Chapter 3 What the transit light curve tells us

Introduction

We saw in Chapter 1 that the depth of a transit is proportional to the square of the ratio of the planet radius to the host star radius. In Chapter 2 we asserted that the transit light curve allows us to obtain a precise measurement of the orbital inclination, i, of the planet, and consequently allows an exact value of the planet's mass to be deduced from the radial velocity curve of the host star. Now we will explore the analysis of a transit light curve and learn exactly how the orbital inclination and other precise parameters for the planet and its host star are obtained.

In the first section of this chapter we will begin with Kepler's third law and use it to deduce how the transit duration depends on the orbital inclination, i, and the more fundamental parameters of the star and planet. In Section 3.2 we analyze the shape of the transit light curve, deriving quantitative models for the transit shape in the case of a uniform stellar disc and a limb darkened stellar disc. In Section 3.3 we step back from the analytical detail and summarize the logic behind the deduction of values for the radii, stellar and planetary masses, and orbital inclination. Finally, in Section 3.4 we discuss the light curve model fitting process that yields parameters for transiting planet systems.

This chapter underpins the key role of transiting exoplanets in the wider field of planetary astrophysics.

3.1 Kepler's third law and exoplanet orbits

Kepler's third law underlies much of astrophysics. It is a beautifully simple relationship that was deduced from observations of the Solar System planets, and can be derived from Newton's law of universal gravitation and Newton's second law of motion. Kepler's third law provides the basis for the quantitative analysis of planetary orbits, and is the starting point for analyzing exoplanetary transits.

3.1.1 The semi-major axis

One of the easiest things to measure once a transiting exoplanet is discovered is the orbital period, P (cf. Equation 2.1). The orbital period is related to the semi-major axis of the orbit, a, and the total mass of the two bodies via Kepler's third law:

$$\frac{a^3}{P^2} = \frac{G(M_* + M_{\rm P})}{4\pi^2}. \qquad \text{(Eqn 1.1)}$$

Generally $M_* \gg M_{\rm P}$, and we can obtain a fairly good estimate of the mass of the star from its spectral type. Consequently, the semi-major axis, a, is known

approximately once the orbital period and the spectral type of the host star have been measured:

$$a \approx \left(GM_* \left(\frac{P}{2\pi} \right)^2 \right)^{1/3}. \qquad \text{(Eqn S2.2)}$$

3.1.2 The orbital speed

With values for a and P, the orbital speed can be deduced. This is simplest to derive for the case of a circular orbit, but is equally true for the general case of an elliptical orbit. Kepler's second law states that a planet orbiting in an elliptical orbit sweeps out equal areas in equal times, i.e. it moves fastest at periapsis (cf. Figure 1.14) when it is closest to the star, and slower when it is further from the star.

For a circular orbit, the orbital speed is a constant and is given by

$$v = \frac{2\pi a}{P}. \qquad (3.1)$$

If we are calculating the speed of the planet about the barycentre of the star–planet system, then the semi-major axis here is a_{P}, as defined in Section 1.2.

- Which fundamental conservation law of physics underpins Kepler's second law?
- The principle of conservation of angular momentum: since no net external torque is acting on the star–planet system, the total angular momentum remains constant.
- As written, with a in the numerator of the right-hand side, what speed does Equation 3.1 give?
- a is the semi-major axis of the astrocentric orbit, so Equation 3.1 gives the speed of the planet relative to the star.
- In applying Equation 3.1 to radial velocity measurements of the type discussed in Section 1.3, what quantity should be used for the semi-major axis?
- The radial velocity that can be measured is that of the stellar reflex orbit. The semi-major axis of this orbit relative to the barycentre should be used, i.e. a_*. This quantity is much smaller than the semi-major axis of the astrocentric orbit of the planet, a, or the semi-major axis of the barycentric orbit of the planet, a_{P}.
- What is the relationship between the three semi-major axes mentioned in the previous answer?
- $a = a_{\mathrm{P}} + a_*$.

Exercise 3.1 (a) What is the orbital speed of the Earth around the Sun? (You may assume that the orbit is circular.)

(b) What would be the orbital speed of a planet in a circular orbit at a distance of 1 AU from a star of mass $0.5\,\mathrm{M}_\odot$? ■

3.1.3 The orbital inclination, the impact parameter and the transit duration

The limb is the term used to refer to the edges of the disc of any astronomical body.

Phases of a transit

A transit has four **contacts**; these are illustrated in Figure 3.1. **First contact** is when the **limb** of the planet's disc first coincides with the limb of the star's disc as viewed by the observer. **Second contact** is when the entire disc of the planet is just within the stellar disc as viewed by the observer.

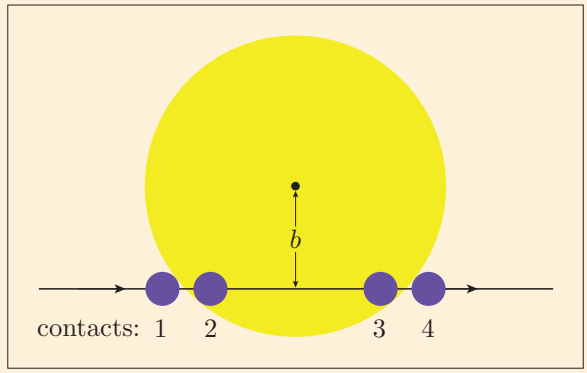

Figure 3.1 The position of the planet's disc relative to the star's disc at the four contact points. If the orbital motion of the planet is in the direction shown, the four planet discs correspond to first, second, third and fourth contacts, respectively. These contact points correspond to measurable features in the light curve.

Third contact occurs when the limb of the planet's disc coincides with the limb of the star's disc, i.e. the last instant when the entire disc of the planet is just within the stellar disc. **Fourth contact** occurs when the trailing limb of the planet's disc crosses the limb of the star's disc, i.e. the last instant of the transit.

In Subsection 2.2.2 we derived the duration of the transit that would be observed from an orbital inclination of $i = 90°$, or equivalently an impact parameter of $b = 0.0$:

$$T_{\rm dur}(b = 0.0) = \frac{P}{\pi} \sin^{-1}\left(\frac{R_*}{a}\right). \qquad \text{(Eqn S2.8)}$$

We also noted that an approximate version of this equation (Equation 2.6) was sufficient for the purposes of the analysis in Chapter 2. Now we are going to examine how to use transit light curves to derive precise parameters, so we will use the exact version (Equation S2.8), and derive the dependence on the orbital inclination. Figure 1.19 allows us to derive the relationship between the impact parameter and the orbital inclination. In Figure 1.19 the observer is viewing along a horizontal line of sight, while the planet's orbital plane is inclined and makes an angle i with the vertical as shown. A version of Figure 1.19 emphasizing the trigonometry is reproduced here as Figure 3.2.

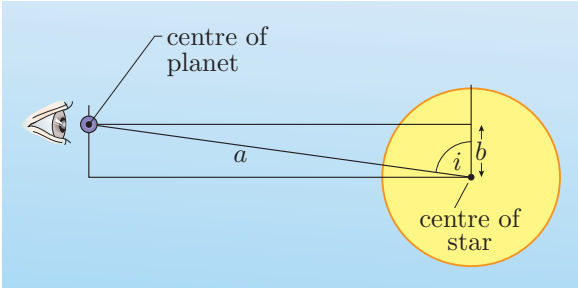

Figure 3.2 The geometry of Figure 1.19, emphasizing the trigonometry determining the impact parameter, $b = a\cos i$.

The impact parameter, b, is the vertical distance at mid-transit of the centre of the planet from the centre of the star as viewed by the observer. This distance is

$$b = a\cos i. \tag{3.2}$$

The transit duration depends strongly on the impact parameter. Figure 3.3 shows the relevant geometry in a snapshot taken at fourth contact. The larger circle represents the disc of the star, while the smaller circle represents the disc of the planet.

Note in some other texts the impact parameter is defined as the dimensionless quantity

$$b = \frac{a\cos i}{R_*}.$$

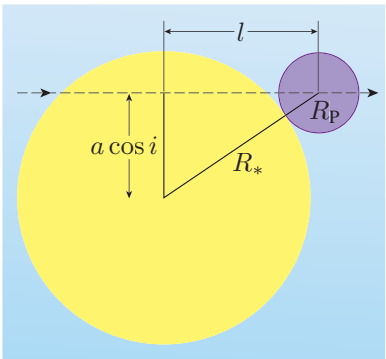

Figure 3.3 Pythagoras's theorem allows us to express the length l in terms of the impact parameter, b, and the radii of the star and planet.

Figure 3.3 shows a right-angled triangle that has a hypotenuse of length $R_* + R_P$ and a vertical side that is equal to the impact parameter, b, and therefore has length $a\cos i$. The third, horizontal, side of the triangle joins the positions of the centre of the planet's disc at mid-transit and fourth contact. By Pythagoras's theorem, the length of this side is

$$l = \sqrt{(R_* + R_P)^2 - a^2\cos^2 i}. \tag{3.3}$$

The path of the planet is not precisely along this straight line, but is the flattest part of the extremely eccentric ellipse formed by the projection of the highly inclined orbit on the plane of the sky.

- For $a^2\cos^2 i > (R_* + R_P)^2$, the quantity under the square root is negative, and the length l would be a purely imaginary number. What physical situation does this correspond to?
- If $a^2\cos^2 i$ exceeds $(R_* + R_P)^2$, then the impact parameter is greater than the sum of the stellar and planetary radii, i.e. the planet's disc moves across the diagram above the star's disc without ever overlapping it, and no transit will occur.

Figure 3.4 illustrates how the transit duration can be deduced from the length l. During the transit, the planet moves along its orbit from point A to point B, subtending an angle α at the star. From the triangle formed by A, B and the centre of the star, $\sin(\alpha/2) = l/a$. With the angle α in radians, for a circular orbit the time elapsed is simply

$$T_{\rm dur} = P\frac{\alpha}{2\pi} = \frac{P}{\pi}\sin^{-1}(l/a),$$

thus

$$T_{\rm dur} = \frac{P}{\pi}\sin^{-1}\left(\frac{\sqrt{(R_* + R_{\rm P})^2 - a^2\cos^2 i}}{a}\right). \qquad (3.4)$$

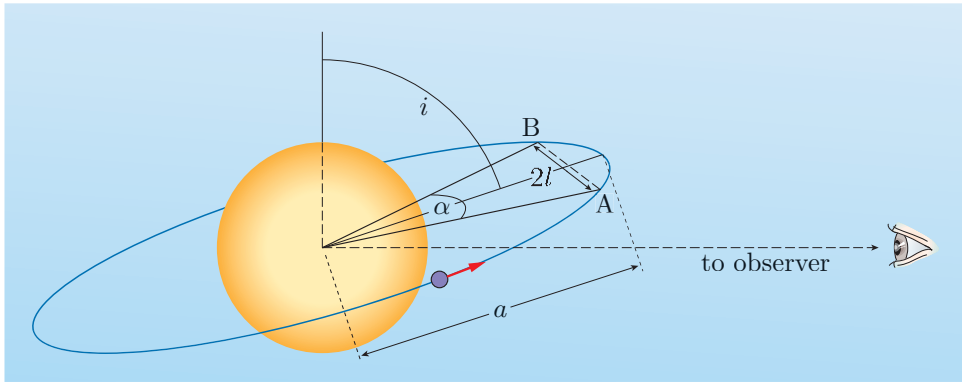

Figure 3.4 During transit, the planet moves from A to B around its orbit. If the orbit is circular, the distance around the entire orbit is $2\pi a$, and the arc length between A and B is $a\alpha$, with the angle α in radians. The distance along a straight line between A and B is $2l$. From the triangle formed by A, B and the centre of the star, $\sin(\alpha/2) = l/a$.

This is the exact general expression for the duration of a transit for an exoplanet in a circular orbit. For a non-circular orbit, the time taken to move from point A to point B in Figure 3.4 is not simply proportional to the angle α.

- What could we use instead of the above to work out the transit duration for an elliptical orbit?
- Kepler's second law tells us that equal areas are swept out in equal times. If the elliptical orbit is known, we can use this principle to work out the time between any two points on the orbit.
- Why is it appropriate to use a rather than $a_{\rm P}$ in Equation 3.4?
- Because it is the astrocentric motion that determines when the contact points occur. If we used $a_{\rm P}$, the motion of the star would cause an underestimate of the transit duration.

Exercise 3.2 (a) Show that if $a \gg R_* \gg R_{\rm P}$, then Equation 3.4 simplifies to

$$T_{\rm dur} \approx \frac{P}{\pi}\left[\left(\frac{R_*}{a}\right)^2 - \cos^2 i\right]^{1/2}.$$

(b) Use this to work out the orbital inclination of a planet discovered around a Sun-like star (i.e. $M_* = 1\,{\rm M}_\odot$, $R_* = 1\,{\rm R}_\odot$) with $P = 6$ days and $T_{\rm dur} = 2$ hours. Radial velocity measurements indicate that the orbit of the planet is circular.

(c) What would you deduce about a transit candidate with $T_{\rm dur} = 4$ hours and other characteristics identical to those given above? ■

We showed in Exercise 3.2 that if we assume a circular orbit and $a \gg R_* \gg R_\mathrm{P}$, then Equation 3.4 simplifies to

$$T_\mathrm{dur} \approx \frac{P}{\pi}\left[\left(\frac{R_*}{a}\right)^2 - \cos^2 i\right]^{1/2}. \tag{3.5}$$

If we know the spectral type of the star, we can infer approximate values of M_* and R_* by assuming that they are typical for the spectral type. Then the transit duration, T_dur, and the orbital period, P, are sufficient to deduce (via Equation 3.5) the approximate orbital inclination, i.

- How is the value of a to be used in Equation 3.5 determined from the quantities mentioned in the paragraph above?
- Kepler's third law gives a in terms of M_* and P.

Furthermore, once the orbital inclination is estimated in this way, we can use the time elapsed between first and second contact, t_{1-2}, to obtain a second estimate of the radius of the planet. At transit ingress the planet's disc moves from first contact to second contact as shown in Figure 3.1. The time taken to do this depends on the orbital inclination, the sizes of the two discs, and the orbital speed. The transit duration and the orbital period give us the first and last of these quantities, so a previously known value for R_* and a measurement of t_{1-2} implies a value of R_P. Of course, the transit depth has already given us an estimate of the ratio of the radii:

$$\frac{\Delta F}{F} = \frac{R_\mathrm{P}^2}{R_*^2}. \tag{Eqn 1.18}$$

Thus a light curve of sufficient quality to provide a precise measurement of t_{1-2} as well as the transit depth, ΔF, allows two independent estimates of the planet's radius, given the radius of the star. If these two estimates do not agree, then a new value can be assumed for R_* to see if it leads to more consistent results.

3.2 The shape of the transit light curve

In our discussion so far we have related measurements of particular features in the light curve to the physical parameters of the star and its transiting planet. The features discussed so far are:

- the transit depth, ΔF;
- the transit duration, T_dur;
- the duration of ingress, t_{1-2} (or equivalently egress, t_{3-4}).

Figure 3.5 shows the first light curve of HD 209458 b's transit taken using the Hubble Space Telescope. The data have a sampling cadence of 80 s, and the out-of-transit data points for an individual measurement of the stellar flux, F, have a signal-to-noise ratio of almost 10^4: the scatter among the measurements of F is $1.1 \times 10^{-4} F$. This exquisite light curve provides the opportunity to precisely measure each of the three quantities listed above.

- Is the light curve in Figure 3.5 completely described by the three quantities listed above?

❍ No. The high precision light curve in Figure 3.5 clearly has a curved transit floor. This curvature is not characterized by any of the three quantities above.

- What does the shape of the transit floor imply about the brightness distribution of the stellar disc?

❍ The transit depth varies smoothly, being deepest at the centre, implying that most light is lost when the planet is farthest from the limb of the stellar disc. This implies that the stellar disc gets smoothly brighter from the limb to the centre.

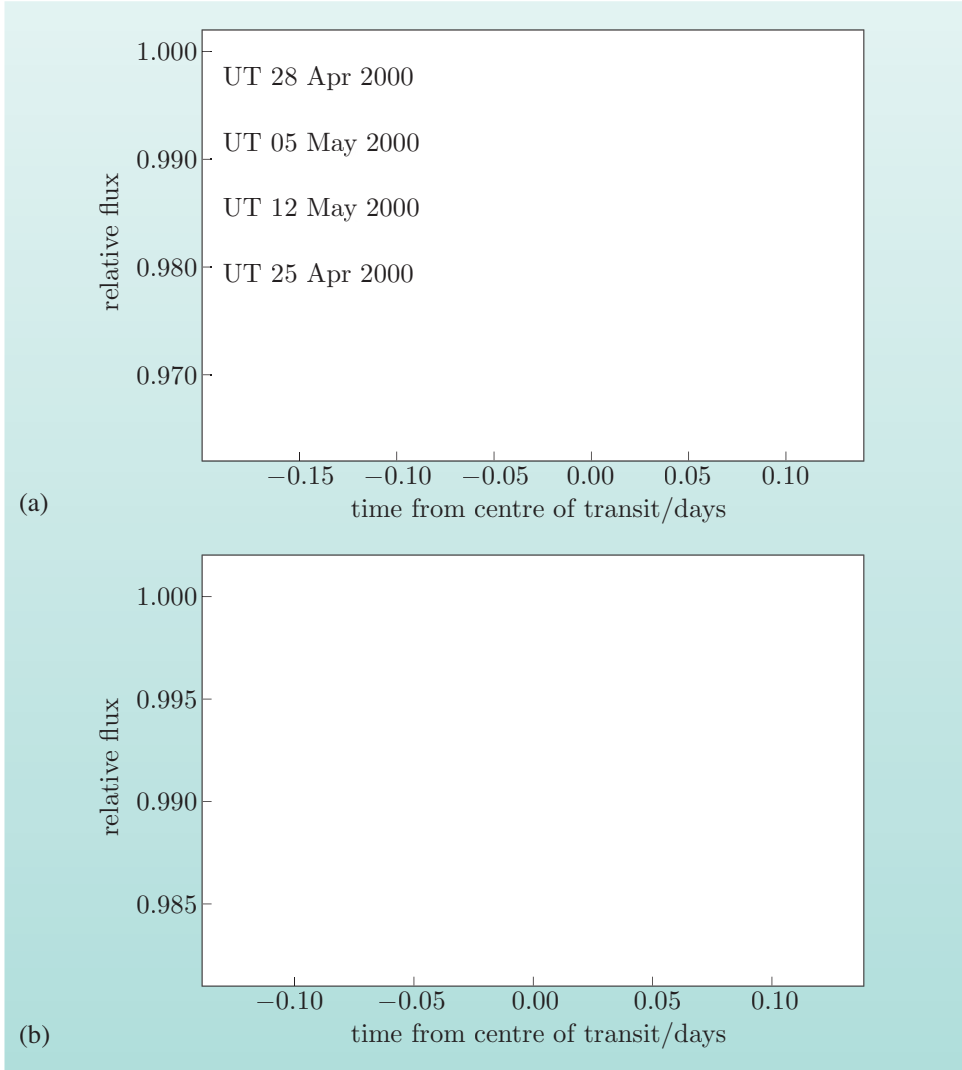

Figure 3.5 The first exquisitely high-precision transit light curve. (a) Hubble Space Telescope photometry of HD 209458 b during four different transits in April and May 2000. The four light curves have been offset from each other for clarity. The gaps in each individual light curve are caused by the Earth regularly occulting HD 209458 b as Hubble proceeds around its low Earth orbit. (b) The phased light curve obtained by combining the four sets of observations. The data have been normalized so that the out-of-transit level is exactly 1.000. The depth of the transit is 1.64%, and the noise level is orders of magnitude smaller than this.

Exercise 3.3 (a) Calculate the signal-to-noise ratio for the transit shown in Figure 3.5, where the signal is the transit depth, ΔF, and the noise is estimated by the scatter between out-of-transit points.

(b) Estimate the maximum scatter in the measurements of F that would allow detection of the curvature in the transit floor for the transit shown in Figure 3.5. ■

3.2.1 Limb darkening

Figure 3.5 makes it abundantly clear that the transit is not a flat-bottomed dip in the light curve. The curvature in the transit floor implies that a smoothly varying fraction of the stellar flux is lost during the transit, so

$$\frac{\Delta F}{F} = \frac{R_{\rm P}^2}{R_*^2} \qquad \text{(Eqn 1.18)}$$

cannot be strictly true unless one or both of the radii vary. It is extremely difficult to imagine any mechanism that would cause such a variation to occur at the time required to produce a symmetrical transit as viewed from a planet orbiting a random G type star; instead, we need to examine the assumptions implicit in Equation 1.18. As already noted, the shape of the transit floor suggests that the stellar disc is brightest at the centre and gets smoothly darker towards the limb. In reaching Equation 1.18 we implicitly assumed that the stellar disc has a uniform brightness, but in fact we know that this is not the case for the Sun or for stars in general, as discussed in the box on limb darkening below.

Limb darkening

Figure 3.6 shows a broad-band optical image of the Sun. There are some immediately obvious features in this image: the dark **sunspots**. The solar disc appears noticeably brighter at the centre, and appears progressively dimmer and redder towards the limb. This **limb darkening** is easily understood in terms of the radiative transfer physics of stellar atmospheres, and has important consequences for the shape of a transit light curve.

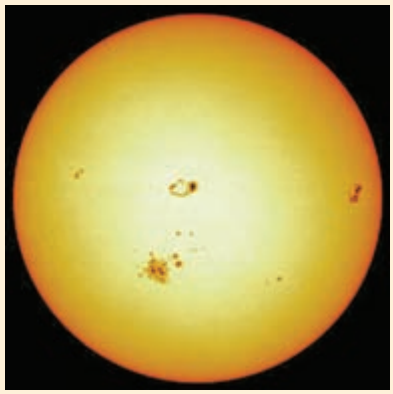

Figure 3.6 A NASA image of the Sun. The various dark spots are sunspots. The solar disc appears noticeably brighter at the centre, and appears progressively dimmer and redder towards the limb. This phenomenon is limb darkening.

Main sequence stars do not have solid surfaces: their outer layers are composed of plasma. The light that escapes from them comes from a variety of depths within the stellar atmosphere. The probability of escape of a photon emitted within a particular layer of the atmosphere is dependent on the **optical depth** of that layer. At a given frequency, ν, the optical depth, τ_ν, is the integral of the opacity, κ_ν, multiplied by the density, $\rho(s)$, along the path taken by a light ray; i.e. for light emitted at position X travelling towards an observer at infinity, as indicated in Figure 3.7, we have

$$\tau_\nu = \int_{\mathrm{X}}^{\infty} \rho(s)\,\kappa_\nu\,\mathrm{d}s, \tag{3.6}$$

where s indicates the position along the path.

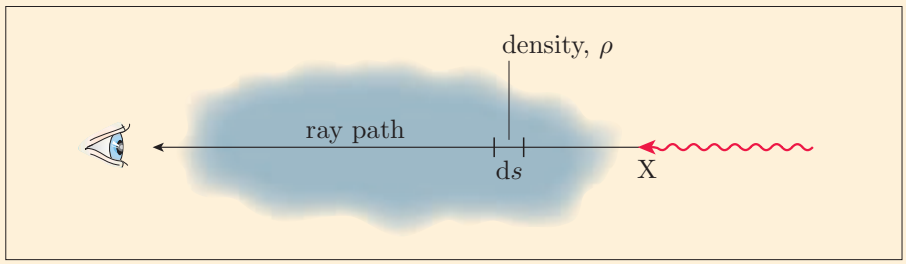

Figure 3.7 The optical depth is the integral along the ray path of the density multiplied by the opacity.

Because the opacity is dependent on the frequency (or equivalently the wavelength) of the radiation, the optical depth is also dependent on this. A particular physical depth in a stellar atmosphere will have a different optical depth depending on the frequency of radiation being considered. The probability of a photon travelling along the path without being absorbed or scattered is $\mathrm{e}^{-\tau_\nu}$, so the ratio between the emitted intensity, I_{emitted}, and the emergent intensity, I, is

$$\frac{I}{I_{\mathrm{emitted}}} = \mathrm{e}^{-\tau_\nu}. \tag{3.7}$$

At any position within the stellar atmosphere, photons are being continually emitted, absorbed and scattered. By definition, any photon detected by the observer at infinity travelled in the direction towards the observer. This means that a photon that emerges from the centre of the stellar disc is travelling radially outwards through the stellar atmosphere, as illustrated in Figure 3.8a.

On the other hand, a photon emerging from anywhere else on the stellar disc is travelling at an angle γ to the outwardly directed radius vector. The two photons indicated in Figure 3.8a have been emitted at the same depth in the stellar atmosphere, but to reach the observer, the photon emitted near the limb of the star must travel through a much greater path length, s, of the stellar atmosphere. For the photons emitted at depth h, the path length through the stellar atmosphere is

$$s \approx \frac{h}{\cos\gamma} = \frac{h}{\mu}, \tag{3.8}$$

where $\mu = \cos\gamma$ and the **plane-parallel** approximation has been made, i.e. we have assumed $h \ll R_*$ so the geometry is essentially that of Figure 3.8b.

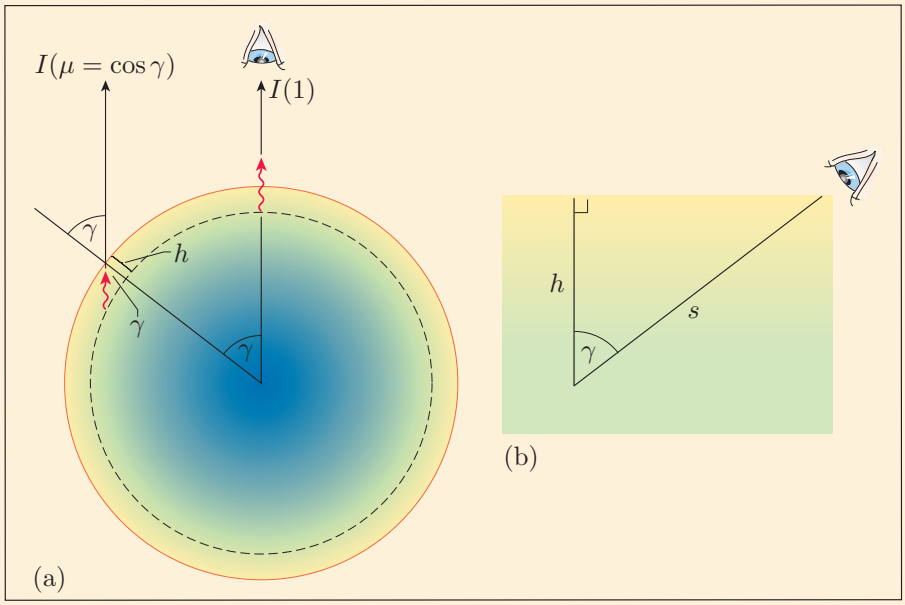

Figure 3.8 Cross-sections through a star indicating the depth and direction of travel of escaping photons. (a) Light travelling towards an observer beyond the top of the page at infinity is emitted in a radially outwards direction at the centre of the stellar disc. Conversely, light travelling towards this same observer from elsewhere on the disc is travelling at an angle γ to the radially outwards direction. The radially outwards direction minimizes the path length traversed from any depth, h. Light from a given depth is consequently more likely to escape for small angles γ. Thus the centre of the disc is brightest and the limb of the disc is faintest. (b) Light escapes only from the very outermost part of the star. The geometry is therefore effectively the plane-parallel atmosphere shown here, with $h \ll R_*$.

The optical depth for a given physical depth increases towards the limb of the star. A smaller fraction of the photons emitted at any particular depth escape to reach the observer from the limb of the star. Consequently, the stellar disc appears progressively dimmer towards the limb. The shorter path length travelled through the atmosphere by photons travelling radially outwards also means that more photons escape from deeper within the atmosphere at the centre of the disc. Generally, deeper layers are hotter, and radiation emitted there consequently approximates a bluer black body spectrum. This means that the disc appears progressively redder towards the limb.

The Sun is the only star for which we can currently measure the limb darkening directly and in detail. Unfortunately, the exact form of the limb darkening is a function of both opacity and emissivity at each depth within the atmosphere. These are dependent on the wavelength being considered, and the composition and thermodynamic properties of the atmosphere at

each point. Thus the limb darkening will be dependent on the spectral type of the star, and on its detailed composition. Some empirical results for the solar limb darkening at two different wavelengths are shown in Figure 3.9.

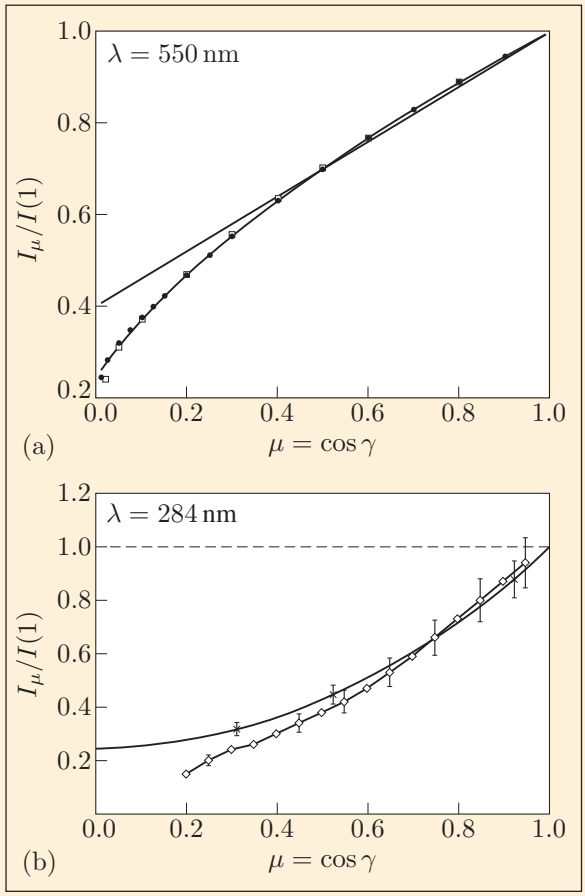

Figure 3.9 Limb darkening for the Sun. Each panel shows $I(\mu)$ as a function of μ for a particular wavelength. (a) Squares are measurements made at 550 nm; dots are the corresponding results from solar models. The solid lines give linear and logarithmic limb darkening relationships (see text). (b) At 284 nm, two alternative sets of measurements are shown; the upper one produced results for all angles shown. The crosses with error bars indicate the measurement uncertainties for selected angles. For $\mu < 0.55$, the two sets of measurements do not agree terribly well.

Even for the Sun, and for a single wavelength, it is difficult to obtain definitive results. In Figure 3.9a the dots show results from a computational model of radiative transfer through the solar atmosphere. The results agree well with the measurements. We cannot make direct observations of limb darkening in stars of spectral type and composition different from the Sun, so the results from computational models are generally used to guide our assumptions about their limb darkening. Researchers often refer to these assumptions as **limb darkening laws**, though they are simply curves adopted as adequately good approximations to the (unknown) stellar limb darkening profile. Limb darkening laws are not laws in the same sense as

Kepler's laws or the laws of thermodynamics. Figure 3.9a shows two relationships, i.e. two possible limb darkening laws, between $I(\mu)$ and μ, plotted as solid lines. The first of these is the simplest linear limb darkening relationship:

$$\frac{I(\mu)}{I(1)} = 1 - u(1-\mu), \tag{3.9}$$

where u is the **limb darkening coefficient** that governs the gradient of the intensity drop between the centre and limb of the disc. The second relationship illustrated in Figure 3.9a is a logarithmic relationship:

$$\frac{I(\mu)}{I(1)} = 1 - u_l(1-\mu) - \nu_l\,\mu\ln\mu, \tag{3.10}$$

which has two limb darkening coefficients, u_l and ν_l.

Two other relationships that may be adopted are the quadratic law

$$\frac{I(\mu)}{I(1)} = 1 - u_{\rm q}(1-\mu) - \nu_{\rm q}(1-\mu)^2 \tag{3.11}$$

and the cubic law

$$\frac{I(\mu)}{I(1)} = 1 - u_{\rm c}(1-\mu) - \nu_{\rm c}(1-\mu)^3, \tag{3.12}$$

with their coefficients $u_{\rm q}$, $\nu_{\rm q}$ and $u_{\rm c}$, $\nu_{\rm c}$, respectively. The more complex relationships given in Equations 3.10, 3.11 and 3.12 have more flexibility, allowing them to fit empirical data more closely, but the cost of this is having two coefficients that must be determined somehow, or simply fixed arbitrarily.

All of these relationships have been used in modelling exoplanet transit light curves. For a particular star, any one of them with the appropriate limb darkening coefficient(s) might provide the best match to the actual (*a priori* unknown) stellar limb darkening.

The wavelength-dependence of the stellar limb darkening is immediately apparent in the multicolour transit light curves shown in Figure 3.10. The longest-wavelength data are shown in the uppermost transit curve, which has a very gently curving transit floor, while the shortest-wavelength transit has a transit floor that is so curved that it is impossible to discern the phases of the second and third contact points from visual inspection of it alone. Each of the observed transits in Figure 3.10 has a best-fitting model transit overplotted. These models are shown overplotted without offsets in Figure 3.11a, clearly showing the wavelength-dependence of transit shape. Figure 3.11b shows the counts spectrum from which these light curves were constructed, and indicates the wavelength ranges that comprise each light curve.

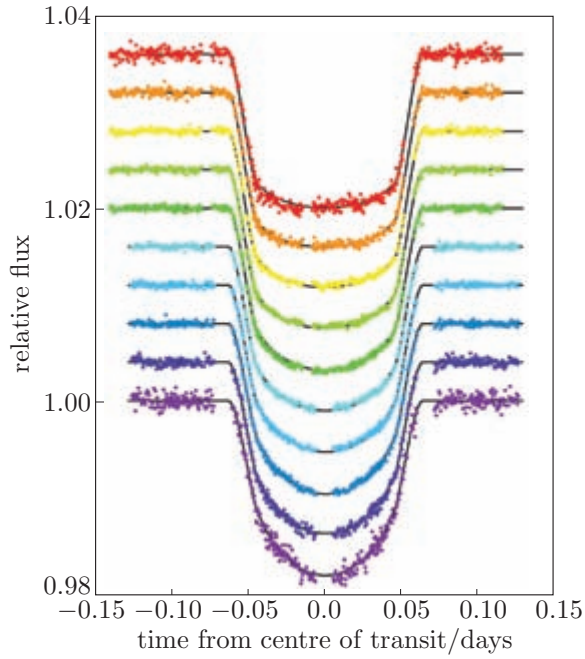

Figure 3.10 Multicolour transit light curves for HD 209458 b taken with the Hubble Space Telescope. All the curves have been normalized to an out-of-transit flux level of 1.00, and the curves have been offset for clarity. The longest wavelength data are at the top, and wavelength decreases for each successive curve. The wavelength ranges corresponding to each colour-coded curve are indicated in Figure 3.11. The solid lines are the model fits to the data.

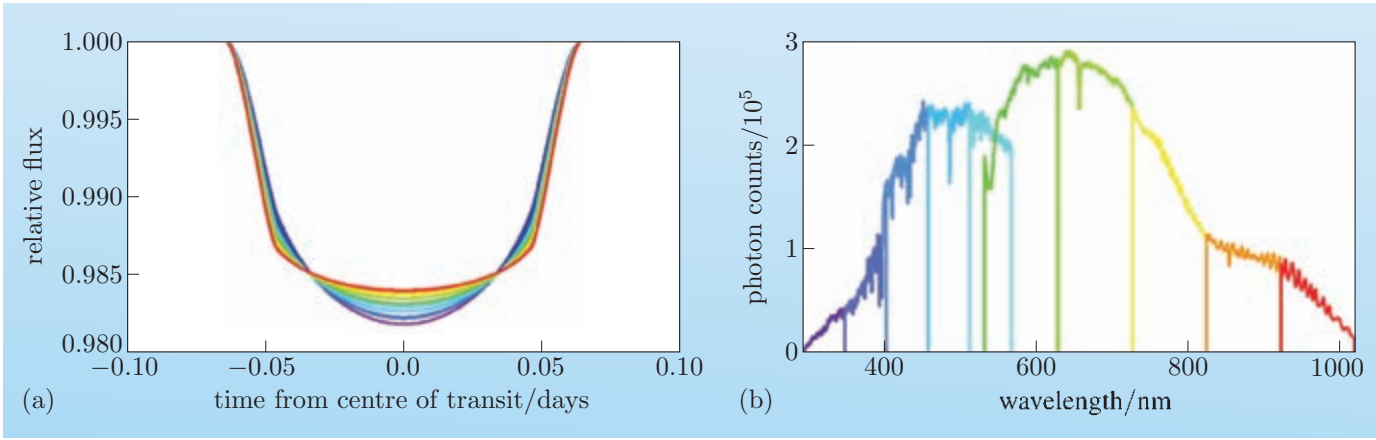

Figure 3.11 (a) The fits to the multicolour transit light curves for HD 209458 b plotted without the offsets. The transit is most flat-bottomed for the reddest wavelengths, where limb darkening is weaker. (b) The counts spectrum of the data used to construct these light curves.

Exercise 3.4 In the paper 'Homogeneous studies of transiting extrasolar planets — I. Light curve analyses' (2008, *Monthly Notices of the Royal Astronomical Society*, **386**, 1644–66) by John Southworth, he fits five different limb darkening laws to the transit light curves of 14 transiting exoplanets. For the exoplanet WASP-1 b, the limb darkening coefficients are determined as follows for the I band light curve:

linear law: $u = 0.215$;

logarithmic law: $u_l = 0.14$, $\nu_l = -0.12$;

quadratic law: $u_\mathrm{q} = 0.29$, $\nu_\mathrm{q} = -0.13$.

For light emerging close to the limb of the star and therefore travelling at an angle $\gamma = 80^\circ$ to the outwardly directed radius vector, what is the fractional intensity,

$I(\mu)/I(1)$, in each case? Does it make a difference which limb darkening law is adopted? ■

3.2.2 The eclipsed area as a function of time

In Figure 3.10 we showed the results of model transit light curves computed and fit to the observed data. An essential step in computing a model transit light curve is to derive a general expression for the area of the star that is occulted at each instant of the transit event. We have seen in Subsection 3.2.1 that the intensity of the disc of the star is a function of μ, or alternatively we could express this intensity distribution as a function of r, the distance from the centre of the stellar disc. We will derive an expression for the occulted area, A_e, that is amenable to being modified to allow for this axially symmetric intensity distribution. The approach follows that of the paper 'Analytic light curves for planetary transit searches' (2002, *Astrophysical Journal Letters*, **580**, L171–5) by Kaisey Mandel and Eric Agol.

The first step in the derivation is to obtain a general expression for the projected separation of the centres of the stellar and planetary discs as a function of the orbital parameters and time. For simplicity we restrict ourselves to circular orbits, and for conciseness we will use the orbital angular speed, ω, where

$$\omega = \frac{2\pi}{P}. \tag{3.13}$$

The orbital phase angle (in radians) at a time t is therefore ωt, and this is related to the orbital phase, ϕ, by

$$\phi = \frac{\omega t}{2\pi}. \tag{3.14}$$

The fiducial phase $\phi = 0$ coincides with inferior conjunction of the planet, i.e. mid-transit ($\omega t = 0$; cf. Figure 3.12a). The stellar orbit is inclined to the line of sight, with orbital inclination i. At inferior conjunction, the projected separation between the centres of the star and planet is $b = a\cos i$, and at any other instant, the separation $s(t)$ is given by the right-angled triangle shown in Figure 3.12b. The locus followed by the planet across the disc of the star is not a straight line, but is part of the ellipse formed by the projection of the orbit on the plane of the sky. The planet moves in a circular orbit, so in astrocentric coordinates, the component of its displacement in the plane of the sky is $a\sin\omega t$, while the other component of its true displacement is $a\cos\omega t$. The full displacement in the plane of the sky is observed, while the other component of the displacement is foreshortened, so that the observed displacement is $a\cos i\cos\omega t$, as illustrated in Figure 3.12b. Using Pythagoras's theorem on the right-angled triangle in Figure 3.12b, we have

$$s^2(t) = (a\sin\omega t)^2 + (a\cos i\cos\omega t)^2. \tag{3.15}$$

Taking the square root and simplifying, we obtain

$$s(t) = a\left(\sin^2\omega t + \cos^2 i\cos^2\omega t\right)^{1/2}. \tag{3.16}$$

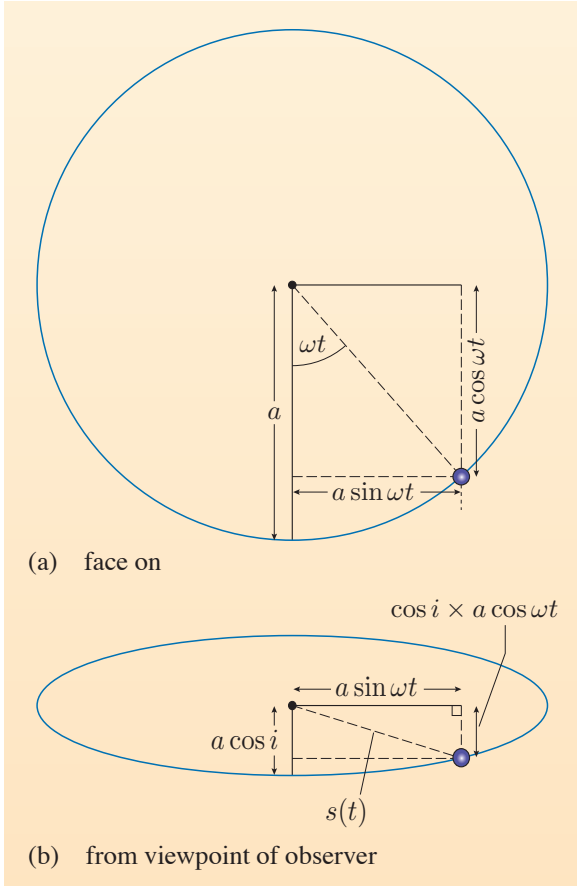

Figure 3.12 (a) The orbit as viewed from above. The displacement between the centres of the star and planet is a vector addition of a component $a \sin \omega t$ in the plane of the sky, and an orthogonal component $a \cos \omega t$. (b) The orbit as viewed along the line of sight of the observer. The projected separation of the centres of the star and planet is $s(t)$. The component of this in the plane of the sky is not foreshortened, but the other component is foreshortened by a factor of $\cos i$.

Exercise 3.5 A planet transits across a star like the Sun ($M_* = 1\,\mathrm{M}_\odot$, $R_* = 1\,\mathrm{R}_\odot$) with an impact parameter $b = 3\,\mathrm{R}_\odot/4$. The inclination angle is $i = 86.5°$. Calculate the position of the planet at three epochs: (i) 1 hour before mid-transit, (ii) at mid-transit, (iii) 1 hour after mid-transit. Then use this information to sketch the locus of the planet transit across the stellar disc. (*Hint*: Use Kepler's third law to calculate the orbital period of the planet, then calculate the orbital phase corresponding to the three times. Use Equation 3.16 to work out the horizontal and vertical components of the planet's position in terms of the stellar radius.) ■

Figure 3.13a shows the two overlapping discs and the variables that we will use to calculate the eclipsed area, A_e, in the difficult case where the planet's disc covers part of the stellar limb. The origin is placed at the centre of the star's disc. Two angles, α_1 and α_2, are indicated. The distance $s(t)$ is parameterized in terms of the stellar radius,

$$s(t) = \xi R_*, \tag{3.17}$$

and the ratio of the radii of the two discs is defined to be equal to p, so we also have

$$R_\mathrm{P} = pR_*. \tag{3.18}$$

There are several shaded areas indicated in Figure 3.13b. The light blue area is a **sector** of the stellar disc, while the dark blue area is a sector of the planet's disc.

The green area is a large triangle formed by an extension of the stellar sector. The area indicated in purple is half of the area that we actually wish to calculate, while the area in red is an irregular shape formed by subtracting the stellar disc sector from the large triangle. The (desired) purple shape is the planetary disc sector with the irregular red shape subtracted. Using this geometric reasoning, we can write a pseudo-equation for the desired area A_e:

$$\begin{aligned} A_e &= 2 \times (\text{small dark blue sector} - \text{red shape}) \\ &= 2 \times (\text{small dark blue sector} \\ &\qquad - (\text{large green triangle} - \text{large light blue sector})) \\ &= 2 \times (\text{small dark blue sector} \\ &\qquad + \text{large light blue sector} - \text{large green triangle}). \end{aligned} \tag{3.19}$$

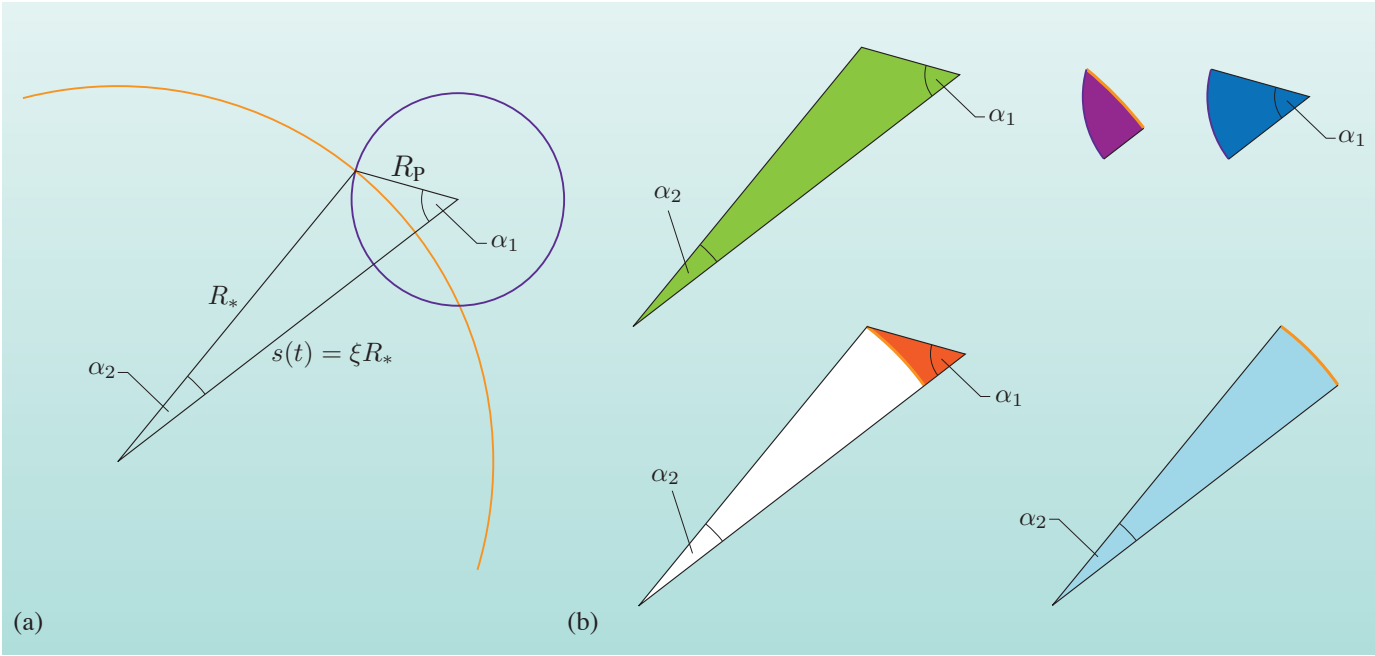

Figure 3.13 The partially overlapping discs of the star and the planet. The area of the stellar disc occulted by the planet during ingress and egress can be calculated using the geometry of the shapes shaded in colour here.

Each of the shapes on the right-hand side of Equation 3.19 has an area that is easy to express algebraically:

$$\text{area of small dark blue sector} = \pi R_P^2 \times \frac{\alpha_1}{2\pi} = \frac{p^2 R_*^2 \alpha_1}{2}, \tag{3.20}$$

$$\text{area of large light blue sector} = \pi R_*^2 \times \frac{\alpha_2}{2\pi} = \frac{R_*^2 \alpha_2}{2}, \tag{3.21}$$

where α_1 and α_2 are in radians. The area of the large green triangle can be written down using the sine rule:

$$\begin{aligned} \text{area of large green triangle} &= \frac{R_* \times \xi R_* \sin \alpha_2}{2} \\ &= \frac{\xi R_*^2}{2} \sin \alpha_2. \end{aligned} \tag{3.22}$$

The sine rule allows us to apply

$$\text{area} = \frac{\text{base} \times \text{perp. height}}{2},$$

where perp means perpendicular.

The cosine rule relates the cosine of an angle to the sides of an arbitrary triangle containing that angle.

The angles α_1 and α_2 can each be evaluated using the cosine rule, referring to the large green triangle in Figure 3.13:

$$\cos\alpha_1 = \frac{p^2 + \xi^2 - 1}{2\xi p}, \tag{3.23}$$

$$\cos\alpha_2 = \frac{1 + \xi^2 - p^2}{2\xi}. \tag{3.24}$$

Finally, by Pythagoras's theorem,

$$\sin\alpha_2 = \frac{\sqrt{4\xi^2 - (1 + \xi^2 - p^2)^2}}{2\xi}. \tag{3.25}$$

Using these results, we can express Equation 3.19 as

$$\begin{aligned} A_\mathrm{e} &= 2 \times \left(\frac{p^2 R_*^2 \alpha_1}{2} + \frac{R_*^2 \alpha_2}{2} - \frac{\xi R_*^2 \sqrt{4\xi^2 - (1 + \xi^2 - p^2)^2}}{4\xi} \right) \\ &= R_*^2 \left(p^2 \alpha_1 + \alpha_2 - \frac{\sqrt{4\xi^2 - (1 + \xi^2 - p^2)^2}}{2} \right). \end{aligned} \tag{3.26}$$

Equation 3.26 gives an expression for the eclipsed area of the star for the case where the planet intersects the limb of the star.

- What value does A_e have when the disc of the planet falls entirely within the disc of the star?
- In this case $A_\mathrm{e} = \pi R_\mathrm{P}^2$, or in the notation introduced above, $A_\mathrm{e} = \pi p^2 R_*^2$.
- What value does A_e have when the disc of the planet falls entirely outside the disc of the star?
- In this case none of the star is occulted and $A_\mathrm{e} = 0$.

To write down a general expression for A_e, we need to express the cases of the planet's disc falling entirely within and outside the stellar disc in terms of our variables ξ and p. We can use Equations 3.16 and 3.17 to express ξ in terms of time (or equivalently orbital phase), and thus obtain $A_\mathrm{e}(t)$. In Subsection 3.1.3 we argued that the time between first contact and second contact gives a measure of the size of a transiting planet, but we did not develop this idea mathematically. The work that we are doing here provides the relevant mathematics.

The planet falls entirely outside the stellar disc if the distance between the centres of the two discs exceeds the sum of their radii, i.e.

$$A_\mathrm{e} = \begin{cases} 0 & \text{if } R_* + R_\mathrm{P} < s, \\ 0 & \text{if } R_*(1 + p) < \xi R_*, \\ 0 & \text{if } 1 + p < \xi. \end{cases}$$

The planet falls entirely within the stellar disc if the distance between the centres of the two discs is less than the difference of their radii, i.e.

$$A_\mathrm{e} = \begin{cases} \pi R_\mathrm{P}^2 & \text{if } R_* - R_\mathrm{P} \geq s, \\ \pi p^2 R_*^2 & \text{if } R_*(1 - p) \geq \xi R_*, \\ \pi p^2 R_*^2 & \text{if } 1 - p \geq \xi. \end{cases}$$

Gathering this together, we have

$$A_{\rm e} = \begin{cases} 0 & \text{if } 1+p < \xi, \\ R_*^2\left(p^2\alpha_1 + \alpha_2 - \frac{\sqrt{4\xi^2-(1+\xi^2-p^2)^2}}{2}\right) & \text{if } 1-p < \xi \leq 1+p, \\ \pi p^2 R_*^2 & \text{if } 1-p \geq \xi. \end{cases} \quad (3.27)$$

The eclipsed area, $A_{\rm e}$, is thus a known function of R_*, p and $\xi(t)$, i.e. $A_{\rm e}(R_*, p, \xi)$, or equivalently of R_*, $R_{\rm P}$ and $s(t)$, i.e. $A_{\rm e}(R_*, R_{\rm P}, s)$.

Worked Example 3.1

A planet of radius $0.1\,{\rm R}_\odot$ transits across the face of a star like the Sun ($M = 1\,{\rm M}_\odot$, $R = 1\,{\rm R}_\odot$). The semi-major axis of the orbit is $20\,{\rm R}_\odot$, and the inclination of the orbit is $i = 89.5°$. At a time 2.0 h before the mid-transit, the planet is crossing the limb of the star.

(a) Determine the period of the planetary orbit, and hence the phase corresponding to the time mentioned above.

(b) Calculate the ratio of the radii of the planet and star, p, and the ratio of the distance of the planet from the centre of the star's disc to the radius of the star, ξ, at the time mentioned above.

(c) Verify that the planet is indeed just crossing the limb of the star at this time.

(d) Calculate the eclipsed area of the star at this time.

Solution

(a) Using Kepler's third law (Equation 1.1), the period of the planetary orbit is

$$\begin{aligned} P_{\rm orb} &= \left(\frac{4\pi^2 a^3}{GM_*}\right)^{1/2} \\ &= \left(\frac{4\pi^2 \times (20 \times 6.96 \times 10^8\,{\rm m})^3}{6.673 \times 10^{-11}\,{\rm N\,m^2\,kg^{-2}} \times 1.99 \times 10^{30}\,{\rm kg}}\right)^{1/2} \\ &= 8.95 \times 10^5\,{\rm s}. \end{aligned}$$

The orbital period is therefore 249 hours (or 10.4 days).

2.0 hours before the mid-transit corresponds to a phase offset of

$$\omega t = \frac{2\pi t}{P_{\rm orb}} = \frac{2\pi \times 2.0\,{\rm h}}{249\,{\rm h}} = 0.051 \text{ radians}.$$

(b) From Equation 3.18, the ratio of the radii of the planet and star is

$$p = \frac{R_{\rm P}}{R_*} = 0.1.$$

From Equation 3.17, the ratio of the distance of the planet from the centre of the star's disc to the radius of the star is $\xi = s(t)/R_*$. Now, the distance $s(t)$

To four significant figures, the phase offset is 0.050 52 radians, which one could argue for rounding to either 0.050 radians or 0.051 radians as we have done. $A_{\rm e}$ (calculated below) would be $0.0124\,{\rm R}_\odot^2$ rather than $0.0105\,{\rm R}_\odot^2$ if the phase offset were rounded to 0.050 radians.

is given by Equation 3.16 as

$$\begin{aligned} s(t) &= a\left(\sin^2\omega t + \cos^2 i\cos^2\omega t\right)^{1/2} \\ &= 20\,\mathrm{R}_\odot \times \left(\sin^2(0.051\,\mathrm{rad}) + \cos^2(89.5^\circ)\cos^2(0.051\,\mathrm{rad})\right)^{1/2} \\ &= 1.025\,\mathrm{R}_\odot. \end{aligned}$$

So

$$\xi = \frac{1.025\,\mathrm{R}_\odot}{\mathrm{R}_\odot} = 1.025.$$

(c) The condition for the planet to be crossing the limb of the star is that $1-p<\xi\leq 1+p$. Since $1-p=0.9$ and $1+p=1.1$, the condition is satisfied, and the planet is indeed crossing the limb of the star.

(d) To calculate the fractional area of the star that is occulted at this time, we need the second case from Equation 3.27. To solve that, we first need to calculate the angles α_1 and α_2, which may be obtained from Equations 3.23 and 3.24:

$$\begin{aligned} \alpha_1 &= \cos^{-1}\left(\frac{p^2+\xi^2-1}{2\xi p}\right) = \cos^{-1}\left(\frac{0.1^2+1.025^2-1}{2\times 0.1\times 1.025}\right) \\ &= 1.272 \text{ radians}, \\ \alpha_2 &= \cos^{-1}\left(\frac{1+\xi^2-p^2}{2\xi}\right) = \cos^{-1}\left(\frac{1+1.025^2-0.1^2}{2\times 1.025}\right) \\ &= 0.0957 \text{ radians}. \end{aligned}$$

So the eclipsed area is

$$\begin{aligned} A_\mathrm{e} &= R_*^2\left(p^2\alpha_1+\alpha_2-\frac{\sqrt{4\xi^2-(1+\xi^2-p^2)^2}}{2}\right) \\ &= \mathrm{R}_\odot^2\times\left((0.1^2\times 1.272)+0.0957-\frac{\sqrt{(4\times 1.025^2)-(1+1.025^2-0.1^2)^2}}{2}\right) \\ &= \mathrm{R}_\odot^2\times(0.0127+0.0957-0.0979) \\ &= 0.0105\,\mathrm{R}_\odot^2. \end{aligned}$$

The eclipsed area is therefore a fraction of the stellar disc equivalent to $(0.0105\,\mathrm{R}_\odot^2)/(\pi\,\mathrm{R}_\odot^2)=0.0033$ or 0.33%. For comparison, the fully eclipsed area would be 0.1^2 or 1% of the stellar disc.

3.2.3 The light lost from an axially symmetric stellar disc

We use the simpler variable name, r, to describe the normalized radial coordinate needed in some intricate algebra later.

The total flux, F, from the stellar disc is the intensity, I, integrated over the surface of the disc. We restrict our analysis to axially symmetric intensity distributions, so $I=I(r')$, where r' is measured from the centre of the stellar

disc, as shown in Figure 3.14. Thus

$$F = \int_{\text{disc}} I(r')\,\mathrm{d}A = \int_{r'=0}^{r'=R_*} I(r')\,2\pi r'\,\mathrm{d}r'. \tag{3.28}$$

Some authors use the *astrophysical flux*, which is the disc-averaged specific intensity, and differs from our definition by a factor of π.

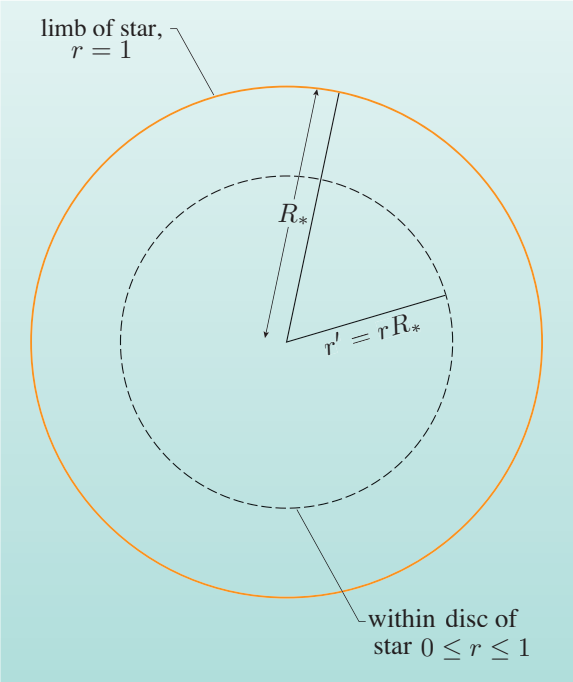

Figure 3.14 The variables r' and $r = r'/R_*$ used to describe locations on the axially symmetric stellar disc.

- Why have we imposed axial symmetry, rather than assuming a general case where I has an azimuthal-dependence as well as a radial-dependence?
- Our knowledge of the brightness distribution of the Sun and our analysis of stellar atmospheres suggest that the stellar disc will be limb darkened so that $I = I(r')$, but we have no reason to expect a substantial breaking of the axial symmetry.

For a stellar disc with uniform brightness, the intensity is a constant, $I = I_0$ across the entire disc, and Equation 3.28 simplifies:

$$\begin{aligned} F &= \int_{r'=0}^{r'=R_*} I_0\,2\pi r'\,\mathrm{d}r' \\ &= 2\pi I_0 \int_{r'=0}^{r'=R_*} r'\,\mathrm{d}r' \\ &= 2\pi I_0 \left[\frac{r'^2}{2}\right]_{r'=0}^{r'=R_*} \\ &= \pi I_0 R_*^2. \end{aligned} \tag{3.29}$$

We can now write down an expression for ΔF as a function of time. It is simply the missing flux that is occulted by the planet's disc, i.e.

$$\Delta F = \int_{\text{occulted area}} I(r')\,\mathrm{d}A, \tag{3.30}$$

where the integral is over the area of the stellar disc that is occulted by the planet. Performing this integral over the irregular purple shape shown in Figure 3.13 is not in general a straightforward step, but for the simplest case of a uniform stellar disc it is easy: Equation 3.30 becomes

$$\begin{aligned}\Delta F &= \int_{\text{occulted area}} I_0 \, \mathrm{d}A = I_0 \int_{\text{occulted area}} \mathrm{d}A \\ &= I_0 A_\mathrm{e}. \end{aligned} \tag{3.31}$$

With this rather trivial mathematical step, we have completed the work of deriving the analytic description of the shape of a transit light curve, albeit in the simplified case of a planet in a circular orbit transiting an idealized star whose disc has uniform brightness.

Exercise 3.6 (a) Derive an expression for the flux, $F(t)$, observed when a planet of radius R_P, with orbital inclination i, in a circular orbit of radius a and period P, transits an idealized star of radius R_* that has negligible limb darkening. You may use any of the relationships derived so far in this book. You do not need to exhaustively make the substitutions to produce a single expression relating $F(t)$ to i, a and P, but you need to define all the symbols used in your expression for $F(t)$ in terms of these parameters.

(b) Verify that for $p = 0.1$ and $\xi = 0.2$, your expression is consistent with

$$\frac{\Delta F}{F} = \frac{R_\mathrm{P}^2}{R_*^2}. \tag{Eqn 1.18}$$

(c) Assume that the orbital inclination has already been determined from the transit duration, T_dur, as we saw in Subsection 3.1.3, and that the period, P, is known. Use your expression to explain how R_P can be deduced in terms of R_* from measurements of the timings of first and second contacts, i.e. show that the ratio R_P/R_* can be determined by a method that is entirely independent of the transit depth.

(d) Parts (b) and (c) illustrate two independent methods for determining the ratio R_P/R_*. Discuss which of these two methods is most affected by the limb darkening that will occur in a real star. ■

The next bit is clever. Figure 3.14 shows a normalized axial coordinate, r, defined such that it is 0 at the centre of the stellar disc and 1 at the limb, i.e.

$$r = \frac{r'}{R_*}. \tag{3.32}$$

Figure 3.15 demonstrates how the eclipsed area is composed of a series of partial annuli, one of which is shown. It is clear from the figure that the eclipsed area within the circle of radius rR_* is exactly the area that would have been eclipsed were the star of radius rR_* rather than of radius R_*. (If this is not immediately apparent, imagine the figure without the circle of radius R_*, and compare it to Figure 3.13.) Therefore we can use the function that we have already derived for A_e with a change of variables to allow us to perform the integration over the axially symmetric stellar disc. If your abstract reasoning skills are not on form, you may wish to skip the following derivation and simply accept the result. The element of area shaded in red in Figure 3.15 is the additional eclipsed area $\mathrm{d}A(r)$

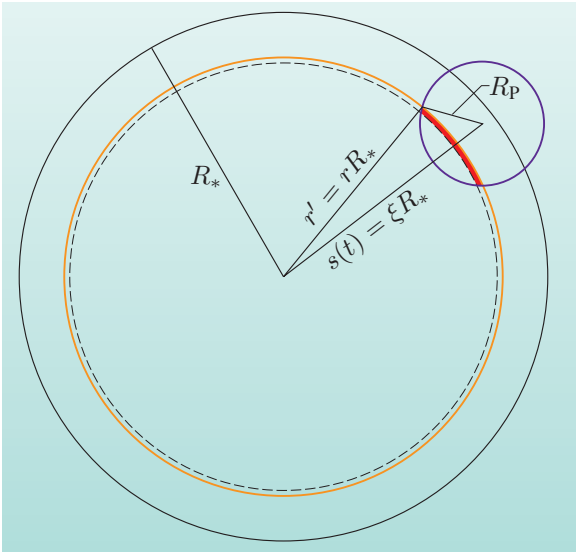

Figure 3.15 The eclipsed area $A_{\rm e}$ is made up of a series of partial annuli. Within each of these annuli $I(r)$ is constant, so ΔF for an axially symmetric brightness distribution can be evaluated by summing over these annuli.

when incrementing r by an amount $\mathrm{d}r$. But we have already observed that this element area is also the difference between $A_{\rm e}$ appropriate for a star of radius $R_* = r'$ and for a star of radius $R_* = r' + \mathrm{d}r'$. Consequently, the area shaded in red is

$$\mathrm{d}A(r) = \frac{\mathrm{d}A_{\rm e}}{\mathrm{d}r'}\,\mathrm{d}r'. \tag{3.33}$$

We can also write an expression for the eclipsed area for a star of radius rR_*, i.e. a star for which the annulus shown in Figure 3.15 marks its outer boundary. To do this we will need to change variables, as we used variables p and ξ, which were normalized with respect to the radius of the star. Since $p = R_{\rm P}/R_*$, changing the radius of the star to rR_* implies that p becomes p/r; similarly, ξ becomes ξ/r. Consequently, the eclipsed area of our hypothetical smaller star is given by

$$\begin{aligned}\int_0^{r'=rR_*} \frac{\mathrm{d}A_{\rm e}}{\mathrm{d}r'}\,\mathrm{d}r' &= A_{\rm e}\left(rR_*, \frac{p}{r}, \frac{\xi}{r}\right) \\ &= r^2 A_{\rm e}\left(R_*, \frac{p}{r}, \frac{\xi}{r}\right).\end{aligned} \tag{3.34}$$

Here we have made use of the fact that the only explicit occurrence of R_* in Equation 3.27 is a factor of R_*^2 multiplying the entire expression in each non-zero case. The implicit occurrences of R_* as the unit length in our normalized coordinate system have been taken care of by our change of variables. Comparing Equations 3.33 and 3.34, it is clear that the integrand on the left-hand side of the latter is equal to the right-hand side of the former. Consequently, we can say that

$$\mathrm{d}A(r) = \frac{\mathrm{d}}{\mathrm{d}r}\left[r^2 A_{\rm e}\left(R_*, \frac{p}{r}, \frac{\xi}{r}\right)\right]\mathrm{d}r. \tag{3.35}$$

This expression gives us exactly what we need to perform the integral required to obtain the flux deficit, ΔF, for a transit across a limb darkened star.

Going back to the completely general expression for ΔF, Equation 3.30, we can say for the limb darkened axially symmetric case that

$$\begin{aligned}\Delta F &= \int_{r=0}^{r=1} I(r)\,\mathrm{d}A(r) \\ &= \int_{r=0}^{r=1} I(r)\,\frac{\mathrm{d}}{\mathrm{d}r}\left[r^2 A_\mathrm{e}\left(R_*, \frac{p}{r}, \frac{\xi}{r}\right)\right]\mathrm{d}r. \end{aligned} \tag{3.36}$$

This expression therefore allows us to calculate the exact shape of a transit light curve by specifying the limb darkening law, the sizes of the star and planet, and the orbital parameters P, a and i.

3.3 The analytic underpinnings of transit light curve fitting

With ample justification it has been said that the core activity of physicists is fitting models to data. In constructing a mathematical model, the important factors governing a measured value must be identified and quantified, and equations specifying how these factors combine to produce the measurable quantity must be derived. In Section 3.2 we constructed a mathematical model to predict transit light curves. In Section 2.6 we briefly discussed the process of model-fitting; it is through this process that the parameters of transiting exoplanet systems are deduced from the observed data. In the next section we will describe how the observed transit light curve is used in conjunction with a mathematical model to determine the parameters of the observed system. Note that the model derived in Section 3.2 is not the *only* way to calculate a theoretical transit light curve for a limb darkened star. For example, a completely complementary approach is to build a model star–planet system in a computer simulation. In such a simulation, the star is covered with a grid and the geometry of the problem is used to calculate whether the light emitted by each grid point is visible to the observer at each orbital phase. This approach has strengths and weaknesses compared to employing the analytic model of Section 3.2; we focus on the latter purely because it *is* analytic, and therefore we can write down the salient equations and analyze it. In this section we show how the various features of the transit light curve allow the parameters of the star and planet to be deduced; in the next section we will illustrate how the model-fitting to accomplish this is performed.

3.3.1 The scale of the system

It is clear that the depth of the transit is dependent on the ratio of the two radii. As we explored in Exercise 3.6, there is a further constraint on the radius of the planet from the time between first contact and second contact, but this does not yield an absolute value for R_P either. In the exercise we used the timings to determine the ratio of the two radii. We could extend the work that we did in Exercise 3.6 to express the size of the orbit, a, in terms of either of the radii. Consequently, the transit alone does not determine the absolute size of anything in the system, it simply determines the ratios $R_\mathrm{P} : R_* : a$. As we noted in Section 3.1, once a transiting system has been identified, we have an excellent

method for determining the orbital period, P, and knowing this, Kepler's third law gives us an expression (Equation S2.2) for a that depends on the total mass, or to a good approximation on the stellar mass, M_*. Astronomers have been studying stars for centuries, and consequently we have reliable ways of estimating their properties based on their spectral type. Adopting values for M_* and P therefore fixes the scale of the system, by providing a value (in physical units) for a.

3.3.2 The radii and consistency checks

With a fixed in physical units, analysis of the transit light curve alone can yield values for R_P and R_*. The values obtained can be checked in several ways.

- First, the value for R_* should be consistent with the stellar properties deduced from the spectral type. Indeed, we discussed in Subsection 2.8.2 how a consistency check of this type is used in the SuperWASP candidate-winnowing process.
- As we mentioned above in Subsection 3.3.1, there are two independent estimates of R_P from two different properties of the transit light curve; if the solution is reliable, these should be mutually consistent.
- The transit duration should be consistent with the values of R_* and a, and the shape of the transit should be consistent with the impact parameter required to match the duration.

3.3.3 The orbital inclination and the mass of the planet

The transit duration coupled with the values of the two radii, R_P and R_*, yields a value for the orbital inclination through Equation 3.4. Values for M_* and P, coupled with a measurement of the stellar reflex orbital radial velocity, A_RV, give us $M_\mathrm{P} \sin i$, as we have seen in Equation S2.6. Using the orbital inclination from the transit duration thus allows the actual value of M_P to be determined. Thus for transiting planets we can obtain the mass, the radius and the orbital inclination.

3.4 Parameter determination from transit light curve fitting

Figure 3.16 shows the results of model-fitting to the data of Figure 2.4 folded on orbital phase. The stellar parameters were fixed at $M_* = 1.1\,\mathrm{M}_\odot$ and $R_* = 1.1\,\mathrm{R}_\odot$, and a linear limb darkening law of the form of Equation 3.9 was adopted. The best-fitting values for the planet HD 209458 b found with these assumptions are formally $R_\mathrm{P} = 1.27 \pm 0.02\,\mathrm{R_J}$ and $i = 87.1^\circ \pm 0.2^\circ$. We will discuss the uncertainty estimates from model fitting in the box 'Uncertainty estimates and ...' later in this section.

3.4.1 Light curve features to be matched

In Section 3.3 we examined analytically how the various parameters can be deduced from properties of the light curve. In the process of light curve

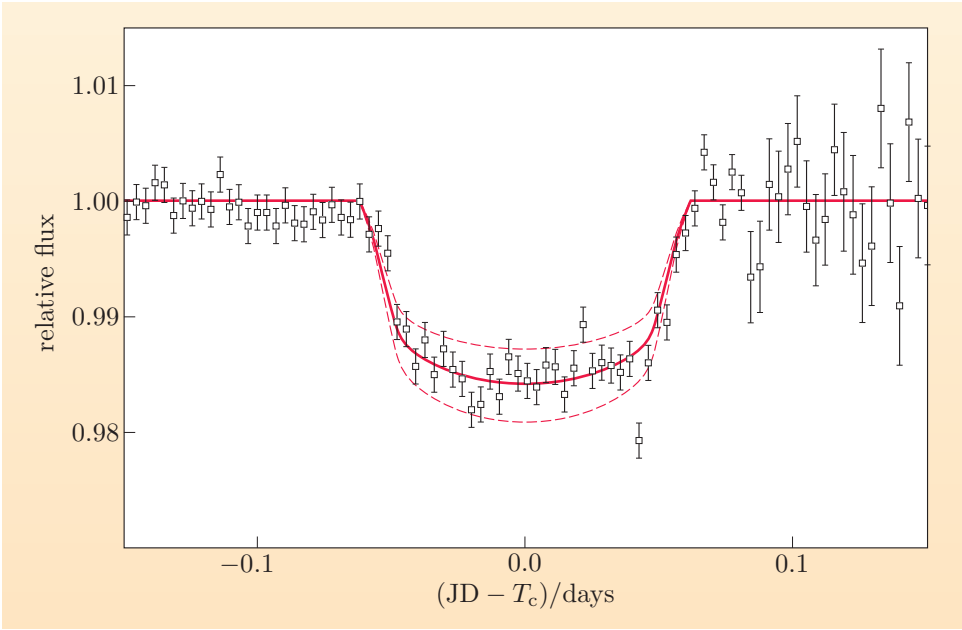

Figure 3.16 The first model fit to an observed exoplanet transit light curve. The observed data are for the transit of HD 209458 b shown in Figure 2.4, and the solid and dashed lines are model transit light curves calculated assuming $M_* = 1.1\,\mathrm{M}_\odot$ and $R_* = 1.1\,\mathrm{R}_\odot$. The solid line corresponds to $R_\mathrm{P} = 1.27\,\mathrm{R}_\mathrm{J}$ and $i = 87.1°$, and provides a good fit to the data, while the dashed lines show the models for planet radii 10% smaller and 10% larger.

modelling, these properties all simultaneously influence the parameters of the best fit. Consequently, instead of obtaining two independent values for R_P, one from the depth and another from the slope of ingress and egress, the best-fit parameter will be the one that produces the minimum χ^2 by producing the overall light curve that best matches the observations. Figure 3.17 schematically illustrates the features that can be measured in a high signal-to-noise ratio transit light curve. These are as follows.

- The depth, ΔF, measured from the out-of-transit level to the mid-transit level.
- The transit duration, which can be defined in various ways. The quantity T_dur that we derived in Equation 3.4 gives the time between first and fourth contacts. T_dur depends on R_P, R_* and i. An alternative definition, which is labelled T_tran in Figure 3.17, measures the time between mid-ingress and mid-egress, which is approximated by the time at which the centre of the planet's disc crosses the limb of the star, and is therefore independent of R_P.
- The duration of ingress, t_{1-2} (or equivalently egress, t_{3-4}). As we have already discussed, this depends on R_P. It also depends on the orbital inclination: for central transits, the motion of the planet is along the normal to the limb of the star, so the distance moved between first and second contact is a minimum. For larger impact parameters, the motion of the planet is at an angle to the normal to the stellar limb and therefore the distance travelled and the elapsed time between first and second contacts are longer.
- The curvature of the light curve between second and third contacts depends on the stellar limb darkening, and on the locus that the planet takes across the stellar disc, which in turn depends on R_* and i.

David Kipping's paper arXiv:1004.3819 discusses seven different definitions of transit duration.

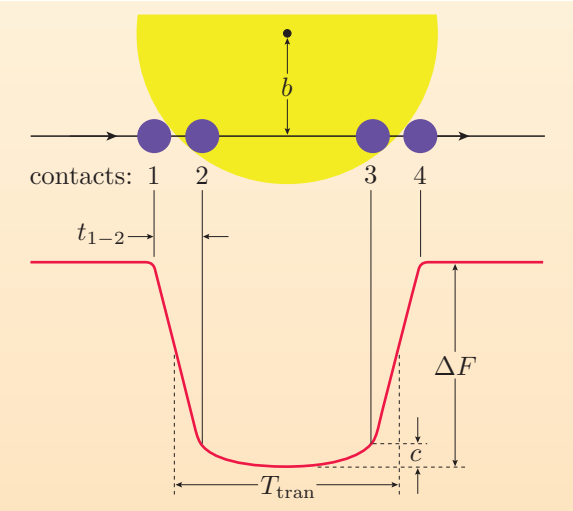

Figure 3.17 Properties of the transit light curve.

In the process of fitting a model light curve to observed data, the parameters of the model are adjusted until the model is able to fit the combination of all of these features.

Exercise 3.7 Sketch transit light curves for the following two contrasting transits:

(a) A large planet ($R_{\rm P}/R_* \sim 0.1$) with a large impact factor ($b \sim 0.9$) and very little limb darkening.

(b) A small planet ($R_{\rm P}/R_* \sim 0.01$) with a small impact factor ($b \sim 0$) and significant limb darkening. ■

3.4.2 The process of light curve fitting

The exact procedure used in fitting a model light curve can be varied as appropriate to the data in hand, to the prior knowledge about the system, and to the preferences, tools and expertise of the researcher. In general, a mathematical model (e.g. that described in Equation 3.36) is adopted, the stellar mass is assumed to fix the scale of the system, and a limb darkening law is adopted. In some cases, the limb darkening coefficient(s) may be fixed, but generally the limb darkening coefficients are treated as free parameters. For low signal-to-noise ratio data like those shown in Figure 3.16, the data cannot define the shape of the curvature of the transit floor to high precision, so the simplest limb darkening law, Equation 3.9, should be used. This linear limb darkening relationship has a single limb darkening coefficient, while the other relationships (Equations 3.10, 3.11 and 3.12) each have two coefficients, which must be either arbitrarily fixed or determined from the fit to the data. Following these initial assumptions, a process of χ^2 minimization is carried out to determine the parameters that produce the best fit to the observed transit light curve. In this way, best-fitting values of $R_{\rm P}$, R_* and i are determined.

Uncertainty estimates and parameter correlations in model-fitting

A complete discussion of this subject is well beyond the scope of this book, but we summarize some of the key points here. In Figure 3.16 the solid line shows the best-fit model, $\mu(t, \text{parameters})$, and we can see by eye that it does do a good job of matching the data points ($x_i \pm \sigma_i$). Examination of the figure reveals that the model does not pass through all of the $\pm 1\sigma$ uncertainty estimates for the individual data points. This is as expected. Figure 2.13 shows us that we expect roughly a third of all measurements to deviate by more than 1σ, so a well-fitting model should pass through roughly two-thirds of the $\pm 1\sigma$ error bars.

Assuming that the uncertainty estimates are correct and the measurement uncertainties are Gaussian distributed, there is therefore an 'expected' goodness of fit. This is characterized by the 'reduced chi-squared' or χ^2_N. For a fit of a model with n_p free parameters to n_x data points with Gaussian uncertainties,

$$\chi^2_N = \frac{\chi^2}{n_x - n_p}, \tag{3.37}$$

where χ^2 is given by

$$\chi^2 = \sum \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2. \tag{Eqn 2.36}$$

With the assumptions above, χ^2_N should have a value of

$$\chi^2_N = 1 \pm \sqrt{\frac{2}{n_x - n_p}} \tag{3.38}$$

for a model that fits the data.

The best-fit parameters correspond to the minimum value of χ^2, and we can use the variation of χ^2 as the parameters are adjusted to assess the uncertainty on the fitted values. If, for example, changes in the orbital inclination over an interval Δi make negligible difference to the value of χ^2, then the fitting procedure is unable to discriminate between values of i in this range. To quantify and generalize this statement, if varying one particular parameter over a defined interval while holding all the other parameters constant produces a change in χ^2_N of $\ll 1$, then the fit cannot discriminate meaningfully between parameter values within the interval. Consequently, again assuming that the uncertainty estimates are correct and the measurement uncertainties are Gaussian distributed, the procedure for estimating the uncertainty in a single fitted parameter is as follows.

1. Find the values of the model parameters that produce the minimum value of χ^2; call this value χ^2_{min}.
2. Check that $\chi^2_N \approx 1$ for these values of the parameters.
3. Hold all but the chosen parameter constant at their best-fitting values.
4. Vary the chosen parameter over a range around its best-fit value, and record the χ^2 corresponding to each input value.

5. Plot a graph of χ^2 versus the chosen parameter. The $\pm 1\sigma$ error estimate is the range within which

$$\chi^2 = \chi^2_{\rm min} + 1. \tag{3.39}$$

Sometimes the parameters are correlated, i.e. the data cannot tightly constrain an individual parameter, but can constrain a combination of two parameters. This can be examined by making a χ^2 contour map. Figure 3.18 shows an example: the contours represent the 1σ, 2σ and 3σ confidence intervals in the $(R_{\rm P}, i)$-plane. If the two parameters, $R_{\rm P}$ and i, were completely independent of each other in their influence on the χ^2 statistic, the intervals would be more-or-less circular. In Figure 3.18 we can see that the contours are skewed, with higher values of $R_{\rm P}$ corresponding to lower values of i; i.e. these two parameters are correlated.

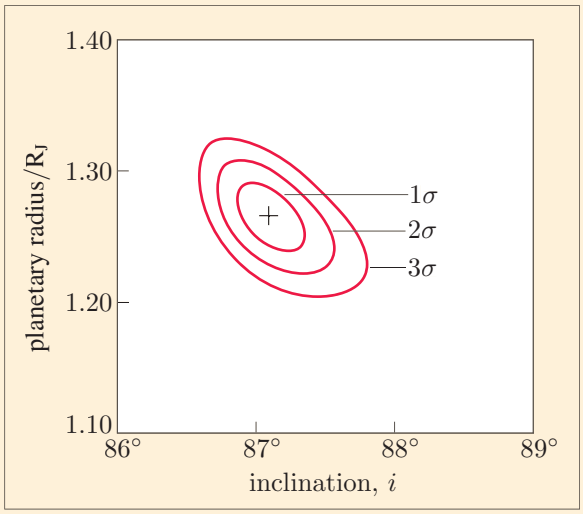

Figure 3.18 The 1σ, 2σ and 3σ confidence intervals in the $(R_{\rm P}, i)$-plane from the fits to the data shown in Figure 3.16.

The results for $R_{\rm P}$, R_* and i need to be critically examined. The most obvious check is to verify that R_* is consistent with the mass assumed for M_*. If it is not consistent, then fitting with a different limb darkening law, or changing the value of fixed limb darkening coefficients, or changing the assumed input stellar mass, or all of the above, should be explored. The box above describes how the formal errors in the deduced parameters may be estimated, assuming a Gaussian distribution in the measurement uncertainties. Realistic estimates of the uncertainties in the parameters of the fit should include the full range of possible outcomes with various justifiable input assumptions. Consequently, it is generally the case that the uncertainty in $R_{\rm P}$ is dominated by the uncertainty in the characteristics of the host star (i.e. M_*, R_* and limb darkening). Even when these factors are carefully considered, the true uncertainties on the parameter values are still likely to exceed these formal estimates, because the measurement uncertainties in the high signal-to-noise light curves are generally not Gaussian distributed; instead, they are dominated by systematic errors.

The source of these systematic errors undoubtedly includes effects that arise in the instrumentation: subtle changes in the state of the optics or detector can lead to changes in sensitivity. These can be at least partially characterized and removed. As more transiting systems are studied intensively and precise parameter determinations are made, it appears that some of the 'errors' may actually be caused by variability of the star itself.

Table 3.1 Selected results from transit light curve fitting: parameters for planets orbiting host stars that appear brightest from Earth.

Planet	$M_P \sin i$ (M_J)	R_P (R_J)	i (degrees)	Refs
HD 209458 b	$0.685^{+0.015}_{-0.014}$	$1.32^{+0.024}_{-0.025}$	86.677 ± 0.060	1, 2, 3
HD 189733 b	1.13 ± 0.03	1.138 ± 0.027	85.76 ± 0.29	4, 1, 5
HD 149026 b	$0.359^{+0.022}_{-0.021}$	$0.654^{+0.060}_{-0.045}$	$85.3^{+0.9}_{-0.8}$	1, 6
HD 17156 b	$3.212^{+0.069}_{-0.082}$	$1.023^{+0.079}_{-0.055}$	$86.2^{+2.1}_{-0.8}$	7
HAT-P-2 b	9.09 ± 0.24	$1.157^{+0.073}_{-0.092}$	$86.72^{+1.12}_{-0.87}$	8
HD 80606 b	3.94 ± 0.11	1.029 ± 0.017	89.285 ± 0.023	9, 10
WASP-18 b	10.43 ± 0.40	1.165 ± 0.077	86 ± 2.5	11, 12
WASP-7 b	$0.96^{+0.12}_{-0.18}$	$0.915^{+0.046}_{-0.040}$	$89.6^{+0.4}_{-0.9}$	13
HAT-P-11 b	0.081 ± 0.009	0.452 ± 0.020	88.5 ± 0.6	14, 15
WASP-14 b	$7.725^{+0.43}_{-0.67}$	$1.259^{+0.080}_{-0.058}$	$84.79^{+0.52}_{-0.67}$	16
XO-3 b	11.79 ± 0.59	1.217 ± 0.073	84.2 ± 0.54	17
HAT-P-8 b	$1.52^{+0.18}_{-0.16}$	$1.50^{+0.08}_{-0.06}$	$87.5^{+1.9}_{-0.9}$	18
HAT-P-1 b	0.524 ± 0.031	1.225 ± 0.059	86.28 ± 0.20	19
WASP-13 b	$0.46^{+0.06}_{-0.05}$	$1.21^{+0.14}_{-0.12}$	$86.9^{+1.6}_{-1.2}$	20

References for Table 3.1:

1: Torres et al., 2008, *ApJ*, **677**, 1324–42
2: Knutson et al., 2007, *ApJ*, **655**, 564–75
3: Southworth et al., 2007, *MNRAS*, **379**, L11–15
4: Boisse et al., 2009, *A&A*, **495**, 959–66
5: Winn et al., 2007, *AJ*, **133**, 1828–35
6: Nutzman et al., 2009, *ApJ*, **692**, 229–35
7: Winn et al., 2009, *ApJ*, **693**, 794–803
8: Pál et al., 2010, *MNRAS*, **401**, 2665–74
9: Pont et al., 2009, *A&A*, **502**, 695–703
10: Fossey et al., 2009, *MNRAS*, **396**, L16–20
11: Southworth et al., 2009, *ApJ*, **707**, 167–72
12: Hellier et al., 2009, *Nature*, **460**, 1098–100
13: Hellier et al., 2009, *ApJ*, **690**, L89–91
14: Dittmann et al., 2009, *ApJ*, **699**, L48–51
15: Bakos et al., 2010, *ApJ*, **710**, 1724–45
16: Joshi et al., 2008, *MNRAS*, **392**, 1532–8
17: Winn et al., 2008, *ApJ*, **683**, 1076–84

18: Latham et al., 2009, *ApJ*, **704**, 1107–19
19: Johnson et al., 2008, *ApJ*, **686**, 649–57
20: Skillen et al., 2009, *A&A*, **502**, 391–4

Table 3.1 summarizes selected published planet transit results. More exhaustive data can be found at various locations on the internet. Despite the rather pessimistic discussion above concerning the error estimates, it is truly amazing that we can measure the radii *of planets that we can't even see* to such precision. These results are one of the great triumphs of modern astrophysics.

Summary of Chapter 3

1. The four contact points of a transit occur when one of the limbs of the planet crosses one of the limbs of the host star.
2. The impact parameter, b, is the closest approach of the centre of the planet's disc to the centre of the star's disc, and is given by

$$b = a\cos i. \qquad \text{(Eqn 3.2)}$$

3. For a circular orbit, the duration of a transit viewed at orbital inclination i is

$$T_{\rm dur} = \frac{P}{\pi}\sin^{-1}\left(\frac{\sqrt{(R_* + R_{\rm P})^2 - a^2\cos^2 i}}{a}\right). \qquad \text{(Eqn 3.4)}$$

4. The optical depth at frequency ν is the path integral of the density, $\rho(s)$, multipled by the opacity, κ_ν:

$$\tau_\nu = \int_{\rm X}^{\infty} \rho(s)\,\kappa_\nu\,{\rm d}s. \qquad \text{(Eqn 3.6)}$$

   Radiation of intensity $I_{\rm emitted}$ is attenuated when passing through material of optical depth τ_ν:

$$\frac{I}{I_{\rm emitted}} = {\rm e}^{-\tau_\nu}. \qquad \text{(Eqn 3.7)}$$

5. Stellar discs appear limb darkened, with maximum intensity at the centre. Radiation travelling radially outwards has escaped, on average, from deeper, hotter layers than radiation travelling obliquely, so the centre of the disc is generally the most blue. Consequently, limb darkening is stronger at shorter wavelengths. Limb darkening depends on wavelength, structure of the star and detailed composition, so varies from star to star. It can be directly measured only for the Sun, and is calculated using computational models for other stars. The common limb darkening 'laws' adopted for the intensity emitted at an angle $\cos^{-1}\mu$ to the stellar radius vector are the linear law

$$\frac{I(\mu)}{I(1)} = 1 - u(1-\mu), \qquad \text{(Eqn 3.9)}$$

   the logarithmic law

$$\frac{I(\mu)}{I(1)} = 1 - u_l(1-\mu) - \nu_l\mu\ln\mu, \qquad \text{(Eqn 3.10)}$$

and the quadratic law

$$\frac{I(\mu)}{I(1)} = 1 - u_{\rm q}(1-\mu) - \nu_{\rm q}(1-\mu)^2. \qquad \text{(Eqn 3.11)}$$

6. Due to limb darkening, transit light curves are not generally flat-bottomed. Consequently,

$$\frac{\Delta F}{F} = \frac{R_{\rm P}^2}{R_*^2} \qquad \text{(Eqn 1.18)}$$

provides only an estimate of the flux deficit, ΔF, during transit. High signal-to-noise ratio light curves allow the measurement of the duration of ingress, t_{1-2}, and egress, t_{3-4}, and the curvature of the transit floor, as well as the duration, $T_{\rm dur}$, and the flux deficit at mid-transit, ΔF. Planet parameters are determined by model fits to the light curve, and the best fit should simultaneously reproduce all aspects of the observed transit.

7. The locus of the centre of the planet's disc relative to the centre of the star's disc is a small part of the foreshortened astrocentric orbit. The distance between the centres of the two discs is

$$s(t) = a\left(\sin^2\omega t + \cos^2 i\cos^2\omega t\right)^{1/2}, \qquad \text{(Eqn 3.16)}$$

where

$$\omega = \frac{2\pi}{P}$$

and $t = 0$ at inferior conjunction of the planet.

8. The eclipsed area, $A_{\rm e}$, of the stellar disc is given by

$$A_{\rm e} = \begin{cases} 0 & \text{if } 1+p < \xi, \\ R_*^2\left(p^2\alpha_1 + \alpha_2 - \dfrac{\sqrt{4\xi^2 - (1+\xi^2-p^2)^2}}{2}\right) & \text{if } 1-p < \xi \le 1+p, \\ \pi p^2 R_*^2 & \text{if } 1-p \ge \xi, \end{cases} \qquad \text{(Eqn 3.27)}$$

where

$$p = \frac{R_{\rm P}}{R_*}, \quad \cos\alpha_1 = \frac{p^2+\xi^2-1}{2\xi p}, \quad \cos\alpha_2 = \frac{1+\xi^2-p^2}{2\xi},$$
$$\xi = \frac{a}{R_*}\left(\sin^2\omega t + \cos^2 i\cos^2\omega t\right)^{1/2} \quad \text{and} \quad \omega = \frac{2\pi}{P_{\rm orb}}.$$

9. The flux from a star out of transit is given by

$$F = \int_0^{R_*} I(r')\,2\pi r'\,{\rm d}r'. \qquad \text{(Eqn 3.28)}$$

During transit, the flux drops by an amount

$$\Delta F = \int_{\text{occulted area}} I(r')\,{\rm d}A. \qquad \text{(Eqn 3.30)}$$

For a limb darkened star, this integral can be evaluated using the relationship

$$\Delta F = \int_{r=0}^{r=1} I(r)\,\frac{\rm d}{{\rm d}r}\left[r^2 A_{\rm e}\left(R_*, \frac{p}{r}, \frac{\xi}{r}\right)\right]{\rm d}r, \qquad \text{(Eqn 3.36)}$$

where r is the normalized axial coordinate, $r = r'/R_*$. This yields the exact shape of a transit light curve once the limb darkening law, the sizes of the star and planet, and the orbital parameters P, a and i are specified.

10. A transit light curve alone cannot determine the absolute size of anything, but it can determine the ratios $R_\mathrm{P} : R_* : a$. Kepler's third law combined with knowledge of $M_\mathrm{total} \approx M_*$ and P allows a to be calculated.
11. A transit light curve fit should be consistent with values of M_* and R_* inferred from the host star spectral type. The uncertainty in the precise host star characteristics can be the largest contributor to the uncertainties in the planet parameters M_P, R_P and i.
12. The best-fit parameters are determined by χ^2 minimization. The $\pm 1\sigma$ error estimate on any single parameter is the range within which

$$\chi^2 = \chi^2_\mathrm{min} + 1, \qquad \text{(Eqn 3.39)}$$

with all the other parameters held constant at their best-fit values. The χ^2 contour map for two parameters reveals whether the fit can independently constrain them. If the contours are skewed, the parameters are correlated: the fit cannot constrain them independently.