

A Mathematical Treatment of Solar Eclipse Formation

Arjun Tan*, Almuatasim Alomari and Marius Schamschula

*Department of Physics,
Alabama A & M University, Normal, AL 35762, U.S.A.*

** E-Mail: arjun.tan@aamu.edu*

Abstract

The solar eclipse is a natural phenomenon caused by the alignment of the Sun, Moon and the Earth in space and dictated by the law of rectilinear propagation of light. It occurs during a new moon, when the Sun is partially or fully blocked by the Moon as observed from the Earth. Qualitative discussions on the solar eclipse are readily found in the literature. In this paper, an in-depth discussion of total and annular solar eclipse formation is given. Also a mathematical exercise on the partial solar eclipse is carried out.

1. INTRODUCTION

The solar eclipse is a rare but spectacular event as observed from the Earth. It stirred the imagination of many ancient cultures, which attributed it to a divine act of some sorts. Yet, it is a purely natural phenomenon caused by the alignment of the Sun, Moon and the Earth in space and dictated by the *law of rectilinear propagation of light* (e.g., [1]). A solar eclipse occurs during a new moon, when the Sun is partially or fully blocked by the Moon as observed from the Earth. If the Sun is fully blocked, we have a *total solar eclipse*; whereas, if the Sun is partially blocked, a *partial solar eclipse* or an *annular solar eclipse* can take place (e.g., Fig. 1, from [2]). Qualitative discussions on the solar eclipse can be found in elementary textbooks in Physical Science (e.g., [3, 4]). In this paper, we give an in-depth discussion of solar eclipse formation from a purely geometrical point of view. We then carry out a quantitative analysis of solar eclipse types.

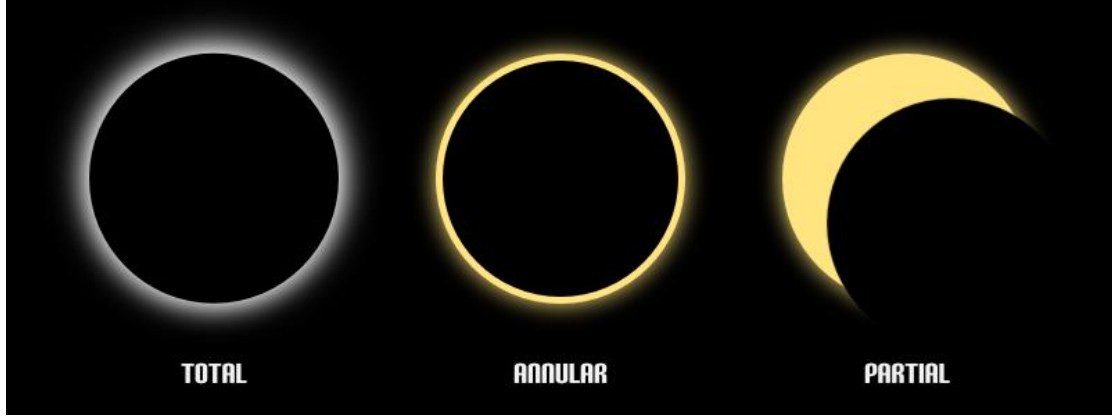


Fig. 1. Images of total, annular and partial solar eclipse (from [2]).

2. PHYSICAL PARAMETERS DETERMINING SOLAR ECLIPSE FORMATION

The solar eclipse involves the three heavenly bodies of the Sun, Moon and the Earth in space. The objects are held by Keplerian motion in two levels. First, the Earth, like any other planet, revolves around the Sun in an elliptical orbit. Second, the Moon revolves around the Earth in a similar Keplerian orbit. Solar eclipse occurs during a new moon when the Sun, Moon and the Earth are perfectly aligned in a straight line. However, this does not happen on every new moon because the orbital planes of the Moon and the Earth are tilted at slightly over 5° in space.

The formation of a solar eclipse is determined by just four physical quantities: (1) The diameter of the Sun D_S ; (2) The diameter of the Moon D_M ; (3) The distance of the Earth from the Sun d_E ; and (4) The distance of the Moon from the Earth d_M . Since the Sun and the Moon both rotate about their respective axes relatively slowly, no flattening of their figures are noticeable, and D_S and D_M are regarded as constants. The distances d_E and d_M are however variable. d_E is shortest at *perihelion* (the nearest approach of the Earth to the Sun) and longest at *aphelion* (the farthest retreat of the Earth from the Sun). Likewise, d_M is shortest at *perigee* (the nearest approach of the Moon to the Earth) and longest at *apogee* (the farthest retreat of the Moon from the Earth). The values of these parameters are taken from references [5] and [6] and entered in Table I.

Also shown in Table I are the angular diameters of the Sun (θ_S) and the Moon (θ_M) as viewed from the Earth. These are derived quantities as defined by $\theta_S = D_S/d_E$; and $\theta_M = D_M/d_M$. The maximum and minimum values of these quantities, as well as their averages, are calculated and entered in Table I. It is an interesting coincidence that the average values of θ_S and θ_M are almost identical. Also the variations of θ_S and θ_M mean that on any occasion the lunar disk may be larger or smaller than the solar disk. When θ_M is greater than θ_S , a total solar eclipse can occur, but not an annular solar eclipse. Likewise, when θ_M is smaller than θ_S , an annular eclipse can occur, but not a total eclipse. Thus the occurrences of total solar eclipse and annular solar eclipse are mutually exclusive events.

Table I. Geometrical Parameters relevant to Solar Eclipse Formation

Symbol	Definition	Value	Value, 10^6 km
D_S	Mean Diameter of Sun	1.392×10^6 km	1.392
D_M	Mean Diameter of Moon	3.475×10^3 km	.003475
d_E	Mean Distance of Earth from Sun	149.6×10^6 km	149.6
$d_{E_{max}}$	Maximum Distance of Earth from Sun (Distance at Earth's Aphelion)	152.1×10^6 km	152.1
$d_{E_{min}}$	Minimum Distance of Earth from Sun (Distance at Earth's Perihelion)	147.1×10^6 km	147.1
d_M	Mean Distance of Moon from Earth	384.4×10^3 km	.3844
$d_{M_{max}}$	Maximum Distance of Moon from Earth (Distance at Moon's Apogee)	405.7×10^3 km	.4057
$d_{M_{min}}$	Minimum Distance of Moon from Earth (Distance at Moon's Perigee)	363.1×10^3 km	.3631
Θ_S	Mean Angular Diameter of Solar Disk at Earth	.533°	
$\Theta_{S_{max}}$	Largest Angular Diameter of Solar Disk at Earth (Angular Diameter at Earth's Perihelion)	.542°	
$\Theta_{S_{min}}$	Smallest Angular Diameter of Solar Disk at Earth (Angular Diameter at Earth's Aphelion)	.524°	
Θ_M	Mean Angular Diameter of Lunar Disk at Earth	.518°	
$\Theta_{M_{max}}$	Largest Angular Diameter of Lunar Disk at Earth (Angular Diameter at Moon's Perigee)	.548°	
$\Theta_{M_{min}}$	Smallest Angular Diameter of Lunar Disk at Earth (Angular Diameter at Moon's Apogee)	.491°	

3. GEOMETRY OF SOLAR ECLIPSE FORMATION

Figure 2 (not to scale) depicts the geometry of a solar eclipse formation. The known distances d_E and d_M are shown in the figure. The unknown distance x is the length of the *umbra*, which is the dark shadow cast by the Moon in space. If the umbral cone converges before reaching the Earth to diverge again (as shown in Fig. 2), we have a *antumbra* [1]. The sub-umbral point is the location on the Earth's surface beneath the umbral cone. At that point, a perfectly annular or total solar eclipse occurs depending upon whether or not the umbral cone converges before reaching the Earth's surface.

The distance x can be determined from Fig. 2. From the umbral triangle in the figure and a similar larger triangle produced by extending the umbral triangle backwards to the Sun, we have

$$\frac{D_M}{D_S} = \frac{x}{d_E - d_M + x} \quad (1)$$

Solving for x , we get

$$x = \frac{D_M}{D_S - D_M} (d_E - d_M) \quad (2)$$

If $x < d_M$ (as in Fig. 2), then an annular solar eclipse is formed at the sub-umbral point. On the other hand, if $x > d_M$, then a total solar eclipse is formed there. Obviously, in the former case, $\Theta_S > \Theta_M$, whereas in the latter case, $\Theta_S < \Theta_M$.

If we substitute the average values of d_E and d_M (from Table I) in Eq. (2), we get x

$=.3734 \times 10^6$ km, which is smaller than the average value of d_M ($.3844 \times 10^6$ km). Thus, on a normal occasion, an annular solar eclipse tends to form, rather than a total solar eclipse. One can state by inference that annular solar eclipses are more common occurrences than total solar eclipses. This is actually corroborated by facts: – the observed frequency ratio of annular to total solar eclipse is about 60:40 [1].

Two extreme cases of solar eclipse formation can be analyzed using Eq. 2. First, the longest umbra is formed when $d_E - d_M$ is maximum, i.e., when d_E is maximum and d_M is minimum, or, when the Earth is at aphelion and the Moon is at perigee. On substituting the values of $d_{E_{max}}$ and $d_{M_{min}}$ from Table I, we get $x = .3797 \times 10^6$ km. This is greater than $d_{M_{min}} = .3631 \times 10^6$ km. Thus, a total solar eclipse rather than an annular solar eclipse can occur. Next, the shortest umbra is formed when $d_E - d_M$ is minimum, i.e., when d_E is minimum and d_M is maximum, or, when the Earth is at perihelion and the Moon is at apogee. In this case, substitution of $d_{E_{min}}$ and $d_{M_{max}}$ from Table I gives $x = .3672 \times 10^6$ km. This is smaller than $d_{M_{max}} = .4057 \times 10^6$ km. Thus, an annular solar eclipse rather than a total solar eclipse may occur on this occasion.

In the latter case, the area of the annular eclipse will be the greatest. The inner and outer angular diameters of the annuli will be those of the lunar and solar disks respectively: $\Theta_M = D_M/d_{M_{max}} = .4908^\circ$; and $\Theta_S = D_S/d_{E_{min}} = .5422^\circ$. The ratio of the inner and outer diameters will be $\Theta_M/\Theta_S = .9012$ or 90.12%. The ratio of the areas of the inner (dark) and outer (bright) disks will be $A_{inner}/A_{outer} = \Theta_M^2/\Theta_S^2 = .8193$ or 81.93%. In other words, the area of the annulus will be 18.07% of that of the (bright) outer solar disk.

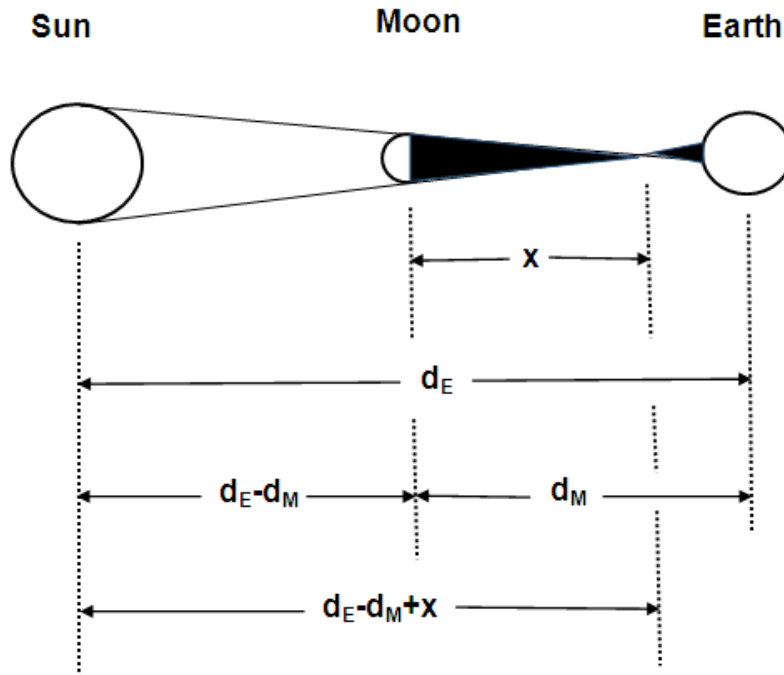


Fig. 2. Geometry of solar eclipse formation.

4. AN EXERCISE ON PARTIAL SOLAR ECLIPSE

Because of the relatively rapid passage of the lunar umbra on the Earth's surface, the duration of a total or annular eclipse seldom exceeds a few minutes [1]. Further, a total or annular eclipse is always preceded by a progressing partial eclipse and followed by a regressing partial eclipse [1]. We now discuss the following mathematical exercise: *Given solar and lunar disks of the same angular diameter, find the separation between the two disks such that a partial solar eclipse of a certain area is formed.* Let α be the fraction of the solar disk eclipsed. Then the area of the visible disk is $(1 - \alpha)\pi a^2$, where a is the radius of either disk.

Figure 3 is a schematic diagram of a partial solar eclipse formation where the bright solar disk on the left is partially eclipsed by the dark lunar disk on the right. The crescent-shaped part of the solar disk on the left is seen as the partial solar eclipse. In Fig. 3, O is the centre of the solar disk; AB the common chord of the two intersecting disks; D is the mid-point of AB; θ is the angle AOD; $p = \overline{AD}$; and $x = \overline{OD}$ is the half distance between the centres of the two disks to be determined. We have, $\cos\theta = x/a$; and $p = \sqrt{a^2 - x^2}$.

Now

$$\text{Area of segment } ABC = \text{Area of sector } OACB - \text{Area of triangle } AOB \quad (3)$$

where

$$\text{Area of segment } ABC = \frac{\alpha}{2}\pi a^2 \quad (4)$$

$$\text{Area of sector } OACB = a^2 \cos^{-1}\left(\frac{x}{a}\right) \quad (5)$$

and

$$\text{Area of triangle } AOB = x\sqrt{a^2 - x^2} \quad (6)$$

Substituting Eqs. (4) – (6) in (3), we have

$$a^2 \cos^{-1}\left(\frac{x}{a}\right) = x\sqrt{a^2 - x^2} + \frac{\alpha}{2}\pi a^2 \quad (7)$$

Eq. (7) is a *transcendental equation* of the form

$$f\left(\frac{x}{a}\right) = g\left(\frac{x}{a}\right) \quad (8)$$

where

$$f\left(\frac{x}{a}\right) = a^2 \cos^{-1}\left(\frac{x}{a}\right) \quad (9)$$

and

$$g\left(\frac{x}{a}\right) = a^2 \left(\frac{x}{a}\right) \sqrt{1 - \left(\frac{x}{a}\right)^2} + \frac{\alpha}{2}\pi a^2 \quad (10)$$

Eq. (7) is customarily solved by the *graphical method*, where $f(x/a)$ and $g(x/a)$ are plotted on the same graph and the intersection point gives the value of (x/a) .

Figure 4 illustrates this method of solution for three fractions of the solar disk eclipsed: $\alpha = 1/4, 1/2$ and $3/4$. $f(x/a)$ is independent of α and has a solitary plot in the graph. $g(x/a)$ is plotted for the three parametric values of α . The intersection points of $f(x/a)$ and $g(x/a)$ mark the values: $x = .6347a$; $.404a$; and $.1975a$ for $\alpha = 1/4, 1/2$ and $3/4$, respectively. The required distance between the two disks are then $2x = 1.269a$; $.808a$; and $.395a$ for a quarter, a half and three-quarters of partial solar eclipse, respectively. Stated separately, we have: (1) A quarter of the solar disk is eclipsed when the distance between the solar and lunar disks is 1.269 times the radius of either disk; (2)

Half of the solar disk is eclipsed when the distance between the disks is .808 times the radius of the disks; and (3) Three-quarters of the solar disk is eclipsed when the distance between the disks is .395 times the radius of either disk.

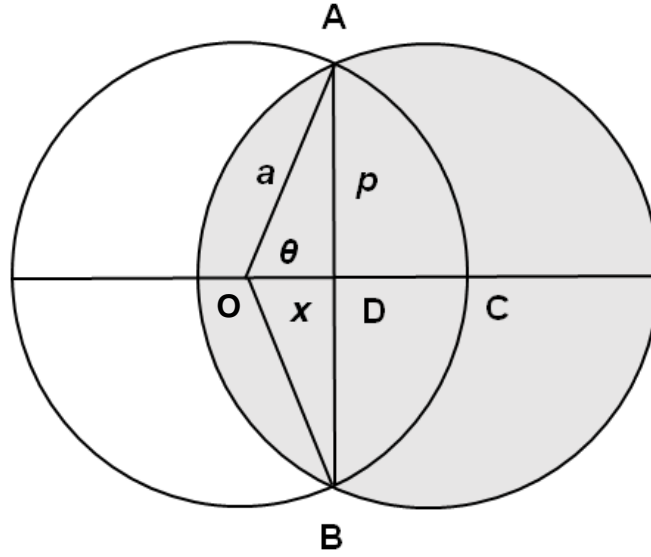


Fig. 3. Schematic diagram of partial solar eclipse.

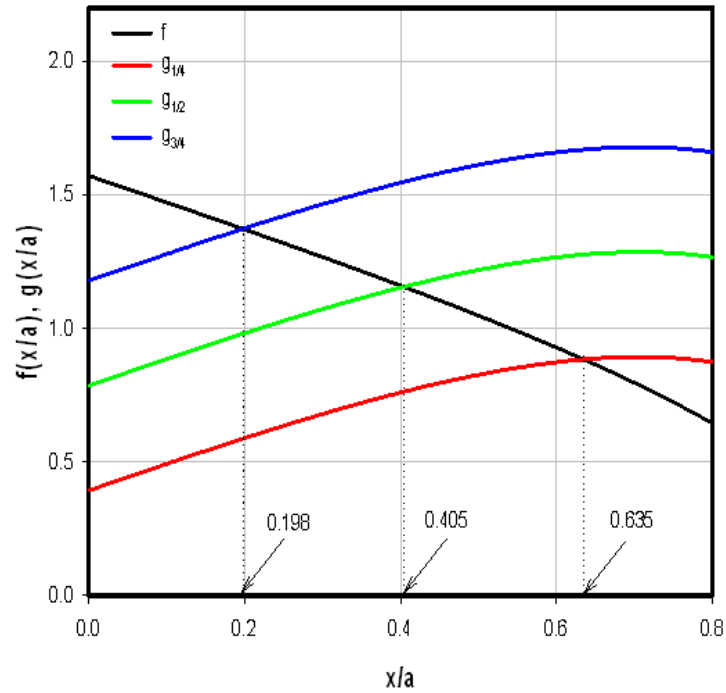


Fig. 4. Graphical solution of transcendental equation (7) for fractions of solar disk eclipsed.

5. SUMMARY AND DISCUSSION

The solar eclipse is a spectacular natural event observed on earth which is entirely explained the law of rectilinear propagation of light, shadow formation and elementary mathematics. The occurrences of total and annular solar eclipses depend on the accidental coincidences of the diameters of the Sun and the Moon and the distances of the Sun and the Moon from the Earth. It is well-known that due to tidal forces, the Moon has been and continues to move away from the earth at the rate of 3.8 cm per year [7]. Consequently, the umbral cone of the Moon is also moving away from the Earth at the same time. It can safely be inferred that one day in the distant future, the umbral cone will always terminate before reaching the Earth when a total solar eclipse will no longer be visible from our planet.

REFERENCES

- [1] https://en.Wikipedia.org/wiki/solar_eclipse.
- [2] <http://space-facts.com/solar-eclipse/>
- [3] J.T. Shipman & J.W. Wilson, *An Introduction to Physical Science* (Heath & Co., 1990).
- [4] K.B. Krauskopf & A. Beiser, *The Physical Universe* (McGraw-Hill, 2003).
- [5] *The New Solar System*, J.K. Beatty, C.C. Peterson & A. Chaikin, ed.(Cambridge, 1999).
- [6] J.R. Lang, *The Cambridge Guide to the Solar System* (Cambridge, 2003).
- [7] https://en.wikipedia.org/wiki/Orbit_of_the_Moon.

