1

Physics from Archimedes to Rutherford

1.1

Mathematics and Physics in Antiquity

In the 3rd millennium BC, the descendants of the Sumerians in Mesopotamia already used a sexagesimal (base 60) place-value numeral system of symbols to represent numbers, in which the position of a symbol determines its value. They also knew how to solve rhetorical algebraic equations (written in words, rather than symbols). By the 2nd millennium BC, Mesopotamians had compiled tables of numbers representing the length of the sides of a right triangle.\(^1\)

Their astronomers catalogued the motion of the stars and planets, as well as the occurrence of lunar eclipses. The day was divided into 24 hours, each hour into 60 minutes, each minute into 60 seconds.

A decisive step in the development of mathematics was made by the Egyptians, who introduced special signs (hieratic numerals) for the numbers 1 to 9 and multiples of powers of 10. They devised methods of multiplication and division which only involved addition. However, their decimal number system did not include a zero symbol, nor did it use the principle of place value.

The roots of modern science grew in the city-states of Ancient Greece scattered across the Eastern Mediterranean. The Greeks established physics as a science, greatly advanced astronomy, and made seminal contributions to mathematics, including the idea of formal mathematical proof, the basic rules of geometry, discoveries in number theory, and the rudiments of symbolic algebra and calculus.

A remarkable testament to their ingenuity is the Antikythera mechanism, the oldest known complex scientific calculator, which was discovered in a wreck off the Greek island of Antikythera and is dated to about 100 BC (see Fig. 1.1). The mechanism, whose complexity is comparable to that of 18th century clocks, could calculate the positions of celestial bodies [1].

---

1) A general theorem stating the relationship of those lengths is attributed to Pythagoras of Samos (c. 570–490 BC). He is also credited with discovering that the harmonic intervals correspond to unique whole number proportions. A special case of Pythagoras’ theorem is given in Baudhayana Sulba-sutra (India, c. 800 BC).
EUCLID of Alexandria (325–265 BC) is credited with establishing *Euclidean geometry*, a mathematical system used by Greek mathematicians and astronomers to describe the heavens. Euclid’s *Elements* is the earliest known systematic and rigorous exposition of geometry — "the first grandiose building of mathematical architecture" [2]. Euclid presented mathematical concepts in logical order, starting with the most basic of assumptions, and using them to form a series of propositions and conclusions of increasing complexity. His work is as relevant today as it was in ancient times! However, there is a small ambiguity in the foundation of the system (the *fifth postulate*). This was explored by mathematicians in the 19th century, particularly by Bernhard Riemann (1826–1866) who laid the foundations of a new form of geometry (*non-Euclidean*) that provides the mathematical basis for *general relativity*.

ARCHIMEDES of Syracuse (287–212 BC) is the greatest mathematician and physicist of antiquity. By resorting to both mathematics and empirical evidence to determine principles behind phenomena, he established physics as a science. Archimedes used a technique familiar from the *differential calculus* to construct a tangent to any given point of a curve, and applied the *method of exhaustion* — originally introduced by Eudoxus of Cnidus (408–355 BC) — to calculate an area by approximating it by the areas of a sequence of polygons, thus anticipating the *integral calculus* [3]. He showed that the surface of a sphere is four times that of the largest circle it contains, and that the volume of a sphere is two-thirds the volume of a circumscribed cylinder. Archimedes also devised a method of calculating the value of \(\pi\) to any desired degree of accuracy. On a recently deciphered palimpsest, Archimedes proposes that two *infinite sets* have the same size because the elements in them can be put in a one-to-one correspondence [5] (such sets are said to have the same *cardinality*).
Archimedes formulated some of the fundamental laws of mechanics (e.g., the law of the lever), and laid the foundations of hydrostatics. Archimedes’ law in hydrostatics, which is described in his treatise On Floating Bodies, states that a body immersed in a fluid experiences a buoyant force equal to the weight of the fluid it displaces. Since water is practically incompressible, an immersed body would displace an amount of water equal to its own volume. By dividing the mass of the body by the volume of water displaced, the density of the body could be determined.

Aristarchus of Samos (310–230 BC) proposed the heliocentric model of the solar system described in Archimedes’ book The Sandrecker [3] and in Plutarch's treatise Face on the Moon. Aristarchus used the following four observations to determine the diameters of the Sun and Moon and their distances from the Earth [4]: (1) When the Moon is half-illuminated by the Sun, the angle between the lines of sight from the Earth to the Moon and the Sun is 87° (the actual value is 89.853°); (2) The Moon just covers the visible disk of the Sun during a solar eclipse; (3) The shadow of the Earth at the position of the Moon during a lunar eclipse is just wide enough to fit a sphere with twice the diameter of the Moon (this was presumably found by timing a lunar eclipse); and (4) The Moon subtends an angle of 2° (the actual value is 0.519°). By taking these observations as postulates, Aristarchus deduced in turn that: (1) The Earth-Sun distance is about 19 times larger than the Earth-Moon distance; (2) The diameter of the Sun is approximately 20 times larger than the diameter of the Moon; (3) The diameter of the Earth is about 2.8 times larger than that of the Moon; and (4) The distance from the Earth to the Moon is between 22.5 and 30 times larger than the diameter of the Moon. Aristarchus reasoned that, since the Sun is much larger than the Earth, our planet most probably orbited the Sun. It was suggested by Heraclides of Pontus (387–312 BC) that the Earth must rotate on its axis in order to explain the apparent motion of the stars. In 1543, Nicolaus Copernicus (1473–1543) expounded basically the same ideas in his epoch-making treatise De Revolutionibus Orbium Celestium.

Eratosthenes of Cyrene (276–197 BC) determined the size of the Earth by assuming that the Sun is so far away that its rays are essentially parallel. He knew that at noon on June 21 (the summer solstice), the Sun would illuminate the bottom of a deep vertical water well in the city of Syene (now Aswan on the Nile in Egypt), which means that the Sun was then at the zenith in the sky. At the same time, he measured the length of a shadow cast by an obelisk in Alexandria, located north of Syene. He used this length and the known height of the obelisk to infer that the direction to the Sun at Alexandria differed from that at Syene by an angle of seven degrees. In other words, the known distance between Syene and Alexandria was 7/360 of the circumference of the Earth. Expressed in modern units of length, Eratosthenes estimated that the Earth had a radius of 6100 km, amazingly close to the actual mean value of 6371 km.
HIPPARCUS of Nicaea (190–120 BC) is the greatest of the ancient Greek astronomers. He made important contributions to trigonometry, discovered the precession of the equinoxes, calculated the length of the solar year and catalogued the positions of almost one thousand stars to a precision of one degree. Hipparchus classified the visible stars in six groups, called magnitudes, according to their apparent brightness: he divided a fixed time interval after sunset in six equal periods, and assigned the same magnitude to the stars that became visible to the naked eye during each period.

DIOPHANTUS of Alexandria (3rd century AD), "the father of algebra", is best known for his Arithmetica, a series of books on algebraic equations and the theory of numbers (the Greek word for number is arithmos). Diophantus was the first Greek mathematician to recognize fractions as numbers. He also took a fundamental step from rhetorical algebra to modern symbolic algebra, wherein symbols are used for the unknown and for algebraic powers and operations.

The civilization of Ancient Greece spread across the African and Asian territories of the empire created by Alexander the Great (356–323 BC). The Greek scientific ideas, adopted and enriched by Indian and Islamic mathematicians, were reintroduced into Europe by Arab scholars and, in particular, through the work of Leonardo Fibonacci of Pisa (1175–1250). Arguably the single most important contribution of post-Hellenic scholars to modern mathematics is succinctly described by Pierre-Simon Laplace (1749–1827):

"The ingenious method of expressing every possible number using a set of ten symbols (each symbol having a place value and an absolute value) emerged in India. The idea seems so simple nowadays that its significance and profound importance is no longer appreciated. Its simplicity lies in the way it facilitated calculation and placed arithmetic foremost among useful inventions" [6].

A textbook by Indian mathematician Brahmagupta (598–668) is considered the earliest work to treat zero as a number in its own right. Brahmagupta is also credited for extending arithmetic to negative numbers. Persian mathematician Muhammad al-Khwarizmi (780–850) was principally responsible for the spread of the Hindu-Arabic numeral system throughout the Middle East and Europe. The word algebra is derived from the name of one of the basic operations with equations (al-jabr) described in a book he wrote around 820 AD. Omar Khayyam (1048–1131), Persian astronomer, mathematician and poet, contributed to calendar reform and discovered a geometrical method of solving cubic equations by intersecting a parabola with a circle.

2) The three essential features of this number system — (1) nine signs and the concept of zero, (2) a place-value system, and (3) a decimal base — may have already been known to Chinese mathematicians. The Maya of Central America also discovered the zero digit, but used a vigesimal (base 20) number system [2].
1.2 Renaissance under Galileo, Kepler and Descartes

The age of modern physics was inaugurated by Galileo Galilei (1564–1642) through his use of geometry in the study of motion. Galileo discovered that:

- Free objects propagate in straight lines at constant speed (Newton’s first law).
- The trajectory of an object in a gravitational field does not depend on its mass.

Galileo’s Discourses Regarding Two New Sciences [7], which laid the foundations of statics and kinematics, is his ultimate scientific ‘testament’. The book describes Galileo’s discoveries concerning uniformly accelerated systems, free fall under gravity, the motion of pendulums, as well as the law of inertia.

By letting balls of different weights roll down a slope (see Fig. 1.2) — thereby slowing down their fall so that it could be timed — Galileo showed that the balls experienced uniform acceleration [8]. For such a motion he established the time-squared law \( d = \frac{1}{2} at^2 \), where \( d \) is the distance travelled and \( a \) is the acceleration. Galileo used a water clock to measure time intervals quite accurately, and found that the same law holds for freely falling objects, which implies that they experience a constant acceleration \( g \) due to gravity. He separated the motion of a projectile into a vertical component \( y = gt^2/2 \) and a horizontal component \( x = v_x t \), where \( v_x \) is the constant velocity along the \( x \)-axis, and found that \( y = \text{const} \cdot x^2 \), i.e., that its trajectory is a parabola [9].

In his Dialogue Concerning the Two Chief World Systems: Ptolemaic & Copernican, Galileo asserts that the uniform motion of one inertial frame (the coordinate system in which Newton’s first law, the law of inertia, is valid) relative to another cannot be detected by purely mechanical tests performed in each frame (Galilean relativity). For such reference frames the Galilean coordinate transformation \( x' = x - v_x t \) holds in the realm of Newtonian physics.

In 1608 Galileo built the first telescopes that could be used for astronomical observations (see Fig. 1.3). From his observations of the movement of sunspots, Galileo deduced that the Sun rotates around its axis. In his book Starry Messenger, published in 1610, Galileo claims to have observed that our galaxy, the Milky Way, is composed of myriads of stars. He also discovered Jupiter’s satellites and the phases of the planet Venus. This last discovery was interpreted as evidence in favour of the Copernican heliocentric model.

The basic measurement technique in astronomy is to determine the angles to a celestial object from two different places, and then deduce the distance to the object from known distance between the observation points (called the
Galileo’s theorem states that a body will fall freely down the diameter $AB$ of a circle in the same amount of time that it will fall along any chord $AC$ of the circle. The theorem follows from the geometrical relation $AC/AB = \sin \alpha$ and the time-squared law $d = at^2/2$ for the distances traversed. Based on this Galileo inferred that the period of a pendulum, $\tau$, is independent of the amplitude of its swing, since the time to travel along any chord drawn to $B$ will be the same as the time it takes the body to fall freely down twice the length $OD$ of the pendulum [8]. Galileo also observed that $\tau$ depended on the pendulum’s length, but not weight.

Ptolemy of Alexandria (90-168 AD) made a fairly accurate estimate of the distance between the Earth and the Moon by measuring the angles of sight to the latter at different times, which is equivalent to making measurements at two different places because of the Earth’s daily rotation.

Until the invention of the telescope, the instruments constructed by Tycho Brahe (1546–1601) produced the most accurate measurements of the positions of the planets (see Fig. 1.4). After Galileo’s pioneering application of telescopes in astronomy, the ability to measure angles of sight precisely gradually improved to the point that, when combined with a newfound precision in measuring locations on the Earth’s surface, it enabled astronomers to determine, in 1672, the distance from the Earth to the Sun. This was a great breakthrough, for one could then use the diameter of the Earth’s orbit around the Sun as the new baseline, thereby considerably enhancing the difference between the angles of sight. A century and a half later, Thomas Henderson (1798–1844) and Friedrich Bessel (1784–1846) determined the distances to our nearest stellar neighbours.

The astronomical parallax of Sirius, a nearby star, is 0.38 arcsec, about $5 \times 10^{-4}$ the angular diameter of the Moon. That is the apparent size of a coin at a distance of a few kilometers!
In 1609, after a decade of painstaking investigation using Tycho Brahe’s astronomical data, Johannes Kepler (1571–1630) announced [10] the following two laws of planetary motion:

(1) The orbit of Mars is an ellipse with the Sun at one focus;

(2) The planet’s orbital velocity changes in such a way that the line joining Mars to the Sun covers equal areas of the ellipse in equal times.

Kepler’s third law of planetary motion [11], established in 1619, states that:

(3) The square of the orbital period of a planet is proportional to the cube of its mean distance from the Sun.

Kepler’s momentous discovery regarding planetary orbits had to wait for two great advances in mathematics, analytic geometry and the calculus, before a self-consistent theory of celestial mechanics could be formulated. Analytic geometry took a definite shape in 1637 through the work of René Descartes (1596–1650), who fused algebra and geometry to create an essential foundation for the development of the differential and integral calculus. Descartes’ treatise La Géométrie is a culmination of a century-long effort, in particular by Francois Viète (1540–1603), to create modern symbolic algebra. Analytic ge-
Fig. 1.4 These copper-plate engravings by Joan Blaeu were revised from wood-cuts originally published in Brahe’s book Astronomiae Instauratae Mecanica (1598). “This parallactic instrument, which is also called the Ptolemaic rulers, . . . is for observations of the zenith distances of the stars, in the way Ptolemy used to do it with the moon, particularly to find its maximum latitude . . . A few years earlier I had built some other rulers . . . In doing so I followed a particular method, which was partly an imitation of the ancient principle of Hipparchus . . . The use of the instrument depends on the fact that it will measure the altitude . . . as well as azimuths . . .” (pages 28–31 in Astronomiae).

Geometry was invented independently by Pierre de Fermat (1601–1665), who was also a founder, with Blaise Pascal (1623–1662), of the theory of probability. Descartes’ cartesian coordinates made possible the drawing of tangents to curves of any kind. Fermat’s method of finding the maximum or minimum value of any algebraic function involved the modern form of differentiation. “Kepler was the first to introduce the idea of infinity into geometry and to note that the increment of a variable was evanescent [infinitesimal] for values of the variable in the immediate neighbourhood of a maximum or minimum” [12].
1.3 Isaac Newton and his *Principia Mathematica*

The formulation of *classical mechanics* by Isaac Newton (1642–1727) represents a giant milestone in the history of science. In his masterpiece *Philosophiae Naturalis Principia Mathematica* [13], published in 1687, Newton shows that physical phenomena can be explained by a simple set of laws expressed in mathematical form. In particular, he demonstrates that the laws of planetary motion discovered by Johannes Kepler can be accounted for by the same law of gravitation as the motion of terrestrial objects in free fall.

In the years 1664–1666 Newton brought to fruition the ideas of his predecessors, most notably John Wallis (1616–1703) and Isaac Barrow (1630–1677), by developing the *differential calculus* (‘the method of fluxions’) and the *integral calculus* (‘the inverse method of fluxions’). His most crucial insight was that integration and differentiation are inverse operations (the *fundamental theorem of calculus*) [14]. Newton’s generalised binomial theorem

\[(P + PQ)^{m/n} = P^{m/n} + AQ(m/n) + BQ(m - n)/2n + CQ(m - 2n)/3n + \ldots\]

which enabled him to find infinite series for algebraic functions, played a major role in his work on calculus. Newton also invented the *calculus of variations*.

"The same year [1666] I began to think of gravity extending to the Orb of the Moon, and having found out how to estimate the force with which [a] globe revolving within a sphere presses the surface of the sphere, from Kepler’s Rule of the periodical times of the Planets . . . I deduced that the forces which keep the Planets in their Orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve; and thereby compared the force requisite to keep the Moon in her Orb with the force of gravity at the surface of the earth, and found them answer pretty nearly" [15].

In 1679, Newton established that Kepler’s areal law was a consequence of centripetal forces. He also showed that if the orbit is an ellipse under the action of central forces then the radial dependence of the force is the inverse square with the distance from the center. These propositions were stated in 1684 in his treatise *De Motu Corporum in Gyrum* (The Motion of Revolving Bodies) [15].

Newton made a crucial step on the path to his *Principia* when he proved, in 1685, that the gravitational attraction of a spherically symmetric body can be determined by placing all its mass at its center. Newton’s ‘superb theorem’ was crucial for an exact solution of the problem of finding the orbit of a planet under the attraction of the Sun (neglecting the attractions between planets).

In the first section of the *Principia* Newton presents his definitions of mass, ‘quantity of motion’ (momentum), and three types of forces: inertial, impressed and centripetal. Following these definitions is a *Scholium* on absolute time, space and motion. Both absolute time and absolute location are quantities that cannot themselves be observed, but instead have to be inferred from measurements of the corresponding relative quantities; such measurements
are always provisional. Motion in the solar system, e.g., is referred to the fixed stars, and sidereal time is provisionally taken as the preferred approximation to absolute time. "The causes which distinguish true motions from relative motions are the forces impressed upon bodies to generate motion". Newton argued that the concave surface of the water in a rotating bucket is created by absolute rotation (acceleration) with respect to absolute space.

The next section contains Newton’s laws of motion:

**Law 1:** Every body perseveres in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed;

**Law 2:** A change in motion is proportional to the motive force impressed, and takes place along the straight line in which that force is impressed;\(^5\)

**Law 3:** To every action there is always opposed an equal reaction; or the mutual actions of two bodies upon each other are always equal and directed to contrary parts.

The main body of the *Principia* is divided into three books. Book I opens with an exposition of the underlying mathematics, a geometrical form of infinitesimal calculus based on the ‘method of ultimate ratios’ of ‘evanescent quantities’ (infinitesimals) [17]. Such quantities are represented by lines and arcs of curves of arbitrary small length, a procedure that had been used by Archimedes for calculating lengths and areas encompassed by curves.

\(^5\) The familiar \(F = ma\) form of Newton’s second law can be traced back to Jacob Hermann’s *Phoronomia* [16].
In the rest of the *Principia* Newton analyses the motion of bodies in non-resisting (Book I) and resisting (Book II) media under the action of centripetal forces — which is the central theme of his masterpiece — and applies the results to orbiting bodies, projectiles, pendulums and free-fall under gravity. Newton’s law of universal gravitation, which states that all matter attracts all other matter with a force proportional to the product of their masses and inversely proportional to the square of the distance between them, is deduced from the astronomical phenomena discussed in Book III. Newton utilises this law to explain tides and their variations, the precession of the Earth’s axis, the motion of the Moon as perturbed by the gravity of the Sun, and the eccentric orbits of comets.6

*Proposition VI* in Book I establishes Newton’s law of motion under centripetal forces and along an arbitrary trajectory, such that equal areas are swept in equal times with respect to the acceleration center, in accord with *Proposition I* in Book I. In Andrew Motte’s translation from Latin (1729), *Proposition VI* reads: "In a space void of resistance, if a body revolves in any orbit along an immoveable centre, and in the least time describes any arc just then nascent; and the versed sine of that arc is supposed to be drawn bisecting the chord, and produced passing through the centre of force; the centripetal force in the middle of the arc will be as the versed sine directly and the square of the time inversely."

Assume, for the moment, that the ellipse in the diagram in Fig. 1.5 represents an arbitrary trajectory. A particle in orbital motion can be thought of as falling freely toward the acceleration center. The central force at P is then proportional to the displacement from the tangent QR over a short increment of time, divided by the square of this time: \( F \propto a \propto x/t^2 \) (Galileo’s time-squared law). Since the time is proportional to the area swept out, which in the limit \( Q \to P \) is the triangular area \( SP \cdot QT/2 \), the centripetal force must vary along the trajectory as \( 1/SP^2 \) times \( QR/QT^2 \) (where \( QR \) is also called "versed sine"). This proves Newton’s *Proposition VI*.

*Proposition XI Problem VI* reads: "If a body revolves in an ellipsis; it is required to find the law of the centripetal force tending to the focus of the ellipsis." In the ellipse in Fig. 1.5, Newton draws two conjugate diameters DK and PG, with DK parallel to the tangent RPZ. From Q he draws three lines: QR parallel to the focal radius SP, QT perpendicular to SP, and Qx completing the parallelogram QxPR. He then extends Qx until it meets PG at \( v \), and draws PF perpendicular to DK. Newton’s analysis makes use of the following lemmas: (1) PE = AC; (2) All parallelograms circumscribed about any conjugate diameters of an ellipse have equal area; (3) In an ellipse, the squares of the ordinates of any conjugate diameter are

---

6 Gottfried Leibniz was the first to express Newton’s theory of orbital motion in the form of a differential equation [18]. Leonhard Euler synthesized Leibniz’s differential calculus with Newton’s method of fluxions into mathematical analysis.
proportional to the rectangles under the segments which they make on the diameter. The first lemma is Newton’s and the other two are attributed to Apollonius of Perga (262–190 BC), who is famous for developing the theory of conic sections. Since QR is Px and (by Lemma 1) PE is AC, the similarity of the triangles PxV and PEC implies QR = (Pv · AC)/PC. On the other hand, the similarity of the triangles QxT and PEF, combined with Newton’s lemma, leads to QT = (Qx · PF)/AC = (Qx · BC)/CD, where the second equality follows from Lemma 2 (which implies PF · CD = BC · AC). It is now straightforward to show that QT²/QR = (L/2) (Qx² · PC)/(Pv · CD²); here L = 2BC²/AC is the latus rectum of the ellipse. In the limit Q → P, the ratio Qv/Qx → 1. Furthermore, Lemma 3 implies Qv²/(Pv · vG) = CD²/PC². Hence, one finally obtains QT²/QR = (L/2)(vG/PC) → L because vG → 2PC as Q → P.

This completes Newton’s analysis for an elliptic orbit:

\[
\text{Centripetal force } \propto \frac{1}{L \cdot SP^2}
\]

Since L is a constant, the centripetal force is reciprocal to the square of SP, the distance to the focus of the ellipse.\(^7\)

To further illustrate the mathematical reasoning that characterizes the Principia, Newton’s geometrical proof of Proposition LXX in Book I will be presented. The proposition states: "If to every point of a spherical surface there tend equal centripetal forces decreasing as the square of the distances from those points, I say, that a corpuscle placed within that surface will not be attracted by those forces any way" [15]. For the analogous case in electrostatics, Newton’s proposition was tested by the null experiment of Henry Cavendish (see Fig. 1.10).

Newton’s proof of Proposition LXX is based on Fig. 1.6. Consider a cone of infinitesimally small solid angle dΩ that intersects a homogeneous spherical shell of matter in both directions. The intersection areas around the points Q and Q’ are denoted by dS and dS’, respectively. The apex of the cone is at the position of a test body placed inside the shell (point P). The forces of attraction at P, due to the source-surface areas dS and dS’, are equal and opposite because: (1) the gravitational force decreases as the square of the distance, while the surface areas of the source, cut out by the cone of a given solid angle dΩ, increases as the square of the distance; and (2) the angles between the infinitesimally thin two-way cone and the normals to the spherical shell at both intersection points (radial lines OQ and OQ’) are equal. Since the entire solid angle around the point P can be divided into such double cones, the resultant attraction is zero.

\(^7\) In the first ten sections of the Principia Newton speaks of “force”, but he actually calculates accelerations. In his study of orbital motions, Newton does not comment on the cause of gravity; this issue is addressed in the General Scolium on the final pages of the Principia.
1.3 Isaac Newton and his Principia Mathematica

The proofs of Newton’s propositions involve limits, derivatives, integrals, curved paths, acceleration, etc. In other words, the mathematics of the Principia is calculus disguised in the form of geometry, as stated earlier.

Newton realized that the motions of Jupiter’s moons and of the Earth-Moon system in the Sun’s gravitational field would be greatly altered if the observed equality between gravitational and inertial mass did not hold. He demonstrated this equality (known as the weak equivalence principle) to a precision of about one part in $10^3$ using pendulums made from different materials. A generalization of the weak equivalence principle played a crucial role in the formulation of general relativity.

In the General Scholium, added at the end of the second edition of the Principia, Newton writes: "I have not as yet been able to deduce from phenomena the reason for these properties of gravity, and I do not feign hypotheses. For whatever is not deduced from the phenomena must be called a hypothesis; and hypotheses . . . have no place in experimental philosophy. In this experimental philosophy, propositions are deduced from the phenomena and are made general by induction".

Newton originated the science of spectroscopy by demonstrating that white light is composed of a spectrum of colours, and proposed a corpuscular theory of light. He invented and built a new type of telescope, the design of which is the prototype for all modern, large optical telescopes. In the opinion of Albert Einstein, Newton "determined the course of western thought, research and practice like no one else before or since" [19].
1.4 Development of Analytical Mechanics by Euler, Lagrange and Hamilton

Classical mechanics was reformulated by Joseph Lagrange (1736–1813) about a century after the publication of Newton’s *Principia*. Whereas in Newtonian mechanics one deals with forces nowadays expressed in terms of vectors, in Lagrangian mechanics the dynamics of a system is governed by a scalar function $\mathcal{L}(q_i, \dot{q}_i, t)$, called the Lagrangian ($\dot{q}$ is the time derivative of $q$). The trajectory of the system is determined by solving, for each of the generalized coordinates $q_i$ ($i = 1, \ldots, n$), the second-order partial differential Euler-Lagrange equation

$$\frac{d}{dt} \left( \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0$$

(Euler-Lagrange equation) (1.1)

For a non-relativistic particle in one dimension, $\mathcal{L} \equiv E_k - E_p = m\dot{x}^2/2 - V(x)$; here $E_k$ and $E_p$ are the kinetic and potential energies, respectively, and $V(x)$ is a potential function. In this simple case, the Euler-Lagrange equation (1.1) yields Newton’s second law: $\frac{d}{dt}(m\dot{x}) - \frac{d}{dx}(-V) = 0$, that is, $m\ddot{x} = F(x)$. In general, the Lagrangian does not have the form of the kinetic energy minus the potential energy — its actual form for any particular case is a pure guess!

The Lagrangian formalism, which can easily be extended to describe systems outside the realm of Newtonian mechanics, is one of the mathematical cornerstones of modern theoretical physics. In the preface to his *Méchanique Analytique*, published in 1788, Lagrange proudly declares: "The methods that I expound require neither constructions, nor geometrical or mechanical arguments, but only algebraic operations, subject to a regular and uniform course" [20].

The Euler-Lagrange equation was first derived around 1744 using the calculus of variations (see Subsection 1.4.1). Originally invented by Newton while trying to determine the shape of a moving object that would guarantee the least possible resistance, the calculus of variations traces its inception to 1696, the year when Johann Bernoulli (1667–1748) formulated the *brachistochrone problem*. The first major contributions to this venerable branch of mathematics were made by Leonhard Euler (1707–1783) and Joseph Lagrange [21].

"Once the laws of physical theory are expressed as differential equations, the possibility of their reduction to a variational principle is evident from purely mathematical reasoning . . ." [22].

A variational principle is a general rule in physics expressed in terms of the calculus of variations; it states that some functional of dynamical variables (called action) is stationary with respect to small variations of these variables.

The action $S$ of a system is an integral of its Lagrangian:

$$S[q(t)] \equiv \int_{t_1}^{t_2} \mathcal{L}(q_i, \dot{q}_i, t) \, dt$$

(1.2)
The value of $S$ depends on the entire trajectory along which the integration is carried out. Hamilton’s variational principle states that, out of all possible ‘paths’ in configuration space $(q_1 \ldots q_n)$, a particle will travel along the path for which the value of the integral (1.2) is the same (to first order in $\delta q$) as that for any path which differs from it by an infinitesimal displacement:

$$\delta S = 0 \quad \text{Hamilton’s principle}$$

(1.3)

This principle was enunciated by William Rowan Hamilton (1805–1865) in 1834 under the name of the law of varying action [23]. It is straightforward to prove that the Euler-Lagrange equations follow from Hamilton’s principle. The variation of the action reads

$$\delta S = \int_{t_1}^{t_2} dt \delta \mathcal{L} = \int_{t_1}^{t_2} dt \left\{ \frac{\partial \mathcal{L}}{\partial q_i(t)} \delta q_i(t) + \frac{\partial \mathcal{L}}{\partial \dot{q}_i(t)} \delta \dot{q}_i(t) \right\}$$

Writing $\delta \dot{q}_i(t) = \frac{d}{dt} \delta q_i$, the second term can be integrated by parts:

$$\int_{t_1}^{t_2} dt \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta q_i = \left. \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta q_i \right|_{t_1}^{t_2} - \int_{t_1}^{t_2} dt \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta q_i$$

But $\delta q_i(t_1) = \delta q_i(t_2) = 0$, and so the first term on the right is zero. Thus,

$$\delta S = \int_{t_1}^{t_2} \left\{ \frac{\partial \mathcal{L}}{\partial q_i} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right\} \delta q_i dt$$

which vanishes provided that Eq. (1.1) holds.

Hamilton’s principle apparently implies that, before it moves from one point in the configuration space to another, a particle ‘calculates’ the value of $S$ for every possible path linking these points and then follows the one for which $S$ is a minimum. In fact, the particle merely obeys the Euler-Lagrange equations of motion (i.e., Newton’s laws) at each point to minimize the action, which means that every subsection of the actual path must be a minimum. In other words, the particle somehow ‘probes’ the potential within an infinitesimal region of spacetime, and moves in the direction of greatest change.

The quantity $p_i \equiv \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$, introduced by Hamilton, is called the generalized momentum, conjugate to $q_i$. As the name implies, $p_i$ is not always the familiar linear momentum of a particle. For instance, the conjugate momentum of an angular degree of freedom is an angular, rather than a linear, momentum.

If a Lagrangian does not contain a coordinate $q_i$ (though it may contain $\dot{q}_i$), then Eq. (1.1) implies that the generalized momentum conjugate to $q_i$ is conserved. For a system that is invariant, e.g., under translation along a given direction, the corresponding linear momentum is conserved. Thus, momentum conservation is closely connected with the symmetry properties of the system.

---

8) When applied to a function, for instance, the symbol $\delta$ represents the variation of the function (i.e., it leads to a different function), whereas the differential takes us to a different point of the same function.
In an alternative formulation of classical mechanics, proposed in 1835 by Hamilton [24], the generalised coordinates \((q_i)\) and the conjugate momenta \((p_i)\) appear on what is an essentially equal footing:

\[
\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \quad \text{Hamilton’s equations (1.4)}
\]

The Hamiltonian \(H\) is the Legendre transform of the Lagrangian \(L\):

\[
H(q_i, p_i, t) \equiv \sum_i p_i \dot{q}_i - L(q_i, \dot{q}_i, t) \quad \Rightarrow \quad L = p \cdot q - H(q, p, t) \quad (1.5)
\]

where \(\sum_i p_i \dot{q}_i = p_1 \dot{q}_1 + p_2 \dot{q}_2 + \cdots + p_n \dot{q}_n \equiv \mathbf{p} \cdot \mathbf{q}\) is the scalar product of two vectors. The pair \((q, p)\) defines a point in \(2n\)-dimensional phase space. A non-relativistic particle in one dimension has the total energy given by the Hamiltonian \(H(x, p)\)

\[
E = \frac{p^2}{2m} + V(x).
\]

The coupled first-order Hamilton’s canonical equations (1.4) yield \(\dot{x} = p/m\) and \(\dot{p} = -\partial V/\partial x\). One thus obtains the same result as using the Lagrangian formalism: \(m \ddot{x} = -\partial V/\partial x\).

The most compelling reason for studying Hamiltonian dynamics is to acquire a deeper insight into the structure of classical mechanics and its relation to other branches of physics. The relation between the Hamilton-Jacobi theory of classical dynamics (see below) and the short-wavelength limit of wave optics is a bridge between classical and quantum mechanics. The Hamilton-Jacobi theory was originally expounded by Hamilton in his essay [24], and was further developed by Carl Gustav Jacobi (1804–1851) [25]. In Paul Dirac’s opinion: "Hamilton seemed to have some remarkable insight into what was important . . . He found a form of writing the equations of mechanics whose importance would be realized only after a hundred years" [26].

Jacobi developed a technique for carrying out a transformation from \((q, p)\) to a new set of coordinates \(Q = Q(q, p, t)\) and \(P = P(q, p, t)\) for which canonical equations (1.4) still hold (canonical transformation), and for which the equations of motion are directly integrable. Since the new coordinates satisfy

\[
\dot{P}_i = -\frac{\partial \mathcal{H}}{\partial Q_i}, \quad \dot{Q}_i = \frac{\partial \mathcal{H}}{\partial P_i}
\]

this implies that \(\delta \int [\mathbf{p} \cdot \dot{q} - \mathcal{H}(q, p, t)] dt = 0 = \delta \int [\mathbf{P} \cdot \dot{Q} - \mathcal{H}(Q, P, t)] dt\). Note that the two integrands are not necessarily equal. However, a total time derivative of an arbitrary function \(G\) can be added to either integrand without affecting the variations \(\int_{t_1}^{t_2} dG = \text{const.}\). One can therefore write

\[
\sum p_i \dot{q}_i - \mathcal{H}(q_i, p_i, t) = \sum P_i \dot{Q}_i - \tilde{\mathcal{H}}(Q_i, P_i, t) + \frac{dG}{dt}
\]

The generating function \(G\) ensures that the transformation is non-trivial. By choosing \(G\) to depend on a particular mixture of the old and new coordinates,
canonical transformations can be classified into four basic types [27]. Substituting, for instance, \( G \equiv F(q, P, t) - \sum Q_i P_i \) in the last equation results in

\[
\sum_{i=1}^{n} \left[ \left( p_i - \frac{\partial F}{\partial q_i} \right) dq_i + \left( Q_i - \frac{\partial F}{\partial P_i} \right) dP_i \right] + \left( \tilde{\mathcal{H}} - \mathcal{H} - \frac{\partial F}{\partial t} \right) dt = 0
\]

Since the components \( q_i \) and \( P_i \) are separately independent, their coefficients must vanish independently. This yields the canonical transformation equations

\[
p_i = \frac{\partial F(q_i, P_i, t)}{\partial q_i}, \quad Q_i = \frac{\partial F(q_i, P_i, t)}{\partial P_i}, \quad \mathcal{H} + \frac{\partial F}{\partial t} = \tilde{\mathcal{H}}
\]

If \( \tilde{\mathcal{H}} = 0 \), in which case \( \dot{Q}_i = 0 \) and \( \dot{P}_i = 0 \), the third equation becomes

\[
\mathcal{H} \left( q_i, \frac{\partial F}{\partial q_i}, t \right) + \frac{\partial F}{\partial t} = 0 \quad \text{Hamilton-Jacobi equation} \quad (1.6)
\]

Hamilton’s principal function \( F \) is customarily denoted by \( S \). Since \( P_i \) and \( Q_i \) are constants \( (P_i \equiv \alpha_i, Q_i \equiv \beta_i) \), the total time derivative of \( S \) reads

\[
\frac{dS}{dt} = \sum \frac{\partial S}{\partial q_i} \dot{q}_i + \frac{\partial S}{\partial t} = \sum p_i \dot{q}_i - \mathcal{H} \equiv \mathcal{L}
\]

Therefore, \( S \) differs from the indefinite time integral of \( \mathcal{L} \) only by a constant: \( S = \int \mathcal{L} dt + \text{constant} \). If \( \mathcal{H} \) does not depend explicitly on time, one can substitute \( S(q_i, \alpha_i, t) = W(q_i, \alpha_i) - Et \) in (1.6) and obtain \( \mathcal{H}(q_i, \partial W/\partial q_i) = E \).

1.4.1

The Brachistochrone: Gateway to the Calculus of Variations

In the June 1696 issue of the journal *Acta Eruditorum*, Johann Bernoulli posed the following mathematical problem:

- Find the curve connecting two points, at different heights and not on the same vertical line, along which a body acted upon only by gravity will fall in the shortest possible time.

Bernoulli named the curve *brachistochrone*, a term he coined from the Greek words *brachistos* (shortest) and *chronos* (time). The May 1697 issue of *Acta Eruditorum* contained four solutions to this problem, provided by the greatest mathematicians of the time: Isaac Newton, Gottfried Leibniz (1646–1716), Johann Bernoulli, and his elder brother Jacob Bernoulli (1654–1705). Their answers all agreed, but the methods of derivation differed considerably [28].

---

9) Hamilton’s principal function is the generator of a canonical transformation to constant coordinates and momenta. The problem of solving the entire set of canonical equations is replaced by the problem of solving a single partial differential equation [27].
Johann Bernoulli’s solution is of particular interest because it represents the earliest known example of the analogy between optics and mechanics [29]. The solution utilizes Fermat’s principle of least time, which is equivalent to the optical law of refraction \( \sin \alpha_1 / v_1 = \sin \alpha_2 / v_2 \) (Snell’s law); here \( \alpha_1 \) and \( \alpha_2 \) are the angles of incidence and refraction, respectively, and \( (v_1, v_2) \) are the velocities of light in the two media (see Section 1.5).

Consider a ray of light propagating through a sequence of horizontal, thin layers of decreasing optical density. As the speed of light increases, so does the angle of refraction; that is, the ray of light follows a curved path. This inspired Bernoulli to come up with the following optical-mechanical analogy: The constantly increasing speed of a body sliding down a brachistochrone curve corresponds to a light ray propagating along a curved optical path through an optical medium of ever-decreasing optical density.

For a very large number of infinitesimally thin layers, Snell’s law can be expressed as \( \sin \alpha / v = \kappa \), where \( \kappa \) is constant. The speed of a body that falls from rest starting at \( y = 0 \) is \( v = gt \), and the distance of fall is \( y = gt^2 / 2 \) (Galileo’s time-squared law), where \( g \) is the acceleration due to gravity; hence, \( v = \sqrt{2gy} \). Combining this expression with Snell’s law gives \( y = k \sin^2 \alpha \), where \( k = 1/2g \kappa^2 \). If \( v \) is the velocity at a point \((x, y)\), then \( y' = dy/dx \) and \( \sin^2 \alpha = 1/(1 + y'^2) \). From this result and \( y = k \sin^2 \alpha \) follows a differential equation describing the motion of the body: \( y(1 + y'^2) = 2h \), where \( h = k/2 \).

The body’s trajectory is a cycloid described by the parametric equations \( x(\theta) = h(\theta - \sin \theta) \) and \( y(\theta) = h(1 - \cos \theta) \), as can be readily verified by direct substitution in the differential equation derived above. A cycloid curve represents the trajectory of a point on the circumference of a wheel that rolls without slipping along the \( x \)-axis. To completely solve the brachistochrone problem, one has to find the right cycloid that connects the two given points. Newton’s elegant solution, originally published in *Philosophical Transactions* [28], reflects the mastery of mathematical techniques Newton developed in 1685 for solving the first genuine problems of the calculus of variations [29].

Johann Bernoulli was greatly impressed that the solution of his brachistochrone problem was the same as that of the following tautochrone problem (tautos means identical): Determine the curve for which the time taken by an object sliding without friction in uniform gravity to its lowest point is independent of its starting point. The curve is a cycloid, and the time is equal to \( \pi \) times the square root of the radius over the acceleration of gravity. The tautochrone problem was solved geometrically in 1659 by Christiaan Huygens (1629–1695), who found that a cycloid was the curve along which a pendulum would have to descend to be exactly isochronous [28].

In 1656 Huygens invented the pendulum clock. He was the first to derive the formula \( T = 2\pi \sqrt{\ell / g} \), where \( T \) is the period of a pendulum, \( \ell \) its length and \( g \) the gravitational acceleration [30].
1.4 Development of Analytical Mechanics by Euler, Lagrange and Hamilton

The novelty of the brachistochrone problem lies in the fact that the quantity under consideration, the time of descent, depends on the whole curve. This problem stimulated the development of ideas and techniques that led to the branch of mathematics known as the calculus of variations.

The calculus of variations has many applications in physics. The laws of motion and of equilibrium are dominated by maximum and minimum principles. For instance, stable equilibrium of a mechanical system is attained if the system is arranged in such a way that its potential energy is a minimum. Euler and Lagrange were the first to develop "more general methods for solving extremum problems in which the independent element was . . . a whole curve or function or even a system of functions" [31].

In 1744, Euler published a seminal treatise entitled Methodus Inveniendi [32], in which a mathematical method for solving a general class of variational problems was presented. Given a function \( y(x) \) satisfying the boundary conditions \( y(x_A) = y_A, \ y(x_B) = y_B \) and \( x_A < x_B \), he showed that

\[
\frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) = 0
\]

is a necessary condition for the definite integral \( I = \int_A^B f(y, y', x)dx \) to be a minimum. To derive this Euler-Lagrange equation, Euler replaced the integral \( I \) by the sum \( S = \sum_j f(y_j, z_j, x_j)(x_{j+1} - x_j) \), where \( z_j \equiv (\Delta y / \Delta x)_x = x_j \), and asked for the stationary value of \( S \). This problem can be solved by setting to zero the partial derivatives of \( S \) with respect to \( y_k \). The resulting difference equation becomes a differential equation in the limit \( \Delta x_j \to 0 \) [29, 33]. In the treatise, Euler also formulated his principle of least action (see Section 1.5).

In 1755, Joseph Lagrange (born Giuseppe Lodovico Lagrangia) was only nineteen years old when he sent a letter [34] to Euler explaining how Euler’s tedious geometrical reasoning could be replaced by a method that only requires a straightforward use of the principles of the differential and integral calculus. Lagrange’s first published account of the calculus of variations [35] appeared in 1760. In the first essay [35] he formulated a new calculus based on the symbol \( \delta \) (see Eq. (1.3)). This enabled him to solve the following general problem: given an expression involving several variables and their derivatives, find the functional relation between these variables which renders the definite integral of the given expression an extremum. The second essay [35] contains an extensive application of variational techniques to the principle of least action in dynamics. The publication of his Méchanique Analytique marks the culmination of Lagrange’s early work on the foundations of the calculus of variations.

"Euler’s variational principles of physics, rediscovered and extended by [Hamilton], have proved to be among the most powerful tools in mechanics, optics and electrodynamics . . . Recent developments in physics — relativity and quantum theory — are full of examples revealing the power of the calculus of variations" [31].
1  Physics from Archimedes to Rutherford

1.5  Properties of Light

The earliest use of a variational principle was in optics, the branch of physics describing the behaviour and properties of light. The nature of light has been studied from ancient Greek times up to the present day. The law of reflection, which states that the angle of incidence equals the angle of reflection, was formulated by Euclid. He pragmatically considered light as rays, i.e., something that propagates through space in straight lines. In his book *Catoptrica*, Greek mathematician and ingenious inventor Heron of Alexandria (1st century AD) showed that when a light ray is reflected by a mirror, the path actually taken from the object to the observer’s eye is shorter than any other possible path so reflected.

Around 1000 AD, Ibn al-Haytham (965–1039) established optics as an empirical science. He observed that light coming through a tiny hole travelled in straight lines and projected an image onto the opposite wall. In his writings al-Haytham argues that we see objects because they reflect the Sun’s rays of light, that light must travel at a large but finite velocity, and that light refraction is caused by the velocity being different in different substances.

The law of refraction of a ray of light at the interface between two media of different optical density (Snell’s law) states that the ratio of the sines of the angles of incidence and refraction is equivalent to the ratio of refractive indices of the two media: \( \sin \alpha_1 / \sin \alpha_2 = n_2 / n_1 \). This law, formulated by Ibn Sahl (940–1000) and rediscovered in 1601 by Thomas Harriot (1560–1621), can be explained by the principle of least time. According to this principle, stated in 1657 by Pierre de Fermat [36], a ray of light propagates between two points along the path that minimizes the travel time. Since the (phase) velocity of light at position \( r \) in a medium with refractive index \( n(r) \) is given by \( v(r) = c / n(r) \), where \( c \) is the speed of light in vacuo, Fermat’s principle can be expressed as

\[
\delta \int \frac{dr}{v} = \delta \int n(r) dr = 0 \quad \text{Fermat’s principle}
\]

The modern version of Fermat’s principle states that the optical path length must be stationary, meaning that it can be either minimal, maximal or a point of inflection.

Fermat’s principle is analogous to Maupertuis’ principle of least action. Formulated in its present form in 1744 by Leonhard Euler [32], Maupertuis’ principle claims that a particle with constant energy propagates between two points along the path that minimizes the integral of the momentum over the distance traversed: \( \delta \int p \cdot dr = 0 \). Guided by this analogy, Hamilton suggested that the formalisms of optics and mechanics could be unified. The recognition that classical mechanics was the short-wavelength limit of wave mechanics, in the same way as geometrical optics is the short-wavelength limit of wave optics, came only after the interference experiments with electrons had revealed effects that depend on the particle wavelength (see Sect. 2.8).
**Light diffraction** was described by Francesco Grimaldi (1618–1663) in his treatise *Physical Science of Light*, published in 1665. Grimaldi made careful observations of the shadows cast by opaque bodies when illuminated by narrow beams of sunlight, and found that the shadow boundary between light and dark was not sharply defined. While observing the shadow of a thin rod, he observed coloured fringes inside and parallel to the shadow region. Grimaldi concluded that light should be regarded as a fluid possessing wave nature, since fluids also exhibited diffractive behavior.

Robert Hooke (1635–1703) thought that light consists of rapid vibrations propagated with a very great speed [37]. He also believed that, in a homogeneous medium, every vibration would generate a steadily growing spherical disturbance. In 1672, Hooke suggested that the vibrations of light might be perpendicular to the direction of propagation.

Also in 1672, Newton published a treatise entitled *New Theory about Light and Colors* [38], which describes his ground-breaking discovery, made in 1666, that white light can be split into component colours while passing through a prism, and that each pure colour is characterized by a specific refrangibility: "Light itself is a Heterogeneous mixture of differently refrangible Rays . . . As the Rays of light differ in degrees of Refrangibility, so they also differ in their disposition to exhibit this or that colour. Colours are not Qualifications of Light, derived from Refractions, or Reflections of natural Bodies . . . but Original and conname properties".

The wave theory of light developed around 1677 by Christiaan Huygens is based on the idea that each wavefront of an advancing wave of light acts as a source of a spherical wavelet, and that the wavefront at any instant conforms to the envelope of all these wavelets (*Huygens’ principle*) [39]. This theory could explain the laws of reflection and refraction, but did not account for the phenomenon of diffraction. In fact, Huygens’ wavelets are much like single ‘pulses’, rather than disturbances exhibiting periodic motion.

In 1669, Erasmus Bartholin (1625–1698) found that rays of light passing through a transparent crystalline substance known as calcite are split into two refracted rays. While studying this phenomenon, Huygens made the fundamental discovery of polarization: each of the two rays may be extinguished by passing it through a second calcite crystal aligned perpendicular to the first [39]. In 1808, Étienne Malus (1775–1812) discovered polarization of light by reflection. He observed the reflection of the sun from a window pane through a calcite crystal, and noticed that the two images obtained by double refraction varied in intensities as the crystal was rotated about the line of sight. He correctly concluded that the properties hitherto attributed to crystals could also be produced by reflection of light [40].

The wave nature of light was established by Thomas Young (1773–1829). In 1801, he surmised that different colours of light corresponded to different wavelengths, and that wave interference could explain the phenomenon of New-
ton’s rings. "The law is that wherever two portions of the same light arrive at the eye by different routes, either exactly or very nearly in the same direction, the light becomes most intense when the difference of the routes is any multiple of a certain length, and least intense in the intermediate state of interfering portions; and this length is different for light of different colours" [41]. It is clear from this statement that Young appreciated the requirement of coherence. However, he assumed light to be a longitudinal wave disturbance, like sound.

A concentric ring pattern of rainbow colours, known as Newton’s rings, is created by sunlight when it falls normally on the flat surface of a convex lens placed on a glass plate. This pattern is formed because the different wavelengths of light interfere at different thicknesses of the air layer between the two glass surfaces. When viewed with monochromatic light, one observes a series of concentric, alternating light and dark rings centred at the point of contact between the glass surfaces. The light (dark) rings are caused by constructive (destructive) interference between the light rays reflected from both surfaces. Young explained the dark spot in the middle by proposing that "where one of the portions of light has been reflected at the surface of a rarer medium, it must be supposed to be retarded one half of the appropriate interval . . ." He proved this conjecture "by interposing a drop of oil of sassafras between a prism of flint-glass and a lens of crown glass". Since the index of refraction of sassafras oil is intermediate between those of crown and flint glass, "the central spot seen by reflected light was white, and surrounded by a dark ring" [41]. In his Opticks [42], Newton states that the separation between the two glass surfaces at any given distance from the point of contact always changes by the same amount in moving from one dark ring to the next. Based on Newton’s remarkably accurate measurements, Young was able to estimate the wavelengths corresponding to particular colours.

In 1803, Young devised a diffraction experiment to prove the correctness of his interference theory [43]. He produced a narrow beam of sunlight by making a tiny hole in a window shutter. When a strip about 0.85 mm wide, cut from a card, was placed in the beam, a pattern of fringes appeared on the wall behind the strip. Young measured the dimensions of the observed fringes to calculate the difference in path-length, or wavelength of the light. He found that his measurements quantitatively correlated with those of Newton’s rings. Young also made the crucial observation that the pattern of fringes disappeared when a small screen was introduced to obstruct the light passing on one side of the strip.

For two rays of light to interfere, Young argues in his Lectures [44], "it is necessary that they be derived from the same origin, and that they arrive at the same point by different paths, in directions not much deviating from each other . . ." The simplest way to accomplish this is "when a beam of homogeneous light falls on a screen in which there are two very small holes or slits, which may be considered as
Fig. 1.7 Diagram illustrating interference of light in Young’s two-slit interference experiment. From Young’s Lectures on Natural Philosophy and the Mechanical Arts, 1807.

centres of divergence, from whence the light is diffracted in every direction . . .” (see Fig. 1.7). Experiments of this kind have become indispensable for studying the subtle nature of quantum phenomena.

The theory of light diffraction developed by Augustin Fresnel (1788–1827), and described in a series of papers submitted to l’Académie des Sciences from 1815 to 1819, combines Huygens’ envelope construction with Young’s principle of interference of longitudinal waves [45]. In Fresnel’s theory, each point along a diffraction slit becomes the source for a secondary radial wave. These waves interfere among themselves in such a way that their sum is determined by the relative phases as well as the amplitudes of the individual waves [46]. The squared absolute value of the summed amplitude determines the light intensity at each point in space. This theory correctly predicts, e.g., that diffraction around the edge of a small circular disk placed in the path of a light beam should lead to constructive interference at the center of the shadow cast by the disk [47]. Fresnel’s theory of light diffraction was further developed between 1821 and 1823 by Joseph von Fraunhofer (1787–1826).

“It was still to be determined whether two rays originally polarized at right angles would not produce phenomena of the same kind when they met inside the geometrical shadow of an opaque body” [48]. By placing a metal cylinder between the two refracted rays emerging from a calcite crystal, thus partly diffracting each beam, Fresnel and Dominique Arago (1786–1853) observed, in 1816, that light beams do not interfere if they are polarized in perpendicular directions [48]. This prompted Young to suggest that a light wave consists of oscillations occurring perpendicular to the direction of wave propagation [49].

Fresnel took the next, decisive step by asserting that “direct light can be considered as the union . . . of systems of waves polarized in all directions . . . the act of polarization consists not in creating these transverse motions, but in decompos-
ing them in two invariable directions . . ." [50]. To explain the birefringence and polarization in crystals observed by Bartholin and Huygens, Fresnel argued that a beam of monochromatic light entering a calcite crystal would be divided into two beams having orthogonal polarizations and different refractive indices. Fresnel’s theory predicts the intensity of reflected and refracted light, and incorporates the concepts of coherence and elliptical polarization. In 1882, Gustav Kirchhoff (1824–1887) put the theory of light propagation on a firm mathematical foundation by considering light waves as transverse oscillations of the ether [51].

The speed of light was first estimated in the 17th century from the orbital periods of Jupiter’s four largest moons. These periods revealed that the eclipses of Jupiter’s moons were consistently delayed by 22 minutes (17 minutes according to modern measurements) relative to the predicted times of occurrence (see Fig. 1.8). In 1675, astronomer Ole Rømer (1644–1710) reasoned that if light does not travel infinitely fast, then it must take some amount of time to get from Jupiter to the Earth. Based on how much the timing of Jupiter’s moon Io seemed to change and how much the Earth-Jupiter distance varied, he estimated the speed of light to be about 240,000 kilometres per second.

In 1727, James Bradley (1692–1762) determined the speed of light to within one per cent of the currently accepted value (see below) by measuring the aberration of starlight. Stellar aberration is the apparent slight variation in the ecliptic latitude of a star caused by the yearly motion of the Earth. To determine the speed of light, Bradley used the fact that the mean orbital speed of
the Earth is about 30 km/s. Bradley’s work provided the first direct evidence that the Earth revolves around the Sun [52] (see also [53]).

In 1849, Hippolyte Fizeau (1819–1896) became the first person to measure the speed of light without any recourse to astronomical observations. A year later, Léon Foucault (1819–1868) used a rotating-mirror method (see Fig. 1.9) to measure the speed of light in air, and obtained a value of $2.98 \times 10^{10}$ cm s$^{-1}$ — the first accurate measure of this fundamental physical constant. He also showed that light travels slower in water than in air.\(^{11}\)

The velocity of light in moving media was measured by Fizeau in 1851 and 1859. His measurements [54] were of considerable historical importance in the development of relativity theory. In Fizeau’s experiment, a light ray from a source is divided by a silver-coated glass plate into two beams. One beam, which travels upstream through moving water, traverses a rectangular path before entering a telescope. The other beam, which travels downstream through the water, traverses the same rectangle in the opposite direction. The two beams interfere in the telescope producing measurable shifts in the interference fringes. The shifts are proportional to the time difference for light rays traversing the two paths. Fizeau’s experiments, which were repeated with greater precision by Albert Michelson and Edward Morley in 1886 and by Pieter Zeeman in 1914–1922, can be regarded as a test of both Einstein’s law of the addition of velocities and the electrodynamics in moving media.

\(^{11}\) Foucault used a freely suspended pendulum and the gyroscope he himself invented to demonstrate the Earth’s motion around its axis.