^2\theta, \tag{3.41}$$

which gives

$$\frac{1-\sqrt{2\beta^2-1}}{2\beta} < \cos\theta < \frac{1+\sqrt{2\beta^2-1}}{2\beta}. \tag{3.42}$$

Notice that when θ is very close to 0, one does not expect superluminal motion, since $\beta_\perp^{\rm app}$ approaches 0 when θ approaches 0.

3.3 Doppler Transformations

3.3.1 Doppler Factor

Equation (3.33) makes the connection between the *comoving-frame emission time interval* $\Delta t'_e$ and the *observer-frame observation time interval* $\Delta t_{\rm obs}$ (i.e. the observer time interval defined in §3.2.2). The coefficient is called the *Doppler factor*:

$$\mathcal{D} \equiv \frac{1}{\gamma(1-\beta\cos\theta)}. \tag{3.43}$$

It includes two factors: the γ factor is from the Lorentz transformation of the same pair of events (signal emission) in the two frames; and the $(1-\beta\cos\theta)$ factor is from the propagation effect in the same rest frame. The Doppler factor $\mathcal{D}$ is very important. Various comoving-frame quantities are connected to the measured quantities in the observer frame through a certain power of $\mathcal{D}$ (see more below).

From Eq. (3.22),

$$\cos\theta = \frac{\cos\theta'+\beta}{1+\beta\cos\theta'},$$

the Doppler factor can therefore also be written as

$$\mathcal{D} = \frac{1}{\gamma\left(1-\beta\frac{\cos\theta'+\beta}{1+\beta\cos\theta'}\right)} = \frac{1+\beta\cos\theta'}{\gamma(1-\beta^2)} = \gamma(1+\beta\cos\theta'). \tag{3.44}$$

It is useful to summarize some special values of the $\mathcal{D}$ factor:

- $\theta = 0$ and $\theta' = 0$: $\mathcal{D} = (1+\beta)\gamma \simeq 2\gamma$ (last approximation applies when $\beta \lesssim 1$);
- $\theta = \cos^{-1}\beta$ (or $\theta = \sin^{-1}\gamma^{-1}$) and $\theta' = \pi/2$: $\mathcal{D} = \gamma$;
- $\theta = \theta_c \equiv \cos^{-1}\left(\frac{1}{\beta} - \frac{1}{\beta\gamma}\right)$ and $\theta' = \theta'_c \equiv \cos^{-1}\left(\frac{1}{\beta\gamma} - \frac{1}{\beta}\right)$: $\mathcal{D} = 1$;

- $\theta = \pi/2$ and $\theta' = \cos^{-1}(-\beta)$: $\mathcal{D} = \frac{1}{\gamma}$;
- $\theta = \pi$ and $\theta' = \pi$: $\mathcal{D} = \frac{1}{(1+\beta)\gamma} \simeq \frac{1}{2\gamma}$ (last approximation applies when $\beta \lesssim 1$).

In the following, we derive the relations between some comoving-frame quantities and observer-frame quantities through *Doppler transformations*. For simplification, we omit the subscipts "*e*" and "obs". The observer- and comoving-frame quantities are denoted with symbols without and with a prime sign, respectively.

3.3.2 Time, Frequency, and Energy

The time and frequency in the two frames are related to each other through

$$dt = \mathcal{D}^{-1} dt', \tag{3.45}$$

$$\nu = \mathcal{D}\nu'. \tag{3.46}$$

The time relation (3.45) was already derived in Eq. (3.33). The frequency relation (3.46) follows immediately by noticing $\nu \propto (dt)^{-1}$.

It is straightforward to express energy in terms of photon energy $E = h\nu$. As a result, energy should have the same Doppler transformation as frequency, i.e.

$$E = \mathcal{D}E'. \tag{3.47}$$

3.3.3 Length and Volume

The transformations of length and volume are

$$ds = \mathcal{D}ds', \tag{3.48}$$

$$dV = \mathcal{D}dV'. \tag{3.49}$$

Equation (3.48) needs an explanation. As mentioned earlier, the *observed length* of a rod is measured at the same observing time by a particular observer. Let us consider a rod segment with a comoving length ds' moving with a Lorentz factor γ along the direction of the rod length, which has an angle θ with respect to the line of sight (Fig. 3.3). Due to the length contraction effect, the *intrinsic length* of the rod in the lab frame is $d\hat{s} = ds'/\gamma$. An observer along the line of sight does not measure this length. At a particular epoch, she simultaneously records one light signal emitted from the "front end" of the rod and another light signal emitted from the "back end" of the rod (at an earlier emission time), and she measures the length of the rod segment as the distance between the "front end" and the "back end" in the rest frame of the observer (Frame III), i.e. $\overline{B_1A_2}$. Apparently, two light signals emitted simultaneously from the front and back ends will not arrive at the observer at the same time. In order to have a signal emitted from the back end (B_1) at an earlier epoch arrive at the same time to the observer as a signal emitted from the front end (A_2) at a later epoch, one uses the geometric relation $\overline{B_1A_2} \cdot \cos\theta = \overline{B_1C}$ in Fig. 3.3. Noting that $\overline{B_1A_1} = \overline{B_2A_2} = d\hat{s} = ds'/\gamma$, $\overline{B_1C} = cd\hat{t}$, and $\overline{A_2A_1} = \overline{B_2B_1} = vd\hat{t}$, one can derive the length $ds = \overline{B_1A_2}$ through the relation

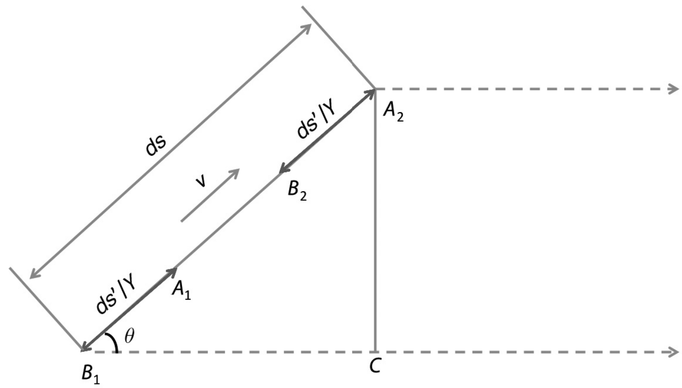

Figure 3.3 The geometry showing the Doppler transformation of the length of a rod moving with an angle θ with respect to the line of sight.

$$\frac{ds - ds'/\gamma}{ds\cos\theta} = \frac{v}{c} = \beta. \tag{3.50}$$

Solving this equation, one gets $ds(1 - \beta\cos\theta) = ds'/\gamma$, and hence, Eq. (3.48).

The unit volume as observed by the observer is $dV = A\,ds$, whereas that in the comoving frame is $dV' = A'\,ds'$, where A and $A' = A$ are the areas perpendicular to the viewing direction of the observer in the two rest frames, respectively. Equation (3.49) is then derived straightforwardly.

3.3.4 Solid Angle

From Eq. (3.22), one can derive

$$\begin{aligned} d\cos\theta &= \frac{(d\cos\theta')(1+\beta\cos\theta') - (\cos\theta' + \beta)\beta(d\cos\theta')}{(1+\beta\cos\theta')^2} \\ &= \frac{(1-\beta^2)(d\cos\theta')}{(1+\beta\cos\theta')^2} = \frac{d\cos\theta'}{\gamma^2(1+\beta\cos\theta')^2} \\ &= \mathcal{D}^{-2}(d\cos\theta'). \end{aligned} \tag{3.51}$$

The unit solid angle can be written as $d\Omega = \sin\theta d\theta d\phi = -(d\cos\theta)d\phi$. Since there is no change in $d\phi$ in the two frames, one can finally derive

$$d\Omega = \mathcal{D}^{-2} d\Omega'. \tag{3.52}$$

3.3.5 Specific Intensity, Emission Coefficient, and Absorption Coefficient

In problems of radiative transfer, the following concepts are widely used (Rybicki and Lightman, 1979):

- Specific intensity I_ν is defined by

$$dE = I_\nu dA dt d\Omega d\nu, \tag{3.53}$$

which carries the meaning of the radiation energy dE crossing an area dA normal to the direction of a given light ray, into unit solid angle $d\Omega$, in unit time dt and unit frequency $d\nu$.

- Specific emission coefficient j_ν is defined by

$$dE = j_\nu dV d\Omega dt d\nu, \tag{3.54}$$

which carries the meaning of the radiation energy dE into unit solid angle $d\Omega$, emitted from unit volume dV, in unit time dt and unit frequency $d\nu$.

- Specific absorption coefficient α_ν is defined by

$$dI_\nu = -\alpha_\nu I_\nu ds, \tag{3.55}$$

which represents the normalized loss of specific intensity in a light beam per unit length.

The three parameters are related in the radiative transfer equation

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \tag{3.56}$$

or

$$\frac{dI_\nu}{d\tau_\nu} = -I_\nu + S_\nu, \tag{3.57}$$

where

$$S_\nu \equiv \frac{j_\nu}{\alpha_\nu} \tag{3.58}$$

is the source function, and τ_ν is the specific optical depth defined by

$$d\tau_\nu = \alpha_\nu ds. \tag{3.59}$$

Based on the transformation relations in §3.3.2–3.3.4, it is straightforward to prove the following Doppler transformations (Exercise 3.4):

$$I_\nu(\nu) = \mathcal{D}^3 I'_{\nu'}(\nu'), \tag{3.60}$$

$$j_\nu(\nu) = \mathcal{D}^2 j'_{\nu'}(\nu'), \tag{3.61}$$

$$\alpha_\nu(\nu) = \mathcal{D}^{-1} \alpha'_{\nu'}(\nu'). \tag{3.62}$$

3.4 Specific Luminosity and Flux

In astrophysical problems, what is directly observed is the specific flux (F_ν) or flux ($F = \int_{\nu_1}^{\nu_2} F_\nu d\nu$) in a frequency range ($\nu_1, \nu_2$). By knowing the distance, one can infer the specific luminosity (L_ν) or luminosity ($L = \int_{\nu_1}^{\nu_2} L_\nu d\nu$).

3.4.1 Steady, Comoving-Frame Isotropic and Homogeneous, Point Emitting Source

Let us consider the simplest case that invokes a *steady* (no time evolution of emissivity), comoving-frame *isotropic* and homogeneous, optically thin emitting source with a comoving emission coefficient $j'_{\nu'}$. The comoving-frame specific isotropic luminosity at ν' can be written as

$$L'_{\nu'}(\nu') = \int\int j'_{\nu'}(\nu')d\Omega' dV' = \int 4\pi j'_{\nu'}(\nu')dV' = 4\pi j'_{\nu'}(\nu')V'. \tag{3.63}$$

In the observer frame, it is straightforward to write

$$\frac{dL_\nu(\nu)}{d\Omega} = \int j_\nu(\nu)dV = \int \mathcal{D}^2 j'_{\nu'}(\nu')\mathcal{D}dV' = \mathcal{D}^3 \frac{dL'_{\nu'}(\nu')}{d\Omega'} = \mathcal{D}^3 \frac{L'_{\nu'}(\nu')}{4\pi}, \tag{3.64}$$

where the comoving isotropic condition has been applied for the last equality.

The transformation of the specific luminosity L_ν depends on the properties of the source. For an extended source where different spatial elements move in different directions (e.g. the case of a conical jet), one should integrate Eq. (3.64) by considering the transformation of the solid angle ($d\Omega = \mathcal{D}^{-2}d\Omega'$), so that at any angle θ between the direction of motion and line of sight in the observer frame, the *specific luminosity of a unit emitting element* at a particular frequency ν reads

$$L_\nu(\nu) = \mathcal{D}L'_{\nu'}(\nu'). \tag{3.65}$$

This can also be derived by $L_\nu(\nu) = \frac{dE}{dt\, d\nu} = \frac{\mathcal{D}dE'}{\mathcal{D}^{-1}dt'\, \mathcal{D}d\nu'} = \mathcal{D}\frac{dE'}{dt'\, d\nu'} = \mathcal{D}L'_{\nu'}(\nu')$. The *luminosity of a unit emitting element* at a particular frequency ν reads

$$L(\nu) = \nu L_\nu(\nu) = \mathcal{D}^2(\nu' L_{\nu'}(\nu')) = \mathcal{D}^2 L'(\nu'). \tag{3.66}$$

The total specific luminosity and luminosity of an extended source need to be calculated by integrating over the entire equal-arrival-time surface, which we will introduce in §3.4.2 and §3.4.3 below.

Next we consider a *point source*, for which all the emitter material is moving towards one direction (no θ dependence in terms of motion). An observer mostly cares about the *isotropic-equivalent specific luminosity*, i.e. the specific luminosity *assuming* that the source is isotropic in the observer's frame (which is not the case for a relativistic moving object). For a point source, the isotropic-equivalent specific luminosity is simply Eq. (3.64) multiplied by $\int d\Omega = 4\pi$, so that

$$L_{\nu,\mathrm{iso}}(\nu) = \mathcal{D}^3 L'_{\nu'}(\nu'). \tag{3.67}$$

The isotropic-equivalent luminosity at a particular frequency ν is

$$L_{\mathrm{iso}}(\nu) = \nu L_{\nu,\mathrm{iso}}(\nu) = \mathcal{D}^4(\nu' L'_{\nu'}(\nu')). \tag{3.68}$$

The luminosity distance D_{L} of a cosmological source is defined through $F(\nu_{\mathrm{obs}}) = L_{\mathrm{iso}}(\nu)/4\pi D_{\mathrm{L}}^2$. As a result, the observed specific flux at Earth can be written as

$$F_\nu(\nu_{\mathrm{obs}}) = \frac{(1+z)L_{\nu,\mathrm{iso}}(\nu)}{4\pi D_{\mathrm{L}}^2}, \tag{3.69}$$

where $\nu_{\rm obs} = \nu/(1+z) = \mathcal{D}\nu'/(1+z)$. For the point source discussed above, one has

$$F_\nu(\nu_{\rm obs}) = \frac{(1+z)\mathcal{D}^3 j'_{\nu'}(\nu')V'}{D_{\rm L}^2}. \tag{3.70}$$

3.4.2 Equal-Arrival-Time Surface

Next we consider a relativistic, extended source in detail. The observed flux should be calculated by integrating the surface brightness across the emission area. One important effect is that, for a relativistic object, photons emitted at different source-frame times at different locations arrive at the observer at the same time. In order to calculate the flux at a particular observer time, one must properly take into account the *equal-arrival-time surface*, or EATS.

We first recall the relationship between the emission time interval dt_e and the observed time interval $dt_{\rm obs}$:

$$dt_{\rm obs} = dt_e(1-\beta\mu), \tag{3.71}$$

where $\mu = \cos\theta$.

Let us consider a simple problem: an infinitely thin, spherical (or conical) shell moves from an origin ($r = 0$) with a constant Lorentz factor Γ (no acceleration[3] or deceleration). Define $t_e = 0$ and $t_{\rm obs} = 0$ when the source is at $r = 0$. One therefore has $t_{\rm obs} = t_e(1-\beta\mu)$. Noting $t_e = r/\beta c$, one can write r as a function of θ (or μ) given a certain constant value of $t_{\rm obs}$, i.e.

$$r = \frac{\beta c t_{\rm obs}}{1-\beta\mu}. \tag{3.72}$$

This is the equation of the *equal-arrival-time surface (EATS)*. An example of EATS for constant Γ motion ($\Gamma = 300$) is displayed as the middle curve (1a) in Fig. 3.4. One can derive the following properties of an EATS (Exercise 3.5):

- At $\theta = 0$ ($\mu = 1$), one has

$$r_{\rm max} = \frac{\beta c t_{\rm obs}}{1-\beta} \simeq 2\Gamma^2 \beta c t_{\rm obs}. \tag{3.73}$$

  The last approximation applies when $\Gamma \gg 1$.
- At $\theta = \cos^{-1}\beta$ ($\mu = \beta$), one has $r = \Gamma^2\beta c t_{\rm obs}$. The perpendicular component $r_\perp = r\sqrt{1-\mu^2}$ reaches the maximum, i.e.

$$r_\perp = r_{\perp,\rm max} \equiv \Gamma\beta c t_{\rm obs}. \tag{3.74}$$

- At $\theta = \pi/2$ ($\mu = 0$), one has $r = \beta c t_{\rm obs}$.
- At $\theta = \pi$ ($\mu = -1$), one has

$$r_{\rm min} = \frac{\beta c t_{\rm obs}}{1+\beta} \simeq \frac{1}{2}\beta c t_{\rm obs}. \tag{3.75}$$

  The last approximation applies when $\beta \lesssim 1$.

[3] In reality, the source needs to undergo an initial acceleration phase in order to reach a terminal Γ. The following discussion is valid if the acceleration time is much shorter than the emission time t_e.

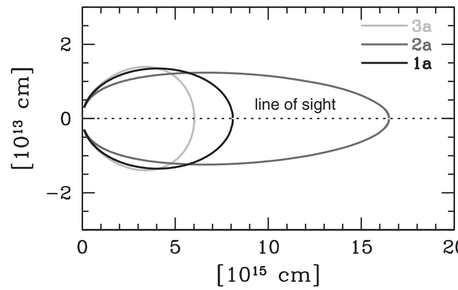

Figure 3.4 The equal-arrival-time surface for a relativistic spherical shell moving with a constant Γ (middle), under acceleration (elongated), and under deceleration (shortened). From Uhm and Zhang (2015).

- The EATS is an ellipsoid with the source at the far focal point. The semi-major axis is

$$a = \frac{1}{2}(r_{\max} + r_{\min}) = \Gamma^2 \beta c t_{\rm obs}, \tag{3.76}$$

and the semi-minor axis is

$$b = r_{\perp,\max} = \Gamma \beta c t_{\rm obs}. \tag{3.77}$$

If the shell is undergoing acceleration or deceleration, the shape of the EATS will be distorted. For an accelerating shell, one needs to release photons earlier at higher latitudes in order to have photons catch up the accelerating shell and reach the observer at the same time. As a result, the EATS is stretched. The trend is opposite for a decelerating shell. In both the acceleration and deceleration cases, the EATS deviates from the simple ellipsoid shape. The elongated (2a) and the shortened (3a) curves in Fig. 3.4 show the examples of acceleration and deceleration. The Lorentz factor is assumed to evolve with radius as $\Gamma(r) = \Gamma_0(r/r_0)^s$ with $r_0 = 10^{14}$ cm. Model (2a) takes $\Gamma_0 = 10^2$ and $s = 0.4$, whereas Model (3a) takes $\Gamma_0 = 10^3$, $s = -0.4$ (Uhm and Zhang, 2015).

3.4.3 Steady, Comoving-Frame Isotropic and Homogeneous, Extended Source

Next we consider an extended source, which is steady and comoving-frame isotropic and homogeneous, and moves with a constant Lorentz factor Γ. In order to calculate the specific luminosity $L_\nu(\nu)$, one needs to integrate Eq. (3.64) over the EATS for a given $t_{\rm obs}$.

According to the EATS equation (3.72) and considering $c\beta t_{\rm obs} =$ constant, one has

$$d\mu = \frac{1-\beta\mu}{\beta r} dr. \tag{3.78}$$

Noting $d\Omega = \sin\theta d\theta d\phi = -d\mu d\phi$, one has

$$\begin{aligned} L_\nu(\nu) &= \int \mathcal{D}^3 \frac{L'_{\nu'}(\nu')}{4\pi} d\Omega = \frac{1}{2}\int_{-1}^{1} \mathcal{D}^3 L'_{\nu'}(\nu') d\mu \\ &= \frac{1}{2}\int_{r_{\min}}^{r_{\max}} \mathcal{D}^3 L'_{\nu'}(\nu') \frac{1-\beta\mu}{\beta r} dr \\ &= \frac{1}{2}\int_{r_{\min}}^{r_{\max}} \frac{L'_{\nu'}(\nu')}{\Gamma^3(1-\beta\mu)^2} \frac{dr}{\beta r}, \end{aligned} \tag{3.79}$$

where $r_{\rm min}$ and $r_{\rm max}$ (Eqs. (3.75) and (3.73)) are the minimum and maximum distances from the engine on the EATS. Plugging in $1 - \beta\mu = c\beta t_{\rm obs}/r$ and performing the integration, one finally gets

$$L_\nu(\nu) = \frac{1}{4}\frac{L'_{\nu'}(\nu')}{\Gamma^3\beta^3(ct_{\rm obs})^2}\left(r_{\rm max}^2 - r_{\rm min}^2\right) = \Gamma L'_{\nu'}(\nu'). \tag{3.80}$$

Compared with Eq. (3.65), this is essentially the specific luminosity of the element at an angle $\theta = 1/\Gamma$ with respect to the line of sight. So, for an extended source with the line of sight piercing through the shell, the dominant contribution of emissivity is from the elements within the $1/\Gamma$ cone.

3.4.4 High-Latitude Emission, Curvature Effect

In GRB problems, often one considers a situation when an emitting shell stops shining abruptly. The observed flux does not stop immediately. Rather, photons from higher latitudes with respect to the line of sight arrive at the observer at progressively later epochs, defining a decaying lightcurve.

The propagation geometry is still defined by Eq. (3.72). However, in this problem, rather than fixing $t_{\rm obs}$ as done when defining the EATS, one has a constant r and allows $t_{\rm obs}$ to vary with μ, i.e.

$$t_{\rm obs} = \frac{r}{\beta c}(1 - \beta\mu), \tag{3.81}$$

and

$$-\,d\mu = \frac{c}{r}dt_{\rm obs}. \tag{3.82}$$

Noticing $d\Omega = -d\mu d\phi$, Eq. (3.64) can be written as $dL_\nu(\nu)/(-d\mu) = (1/2)\mathcal{D}^3 L'_{\nu'}(\nu')$. With Eqs. (3.81) and (3.82), one can derive

$$\begin{aligned}\frac{dL_\nu(\nu)}{dt_{\rm obs}} &= \frac{1}{2}\mathcal{D}^3 L'_{\nu'}(\nu')\frac{c}{r} \\ &= \frac{1}{2}\left(\frac{r}{c}\right)^2\frac{1}{(\Gamma\beta)^3}\frac{L'_{\nu'}(\nu')}{t_{\rm obs}^3},\end{aligned} \tag{3.83}$$

so that

$$L_\nu(\nu, t_{\rm obs}) \simeq \frac{dL_\nu(\nu)}{dt_{\rm obs}}\cdot t_{\rm obs} = \frac{1}{2}\left(\frac{r}{c}\right)^2\frac{1}{(\Gamma\beta)^3}\frac{L'_{\nu'}(\nu')}{t_{\rm obs}^2}. \tag{3.84}$$

Let us consider a comoving-frame power-law spectrum, with

$$L'_{\nu'}(\nu') = A'\nu'^{-\hat\beta} = A'\nu^{-\hat\beta}\mathcal{D}^{\hat\beta}, \tag{3.85}$$

Plugging it in Eq. (3.84), and noticing Eq. (3.81), one finally gets

$$L_\nu(\nu) = \frac{1}{2}A'\left(\frac{r}{c}\right)^{2+\hat\beta}(\Gamma\beta)^{-(3+\hat\beta)}\,\nu^{-\hat\beta}t_{\rm obs}^{-(2+\hat\beta)}. \tag{3.86}$$

Noting $F_\nu \propto L_\nu$, this gives the famous high-latitude emission *curvature effect* relation,

$$\hat{\alpha} = 2 + \hat{\beta}, \tag{3.87}$$

between the temporal decay index $\hat{\alpha}$ and spectral index $\hat{\beta}$, in the convention $F_\nu \propto \nu^{-\hat{\beta}} t_{\rm obs}^{-\hat{\alpha}}$. This relation was first correctly derived by Kumar and Panaitescu (2000a), and reproduced in several later works, both analytically and numerically (e.g. Dermer, 2004; Dyks et al., 2005; Uhm and Zhang, 2015).

There are two assumptions in deriving the relation (3.87): that the spectrum is a power law with index $\hat{\beta}$, and that the emission region moves with a constant Lorentz factor Γ before the emission ceases. Relaxing these two assumptions gives a more general description of the curvature effect.

First, we still consider a constant Γ motion, but with a spectrum that is curved in a non-power-law form, which may be cast in the form

$$L'_{\nu'}(\nu') = L'_{\nu',0}(\nu'_0) F\left(\frac{\nu'}{\nu'_0}\right), \tag{3.88}$$

where $F(\nu'/\nu'_0)$ is an arbitrary function, and ν'_0 is a characteristic frequency. The observed spectrum during the curvature-effect-dominated phase can be characterized by the same function shape, i.e.

$$L_\nu(\nu) = L_{\nu,0}(\nu_0) F\left(\frac{\nu}{\nu_0}\right), \tag{3.89}$$

with (e.g. Zhang et al., 2009b)

$$L_{\nu,0}(\nu_0) \propto L'_{\nu',0}(\nu'_0) t_{\rm obs}^{-2} \propto \mathcal{D}^2 L'_{\nu',0}(\nu'_0) \tag{3.90}$$

(from Eq. (3.84)), and

$$\nu_0 \propto \nu'_0 t_{\rm obs}^{-1} \propto \mathcal{D}\nu'_0. \tag{3.91}$$

For a consistency check, taking $F(\nu'/\nu'_0) \propto \nu'^{-\hat{\beta}} \nu_0'^{\hat{\beta}}$, one has $L_\nu(\nu) \propto L_{\nu_0}(\nu_0)\nu^{-\hat{\beta}}\nu_0^{\hat{\beta}} \propto t_{\rm obs}^{-2}\nu^{-\hat{\beta}} t_{\rm obs}^{-\hat{\beta}} \propto \nu^{\hat{\beta}} t_{\rm obs}^{-(2+\hat{\beta})}$.

For a narrow band (e.g. *Swift* XRT), an intrinsically curved spectrum may be approximated as a power law. When a curved spectrum moves across the band during the curvature effect decay phase, the effective $\hat{\beta}$ would evolve as a function of time. The simple relation (3.87) would still apply approximately, with $\hat{\alpha}(t) \simeq 2 + \hat{\beta}(t)$.

Next, we relax the constant Γ assumption, and consider possible acceleration or deceleration before the emitter stops shining. Since the shape of the EATS depends on the history of the dynamical evolution of the emitter, the decay slope due to the curvature effect would deviate from the simple relation (3.87). In particular, for an accelerating shell, since the EATS is more elongated, at the same latitude the emission comes from an ealier epoch when the emitter had a smaller Γ, so that the emissivity is weaker. The curvature effect decay tail therefore displays a steeper decay than (3.87). Conversely, for a decelerating shell, the higher-latitude emission is enhanced due to a larger radius and higher Γ with respect to the constant Γ case. The decay slope is therefore shallower than (3.87). Figure 3.5 presents the curvature effect lightcurves that correspond to the three EATS calculated in Fig. 3.4 (Uhm and Zhang, 2015).

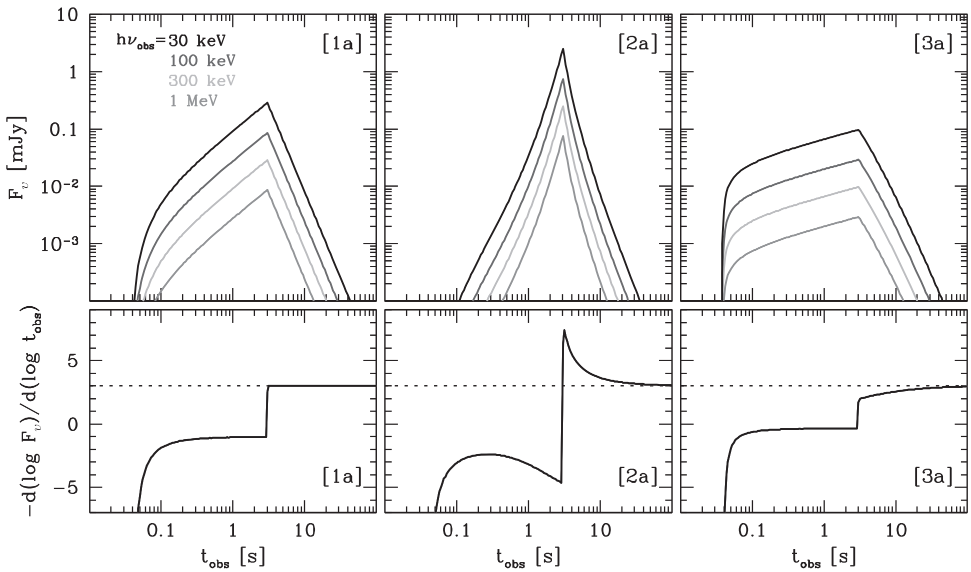

Figure 3.5 The curvature-effect-dominated lightcurves for the cases of constant Γ, acceleration, and deceleration, respectively (see Figure 3.4). From top to bottom, the lightcurves correspond to 30 keV, 100 keV, 300 keV, and 1 MeV, respectively. The asymptotic value (dashed lines in the lower panels) satisfies $\hat{\alpha} = 2 + \hat{\beta}$. From Uhm and Zhang (2015).

Another effect to produce a high-latitude-emission decay slope steeper than Eq. (3.87) is to introduce anisotropy of emission in the comoving frame (e.g. Beloborodov, 2011; Barniol Duran et al., 2016; Geng et al., 2017). The reason is straightforward: when the emission from higher latitudes arrives at the observer, a deficit of emissivity in certain directions would reduce the observed flux with respect to the baseline value, leading to a steeper decay slope.

For the above discussion, the time zero point is defined at the beginning of the shell evolution ($t_{\rm obs} = 0$ at $r = 0$ and $t = 0$). In order to perform a test of the curvature effect and, hence, to diagnose whether an emitter is undergoing bulk acceleration or deceleration, a correct time zero point should be properly selected. In the GRB problems, X-ray flares are usually ideal candidates for diagnosing the curvature effect, thanks to their long decaying tails detected with *Swift*/XRT (Liang et al., 2006b). In practice, correcting the t_0 effect turns out to be non-trivial. An interesting finding from these analyses is that essentially all the X-ray flares have a decay slope steeper than Eq. (3.87) after the t_0 effect is properly taken care of (Uhm and Zhang, 2016a; Jia et al., 2016). The results are consistent with having an emission region that undergoes bulk acceleration (Uhm and Zhang, 2016a). An anisotropic emission region may contribute to the steepening, but bulk acceleration is needed to reach the best fit to the data of both lightcurve and spectral evolution (Geng et al., 2017). Both bulk acceleration and anisotropic emission are consistent with the scenario that the jet is Poynting flux dominated in the emission region, since a fraction of the dissipated Poynting flux energy is converted to the kinetic energy of the flow, and magnetic reconnection tends to generate mini-jets so that the emission is not isotropic in the comoving frame (§9.8 for detailed discussion).

3.5 Some Useful Formulae

There are some commonly used formulae in GRB problems invoking relativistic motion.

3.5.1 Relative Lorentz Factor

One problem is to calculate the relative Lorentz factor between two objects that move in the same direction with different Lorentz factors γ_2 and γ_1.

In this problem, one can apply Eq. (3.14). Letting γ_1 correspond to u, γ_2 correspond to v, and γ_{12} correspond to u', one gets

$$\beta_{12} = \frac{\beta_1 - \beta_2}{1 - \beta_1\beta_2}. \tag{3.92}$$

Since

$$\beta = \left(1 - \frac{1}{\gamma^2}\right)^{1/2} \simeq 1 - \frac{1}{2\gamma^2} \tag{3.93}$$

when $\gamma \gg 1$, one has

$$\beta_{12} \simeq \frac{1 - \frac{1}{2\gamma_1^2} - 1 + \frac{1}{2\gamma_2^2}}{1 - \left(1 - \frac{1}{2\gamma_1^2}\right)\left(1 - \frac{1}{2\gamma_2^2}\right)} \simeq \frac{\gamma_1^2 - \gamma_2^2}{\gamma_1^2 + \gamma_2^2}, \tag{3.94}$$

so that the *relative Lorentz factor* is

$$\gamma_{12} = \frac{1}{(1 - \beta_{12}^2)^{1/2}} \simeq \frac{\gamma_1^2 + \gamma_2^2}{2\gamma_1\gamma_2} = \frac{1}{2}\left(\frac{\gamma_1}{\gamma_2} + \frac{\gamma_2}{\gamma_1}\right). \tag{3.95}$$

3.5.2 Catch-Up Problem

Another problem is related to GRB internal shocks. Suppose that two objects move in the same direction. Object I moves with γ_1 and leaves the engine first. After a time interval Δt, Object II leaves the engine with a larger Lorentz factor, $\gamma_2 > \gamma_1$. The question is when and where Object II catches up with Object I.

Let the collision happen at $t_{\rm col}$ in the lab frame at the radius $R_{\rm col}$ from the origin. One then has

$$(\beta_2 - \beta_1)t_{\rm col} = \beta_1\Delta t, \tag{3.96}$$

so that

$$t_{\rm col} = \frac{\beta_1\Delta t}{\beta_2 - \beta_1}. \tag{3.97}$$

The collision radius is

$$R_{\rm col} = v_2 t_{\rm col} = \frac{\beta_1\beta_2 c\Delta t}{\beta_2 - \beta_1} = \frac{c\Delta t}{\frac{1}{\beta_1} - \frac{1}{\beta_2}}. \tag{3.98}$$

For $\gamma_2 \gg 1$ and $\gamma_1 \gg 1$, one has

$$R_{\rm col} \simeq \frac{c\Delta t}{\frac{1}{2\gamma_1^2} - \frac{1}{2\gamma_2^2}}. \tag{3.99}$$

Let $\gamma_2 = \xi\gamma_1$, with $\xi > 1$, one has

$$R_{\rm col} \simeq \frac{2\xi^2}{\xi^2 - 1}\gamma_1^2 c\Delta t \gtrsim 2\gamma_1^2 c\Delta t. \tag{3.100}$$

This last approximate formula is commonly used to estimate the internal shock radius.

3.6 General Relativity Overview

3.6.1 Postulates and Einstein Field Equations

Postulates

General relativity is based on the following two postulates (Einstein, 1916):

- *General principle of relativity*: All systems of reference, inertial or non-inertial, are equivalent with respect to the formulation of the fundamental laws of physics;
- *Equivalence principle*: The gravitational mass and inertial mass of an object are always equal to each other, so that gravity accelerates all objects equally regardless of their masses.

Einstein Field Equations

The first principle suggests that there exists a set of universal, simple equations, albeit mathematically complicated, to describe motion of objects regardless of whether they are in uniform motion (inertial frame) or in accelerated motion (non-inertial frame). These equations are known as the *Einstein field equations*, which can be cast in a simple tensor form (altogether $4 \times 4 = 16$ equations, 6 of which are independent[4]):

$$G^{\mu\nu} + g^{\mu\nu}\Lambda = \frac{8\pi G}{c^4}T^{\mu\nu}. \tag{3.101}$$

The second principle is the key behind the Einstein field equations. Since the acceleration is the same for all masses in a gravitational field, this so-called "acceleration" can be replaced by a simpler concept, namely, the space-time itself is curved in such a way that objects with different test masses follow the same geodesic trajectories. The theory of gravity then turns into a theory of how the distribution of mass (and, more generally, energy) affects the geometry of space-time and vice versa.

[4] Since the metric tensor is symmetric, there are only 10 independent elements. The so-called *Bianchi identity* states $G^{\mu\nu}_{;\nu} = 0$, which reduces 4 more independent equations.

Energy–Momentum Tensor

The tensor

$$T^{\mu\nu} = \begin{pmatrix} T^{00} & T^{01} & T^{02} & T^{03} \\ T^{10} & T^{11} & T^{12} & T^{13} \\ T^{20} & T^{21} & T^{22} & T^{23} \\ T^{30} & T^{31} & T^{32} & T^{33} \end{pmatrix} \tag{3.102}$$

on the right hand side of Eq. (3.101) is a 4×4 (0 for time and (1,2,3) for space) tensor called the *energy–momentum tensor*. The physical meaning of each element is the following: $T^{00} = \rho c^2$ is the energy density, where ρ is the mass density; T^{12} is the x-component of momentum flux across a unit surface area of constant y, etc; T^{01} is the energy flux divided by c across a unit surface area of constant x, etc. It is a symmetric tensor, i.e. $T^{\mu\nu} = T^{\nu\mu}$.

A cold fluid with density ρ_0 in its rest frame has only one non-zero element:

$$T^{\mu\nu} = \begin{pmatrix} \rho_0 c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.103}$$

In a rest frame where a fluid moves in the x-direction, the new tensor has the form (through Lorentz transformation)

$$T^{\mu\nu} = \rho_0 c^2 \begin{pmatrix} \gamma^2 & -\gamma^2\beta & 0 & 0 \\ -\gamma^2\beta & \gamma^2\beta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.104}$$

For a perfect fluid (not necessarily cold) in its rest frame, the tensor is diagonal:

$$T^{\mu\nu} = \begin{pmatrix} \rho_0 c^2 + e & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}, \tag{3.105}$$

where e is the *internal energy* (energy of the random motion in the comoving frame) *density*, and p is the pressure. The pressure p is simply the flux density of the x-momentum across the unit area in the surface of constant x, etc., so one has $T^{11} = T^{22} = T^{33} = p$.

A general expression of the energy–momentum tensor of a perfect fluid, for which we do not give a proof, reads

$$\begin{aligned} T^{\mu\nu} &= (\rho + e/c^2 + p/c^2) U^\mu U^\nu + p g^{\mu\nu} \\ &= (\rho c^2 + e + p) u^\mu u^\nu + p g^{\mu\nu}, \end{aligned} \tag{3.106}$$

where $U^\mu = (U^0, U^1, U^2, U^3) = \gamma(c, v_x, v_y, v_z)$ is the 4-velocity, $u^\mu = U^\mu / c = \gamma(1, \beta_x, \beta_y, \beta_z)$ is the normalized 4-velocity, and $g^{\mu\nu}$ is the metric tensor, which is explained below.

Metric Tensor

The left hand side of the Einstein field equations is a function of a 4×4 tensor $g^{\mu\nu}$ that delineates how space-time is curved. The tensor $G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R$ is a complicated function of $g^{\mu\nu}$, where $R^{\mu\nu}$ is called the *Ricci tensor*, which is the "contraction"[5] of a fourth-order *Riemann tensor* $R^{\mu}_{\alpha\beta\gamma}$, and R is a *curvature scalar* derived from the Ricci tensor through $R = R^{\mu}_{\mu} = g^{\mu\nu}R_{\mu\nu}$. Getting into the details of the Riemann tensor is tedious and does not benefit the rest of the book; we therefore skip the details and refer the readers to a full description in Misner et al. (1973), or a concise description in Chapter 1 of Rezzolla and Zanotti (2013). The general idea is that $R^{\mu\nu}$ is a tensor that contains second derivatives of the metric ($g^{\mu\nu}$) which describes dynamically how the curved space-time evolves. Notice that in Eq. (3.101) one has the freedom to add a linear term of $g^{\mu\nu}$ (the second term on the left), so Einstein added the famous "Λ term" initially to allow the universe to stay in a steady state. After learning about the discovery of the expansion of the universe from Edwin Hubble, Einstein withdrew the Λ term, and remarked to George Gamow that introducing that term was the biggest blunder of his life. Recent cosmological observations indicate that the universe is being accelerated by an unknown entity called "dark energy", the simplest form of which is Einstein's Λ term. For the strong-field regime relevant to the GRB central engine, the Λ term can be safely neglected.

The *metric tensor* is defined as $g_{\mu\nu}$, the *covariant* form of the *contravariant* tensor $g^{\mu\nu}$. The two forms are connected with each other through

$$g^{\alpha\mu}g_{\alpha\nu} = \delta^{\mu}_{\nu}, \tag{3.107}$$

where

$$\delta^{\mu}_{\nu} = \begin{cases} 1, & \mu = \nu; \\ 0, & \mu \neq \nu \end{cases} \tag{3.108}$$

is the Kronecker delta. The physical meaning of the metric tensor $g_{\mu\nu}$ is that it describes the *distance* in space-time of any two *events*, through the *space-time interval*

$$ds^2 = g_{\mu\nu}dx^{\mu}dx^{\nu} = -c^2 d\tau^2, \tag{3.109}$$

where ds is the four-dimensional differential distance between two "continuous" events,

$$x^{\mu} = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{3.110}$$

and

$$x^{\mu} + dx^{\mu} = \begin{pmatrix} x^0 + dx^0 \\ x^1 + dx^1 \\ x^2 + dx^2 \\ x^3 + dx^3 \end{pmatrix}, \tag{3.111}$$

[5] A higher order tensor can be "contracted" to a lower order one, e.g. $R_{\alpha\beta} = R^{\mu}_{\alpha\beta\mu}$.

and $d\tau$ is the differential of the comoving-frame *proper time*. In order to better understand the "distance" in four-dimensional space-time, one can consider three extreme cases:

- Consider an object at a fixed location but at two different times $t_2 = t_1 + dt > t_1$. The 4-distance is $ds^2 = -c^2 d\tau^2 = -c^2 dt^2 < 0$. In general, the $ds^2 < 0$ regime is called *time-like*;
- Consider a certain epoch ($dt = 0$) when two objects remain at two fixed locations separated by $dr^2 = dx^2 + dy^2 + dz^2$. The 4-distance is $ds^2 = dr^2 > 0$. In general, the $ds^2 > 0$ regime is called *space-like*;
- Consider a beam of light traveling from space-time event (ct_1, x_1, y_1, z_1) to $(ct_2 = c(t_1 + dt), x_2 = x_1 + dx, y_2 = y_1 + dy, z_2 = z_1 + dz)$. The 4-distance is $ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 = 0$. This is *light-like*.

The simplest metric is that of the Minkowski (flat) space-time:

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \eta^{\mu\nu}, \tag{3.112}$$

which is valid strictly when no matter/energy exists in space at all. General relativity states that whenever/wherever mass/energy exists, the metric is modified, and it is space and time dependent. The Einstein field equations are designed to solve for $g_{\mu\nu}$ as a function of x^μ given the energy–momentum tensor $T^{\mu\nu}$ at that coordinate x^μ.

In realistic systems, since the mass/energy distribution is continuously evolving in space and time, solving the Einstein field equations is a daunting task. Unless the system is simple enough, there is no analytical solution, and the equations need to be solved numerically. Technically, even a system containing two point mass sources (e.g. merger of two black holes) suffers from huge computational challenges.

3.6.2 Schwarzschild Metric

Only rare systems have analytical solutions to the Einstein field equations. One simple problem is the gravitational field of a point mass M. The solution was found by Karl Schwarzschild in 1916 within the year after Einstein published his full theory (Schwarzschild 1916):

$$ds^2 = -c^2 d\tau^2 = -\left(1 - \frac{r_s}{r}\right) c^2 dt^2 + \frac{dr^2}{\left(1 - \frac{r_s}{r}\right)} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2, \tag{3.113}$$

or

$$g_{\mu\nu} = \begin{pmatrix} -\left(1 - \frac{r_s}{r}\right) & 0 & 0 & 0 \\ 0 & \left(1 - \frac{r_s}{r}\right)^{-1} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix}, \tag{3.114}$$

where

$$r_s \equiv \frac{2GM}{c^2} \simeq 3\ \text{km}\frac{M}{M_\odot} \tag{3.115}$$

is the *Schwarzschild radius* ($M_\odot$ is solar mass), which defines the *event horizon* of a black hole inside which no light can escape. Here the spatial component is expressed in terms of a spherical coordinate system (r, θ, ϕ), with $r = 0$ defined at the location of the point mass. This solution is in the rest frame of an observer who is at $r = \infty$ from the point source but is at rest with respect to the point source. This solution is not time dependent, since the system is in a steady state (a point source with a constant mass and not moving). It is spatially dependent, but only depends on r (spherically symmetric).

One interesting feature is that, at a fixed point in space ($dr = d\theta = d\phi = 0$), the time measured by a distant observer, dt, is related to the proper time, $d\tau$, through

$$d\tau = \left(1 - \frac{r_s}{r}\right)^{1/2} dt. \tag{3.116}$$

This suggests a gravitational time dilation effect: when a brave astronaut (A) stays close to a black hole, and sends light signals regularly in his own frame (constant proper time $d\tau$, say once every second) towards his friend (B) far away from the black hole, astronaut B would measure a time interval longer than 1 s. The closer to r_s, the longer the time interval dt. In the extreme case when the brave astronaut A is at $r = r_s$, dt would be infinite. Since the wavelength of light is proportional to dt, the light with an original wavelength λ_0 sent by A would be recorded by B at a much longer wavelength $\lambda_{\text{obs}} = (1 - r_s/r)^{-1/2}\lambda_0$. This gives rise to a *gravitational redshift*:

$$z = \frac{\lambda_{\text{obs}} - \lambda_0}{\lambda_0} = \left(1 - \frac{r_s}{r}\right)^{-1/2} - 1. \tag{3.117}$$

3.6.3 Kerr Metric

When a point mass M carries an angular momentum J, the metric is much more complicated, yet an analytical solution is available. After the Schwarzschild solution, it took almost half a century before Roy Kerr discovered this new metric (Kerr, 1963), which is written in the so-called Boyer–Lindquist coordinates:[6]

$$ds^2 = -g_{00}(cdt)^2 - g_{03}(cdt)d\phi + g_{11}dr^2 + g_{22}d\theta^2 + g_{33}d\phi^2, \tag{3.119}$$

where

$$g_{00} \equiv \left(\frac{\Delta - a^2\sin^2\theta}{\Sigma}\right),$$

$$g_{03} \equiv \frac{2a\sin^2\theta(r^2 + a^2 - \Delta)}{\Sigma},$$

[6] The coordinate transformation from Boyer–Lindquist coordinates (r, θ, ϕ) to Cartesian coordinates (x, y, z) is given by (Boyer and Lindquist, 1967)

$$\begin{aligned} x &= \sqrt{r^2 + a^2}\sin\theta\cos\phi, \\ y &= \sqrt{r^2 + a^2}\sin\theta\sin\phi, \\ z &= r\cos\phi. \end{aligned} \tag{3.118}$$

$$
\begin{aligned}
g_{11} &\equiv \frac{\Sigma}{\Delta}, \\
g_{22} &\equiv \Sigma, \\
g_{33} &\equiv \left[\frac{(r^2+a^2)^2 - \Delta a^2 \sin^2\theta}{\Sigma}\right] \sin^2\theta,
\end{aligned}
\tag{3.120}
$$

with

$$r_s \equiv \frac{2GM}{c^2}, \tag{3.121}$$

$$a \equiv \frac{J}{cM}, \tag{3.122}$$

$$\Delta \equiv r^2 - r_s r + a^2, \tag{3.123}$$

$$\Sigma \equiv r^2 + a^2 \cos^2\theta. \tag{3.124}$$

In the tensor form, one can write

$$g_{\mu\nu} = \begin{pmatrix} -g_{00} & 0 & 0 & -g_{03}/2 \\ 0 & \Sigma/\Delta & 0 & 0 \\ 0 & 0 & \Sigma & 0 \\ -g_{03}/2 & 0 & 0 & g_{33} \end{pmatrix}. \tag{3.125}$$

For a Kerr metric, there are two critical radii. The first one is the radius above which an object can stay at rest with respect to a distant observer. In other words, $ds^2 = -(cd\tau)^2$ should be negative when the spatial component in the metric tensor is zero, i.e. $dr = 0$. This "time-like" condition is $g_{00} > 0$, which gives a θ-dependent solution:

$$r > r_0(\theta) \equiv \frac{r_s}{2} + \sqrt{\left(\frac{r_s}{2}\right)^2 - a^2\cos^2\theta} = \frac{GM}{c^2} + \sqrt{\left(\frac{GM}{c^2}\right)^2 - a^2\cos^2\theta}. \tag{3.126}$$

The second critical radius is the true event horizon. This is the boundary for the "time-like" condition in the polar direction $\theta = 0$, where there is no spin effect. This defines an angle-independent characteristic radius

$$r > r_+ \equiv \frac{r_s}{2} + \sqrt{\left(\frac{r_s}{2}\right)^2 - a^2} = \frac{GM}{c^2} + \sqrt{\left(\frac{GM}{c^2}\right)^2 - a^2}, \tag{3.127}$$

which defines the event horizon for all angles θ. The region in between ($r_+ < r < r_0(\theta)$) is the *ergosphere*, inside which objects cannot stay static with respect to a distant observer due to the *frame-dragging* effect, but they can in principle remain in a *stationary* orbit without falling into the black hole (if they have a sufficiently large angular momentum). Energy from the ergosphere can be extracted by the *Penrose mechanism* (Penrose and Floyd, 1971), which states the following: Consider a particle (say, particle A) that enters the ergosphere of a black hole and splits into two particles (say, B and C). Energy conservation states $E(A) = E(B) + E(C)$. Let us assume $E(C) < 0$ so that it will fall into the event horizon; then one has $E(B) > E(A)$, so that particle B can in principle leave the ergosphere, carrying away the spin energy of the black hole. In astrophysical problems, this process can be very efficient with the existence of a large-scale magnetic field threading the ergosphere (Blandford and Znajek, 1977).

3.6.4 Kerr–Newman Metric

The *no hair theorem* of black holes states that all black hole solutions to the Einstein–Maxwell equations can be completely characterized by only three externally observable parameters. Besides mass and angular momentum, another parameter is charge Q. Including all three parameters, the metric becomes the Kerr–Newman metric.

The Kerr–Newman metric carries the same form (Eqs. (3.119) and (3.120)) as the Kerr metric in the Boyer–Lindquist coordinates, except that Eq. (3.123) is replaced by

$$\Delta = r^2 - r_s r + a^2 + r_{\rm Q}^2, \tag{3.128}$$

where

$$r_{\rm Q}^2 = \frac{GQ^2}{c^4}, \tag{3.129}$$

which denotes the length scale corresponding to the charge Q of the mass.

For $J = 0$ (i.e. $a = 0$), the Kerr–Newman metric can be reduced to the Reissner–Nordström metric, which reads (in the standard coordinate system)

$$ds^2 = -\left(1 - \frac{r_s}{r} + \frac{r_{\rm Q}^2}{r^2}\right) c^2 dt^2 + \frac{dr^2}{\left(1 - \frac{r_s}{r} + \frac{r_{\rm Q}^2}{r^2}\right)} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2. \tag{3.130}$$

Exercises

3.1 Show that the Galilean transformation

$$\left\{ \begin{array}{lcl} dt' & = & dt \\ dx' & = & dx - \beta c dt \\ dy' & = & dy \\ dz' & = & dz \end{array} \right. \tag{3.131}$$

does not satisfy Eq. (3.6).

3.2 Use the two postulates of special relativity to prove that the only transformation that satisfies Eq. (3.6) is the Lorentz transformation Eq. (3.4).

3.3 Reverse the direction of the moving shell, and re-derive the relationships of $\Delta t_{\rm eng}$, Δt_e, and $\Delta t_{\rm obs}$.

3.4 Prove Eqs. (3.60)–(3.62).

3.5 Derive the properties of the EATS of a constant Lorentz factor shell.