# Chapter 1
# Celestial and Mechanical Clocks

## 1.1 Cyclic Events in Nature

From the earliest times in the course of human development, a recurring theme has been the inexorable passage of time, bringing with it ever-changing aspects of Nature and the cycle of life and death. Only in the realm of mythology do immortal gods live outside of time in their eternal incorruptible abodes.

The discernment of an underlying order in the evolution of natural phenomena, and the cyclic repetition of the motions of the sun, moon, and stars, may be taken as a measure of man's intellectual development. The ease with which early man was able to recognize that certain changes in nature were cyclic depended on the length of the cycle. That of the daily rising and setting of the sun is so short that it must have soon been accepted with confidence as being in the natural order of things, that if the sun disappeared below the horizon, there was little doubt that it would reappear to begin another day. It is otherwise with the much longer period of the seasons; there is evidence to suggest that for some primitive peoples, a year was so long and the means of recording the passage of time so imperfect that they were unable to perceive a cyclic pattern at all in the changing seasons. They must have watched the changing elements with perpetual wonder. To them the onset of winter must have been filled with dire foreboding, giving rise, according to Frazer (1922) in his classic *The Golden Bough*, to magical, and later religious rites to ensure the return of spring.

This introduces a connection between the timing of important cultural events in the life of early man and that of the cyclic events in nature. This overlays an inherent connection on a biological level: The workings of the human body and indeed of all living creatures follow rhythmic patterns that shadow those in nature. The most obvious are the so-called *circadian rhythm* (from the Latin *circa* (about), and *dies* (day)) with an approximate 24-hour repetitive cycle. Much research has been conducted in recent years on the human asleep–awake cycle, with particular interest in the extent to which the cycle is governed by

F.G. Major, The Quantum Beat

some internal timing mechanism, as opposed to the external environment. The need to adapt one's schedule of activities to be in harmony with nature, a task compounded by the differing cycles of natural events, is evident in the early development of calendars. Of course, the beginnings of agriculture gave a great impetus to this development in order to keep track of the seasons and accurately plan the cultivation of the soil, the planting of seeds, and the eventual harvest. The cyclic succession of the seasons—from the shedding of leaves in the fall to the cold dormancy of winter, the return to life in the spring, and the warm summer that followed—bore witness to some order underlying the vagaries of daily life. For those early societies whose life and livelihood were closely tied to the sea, the periodic rise and fall of the tide reinforced the same perception of unalterable periodic changes underlying unpredictable short-term changes.

A cyclic phenomenon clearly allows time to be quantified; the period of time to complete a cycle, called briefly the *period*, provides a unit in terms of which any given length of time can be expressed as so many of those units. An obvious example is the use of the (solar) day as a unit, defined as the time between the sun passing overhead one day until it comes to the same point the next day. Another common example is the lunar month, which is the time it takes the moon to go from (say) a new moon to the next new moon. As units of time, it is relevant to ask just how constant these units are, and how accurately they can be measured. Such questions are, of course, at the heart of our subject and are taken up in the chapters that follow.

## 1.2 The Calendar

It is unfortunate for those whose primary interest is in keeping track of the seasons that they do not recur after a whole number of days. As we all know, the year is about one-fourth of a day in excess of 365 days. In terms of the planetary motions of the earth, this is the same as saying that the period of the earth in its orbit around the sun does not contain a whole number of rotations of the earth about its axis. It is this simple fact that throughout history has complicated the lives of those charged with keeping the calendar. Another such astronomical fact that has challenged the keepers of the calendar is that the orbital period of the moon does not contain a whole number of days, nor are there a whole number of periods of the moon in one year. However, it happens that after a period of 8 years the moon does return to approximately the same position relative to the earth and sun, that is, to the same lunar phase. This, according to Frazer, accounts for the period of 8 years figuring in certain traditions among some ancient peoples. It is not surprising that the degree to which primitive societies have succeeded in developing a calendar has become a measure of the state of advancement of these societies.

Perhaps the most celebrated of these is the ancient Mayan calendar, a remarkable achievement, often described with such lavish admiration as to convey a sense that this New World culture has surpassed some unspoken expectations. The Maya had in fact two calendars (Morley, 1946): a sacred calendar and a civil calendar with a complicated way of enumerating the days. The sacred year was not divided into groups of days, such as months, but consisted of 260 days enumerated by a number from 1 to 13 followed by one of twenty names. However, curiously, the sequence did not simply run through the numbers from 1 to 13 for each name before running through the numbers again with the next name, which would be tantamount to using 20 "months" of 13 days each. Instead, the name was incremented along with the number in going from one day to the next. After the number 13 was reached, the number sequence was repeated again, incrementing the name at the same time. It would be as if we wrote for a sequence of days 1 Feb., 2 Mar., 3 Apr., etc. It is almost as if the Maya were generating a cryptic code! The Mayan civil calendar was based on groupings of 20 days each, so that there were 18 such "months" and a closing month of 5 days to yield 365 days in a year. If a particular day was specified simultaneously using designations according to both calendars, that specification was repeated every 52 years; that is, within a 52-year span the designation would be unambiguous. For longer periods the Maya developed an enumeration system with base 20, a vigesimal system, which is distinguished in having introduced the *zero* independently of the Old World discovery of that concept. It will be recalled that the place-value system of representing numbers, and therefore arithmetic as we know it, would be impossible without the zero. It is curious that the characters we use to represent the digits, namely what we call Arabic numerals, are not used in the Arabic language; instead, Indian characters are used, in which zero is simply a dot.

The Maya also kept detailed watch on the phases of the moon, and the enumeration of the lunar months played an important part in their elaborate religious calendar. In common with other societies of antiquity, the Maya were in awe of celestial events, which they saw as ominous manifestations in which the mysteries of the universe and the future of human destiny were to be read.

## 1.3 Solar Eclipses as Time Markers

The occurrence of astronomical phenomena such as solar and lunar eclipses, meteors, and comets were recorded with awe from the earliest times and became associated with religious observances or superstitious omens. The seemingly eternal constancy of the motions of the heavenly bodies came to define time and regulate the affairs of many societies, not only in a chronological sense, but also in a mystical astrological sense.

Because of the superstitions that attached to these observations, evidence has been found that records of eclipses reach as far back as 2000 B.C. The sifting of ancient records to discriminate between objectively reported events and those reported spuriously either by accident or by design is a task that has occupied specialists for some time. By now, a large body of data has been compiled from which a chronology of sightings has been constructed, scattered throughout history and over the entire globe.

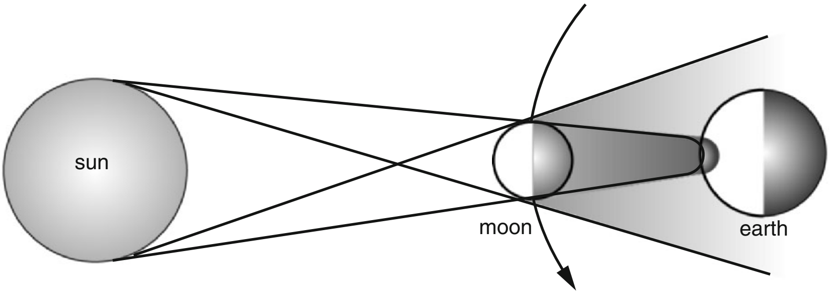


Figure 1.1 The formation of a solar eclipse

If we recall how eclipses are produced, we will be better able to appreciate that recordings of the time and place of their occurrences give sensitive time markers in establishing a long-term chronology. Figure 1.1 depicts the positions of the sun, moon, and earth (not to scale) momentarily along a straight line when a solar eclipse occurs. The three bodies will pass through the aligned condition only when the moon and earth are simultaneously at particular points in their respective orbits. These orbits lie in fixed planes: the plane of the moon's orbit passes through the earth's center, while that of the earth (called the *ecliptic*) passes through the sun's center. These orbital planes are inclined at a constant angle of about 5 degrees; hence there will not be a solar eclipse observed on the earth every lunar month, as would be the case if the orbital planes coincided. However, it can happen that as the earth travels along its orbit around the sun, it will reach a point where the sun and earth are in line with where the moon is just crossing the earth's orbital plane.

As will be recalled from optics, the shadow produced by the moon on the earth consists of regions called the umbra and penumbra, corresponding respectively to total eclipse, in which the complete disc of the sun is obstructed, and partial eclipse, where the moon obstructs only part of the sun's disc. Although the sun is immensely larger in diameter than the moon, it is so much farther away from the earth that to an observer on the earth, the moon's disc can cover the sun's disk that is, both bodies subtend about equal angles at the earth (about 0.5°). Actually, a partial eclipse of the sun will not cause "darkness to fall upon the land" unless it is very nearly total, that is, with over ninety-five percent of the sun's disc obstructed The reason that reliable records of total solar eclipses are such useful time-markers is that they occur only when a very special set of astronomical variables such as the diameters and distances of the three bodies and the positions of the earth and moon in their respective orbits fall within very narrow limits. Moreover, these events, particularly total solar eclipses, are so awe-inspiring and so imprinted themselves on the minds of the ancients that it is not to be expected that many went unnoticed. In fact, the problem is to sift out those sightings spuriously injected into records to lend a supernatural weight to some historical event, such as the death of a king!

It has been argued that when Joshua refers to the sun as having "stopped in the middle of the sky" he may have witnessed a solar eclipse.

A detailed analysis (Schove and Fletcher, 1984) of the chronology of eclipses reported in the historical records, in which times and places of observation are compared to computed predictions, has revealed the remarkable result that the earth's rotation about its axis has been slowing down over the centuries, as has the moon in its orbit around the earth. It has been estimated on the basis of fossil evidence that over geologic time the length of the day has increased from 390 days per year to the present 365.25. If, following Stephenson (Stephenson and Morrison, 1982), we plot the track of totality for a solar eclipse observed at Athens in A.D. 484 and compute a similar plot based on current astronomical data and a constant rotation of the earth, we would find the path 15° west of where it was recorded, corresponding to a difference in time of one hour. Based on more recent precision measurements, about which more will be said in a later chapter, the slowing of the earth's rotation about its axis amounts to about one part in 43,000,000 per century, or about 4.5° of rotation in 1,500 years.

## 1.4 The Tides

If the earth were a uniform hard sphere spinning about its axis in the vacuum of space, it would continue spinning at a constant rate indefinitely. In reality, the earth has topographically complicated bodies of water on its surface, and a molten interior; even the "solid" regions are to a degree plastic. Furthermore, it is not perfectly spherical, having an equatorial bulge with a slight north–south asymmetry. The tidal action in the world's oceans—involving as it does the movement of water, which like most liquids has some viscosity (internal resistance to flow)—can create a drag on the earth's rotation, causing it to slow down. In this process the kinetic energy of the rotational motion of the earth is slowly (and irreversibly) converted to random motion on the molecular scale, that is, heat, in the waters of the oceans.

The predominant cause of the tidal action referred to above is the gravitational pull of the moon, with a smaller contribution from the sun. It arises not so much from the moon drawing towards itself the waters of the earth's oceans by its gravitational pull as from the variation in the gravitational field across the earth that the moon superimposes on top of the smaller variation of the sun's pull. The effect of such a variation can be illustrated by considering what has now become familiar: Astronauts aboard a spacecraft orbiting the earth in a circular orbit. Suppose the spacecraft is equipped with "stabilizing booms," that is, two long straight poles fixed to the spacecraft, one pointing toward the earth and the other in the opposite direction, as shown in Figure 1.2.

As the spacecraft swings quietly along its orbit with all its propulsion systems shut down, the astronauts in the main cabin may float around in a more or less "weightless" state. However, if an astronaut is required to go out to attend to some problem at the ends of the stabilizing booms, then he will notice there that he is

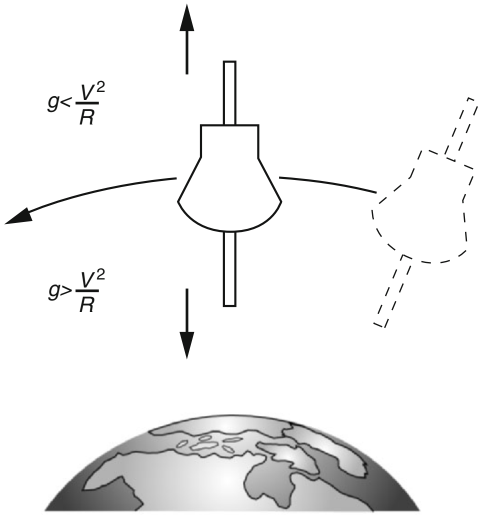


Figure 1.2 Forces acting on an orbiting spacecraft due to variation in the gravitational field across it

no longer weightless. At the end nearest the earth he will have a small but positive weight tending to pull him toward the earth; at the farthest end he will have a small but negative weight, that is, he will have a tendency to be lifted farther out. This means also that the booms themselves experience a stretching force tending to separate the ends. The basic explanation is that for objects in the main cabin there is a balance between the gravitational force of the earth and the dynamical centrifugal force due to the curving trajectory of the spacecraft; a balance that tips in favor of the gravitational force at the end of one boom and the dynamical force at the end of the other. A similar basic argument can be made to explain the fact that tidal motion results in two diametrically opposite bulges in the equilibrium water surface on the earth: one towards the moon and the other away from the moon, as if the system were being stretched along the line joining the centers of the earth and moon. Of course, the actual rise and fall of the tide at any given geographical point is the result of many contributing factors, particularly the topography of the ocean beds, the coast lines, and the resonant response of the tidal motion to the twelve-hour periodic lunar force as the oceans are swept around in the earth's daily rotation.

## 1.5 The Sidereal Day

In specifying the period of rotation of the earth and its variability, a certain frame of reference is of course implied. In our case, the frame of reference is that defined by "the fixed stars." The period of rotation so defined is called the "sidereal day," as contrasted with the "solar day," which is the period between successive transits of the sun across any given meridian (a great circle with a specified longitude). Since the earth sweeps around the sun in a nearly circular orbit while it is spinning around its axis, the time between successive passes of a given meridian under the sun will differ from what the rotation period would be in the absence of the orbital motion. This is made clear by noting that if the earth had zero spin, the sun would still cross a given meridian every time the earth completed a revolution around the sun. The difference in the values of the solar and sidereal days may be easily approximated if we assume a circular orbit. Referring to Figure 1.3, we note that the sense of rotation (whether clockwise or counterclockwise) is the same for the rotation and revolution of the earth. It follows that as the diagram shows, the sidereal day is shorter than the solar day by the time it takes the earth to turn the angle that the sun's direction has turned in one day by virtue of the orbital motion of the earth. This latter angle is (360°/365.25), and the earth rotates at the rate of 360°/(24 × 60) degrees per minute; hence the difference in the length of the two days is $(360/365.25) \times (24 \times 60/360) = 3.95\,\text{min}$ (sidereal).

In making the simplifying assumption that the orbit is circular, we have ignored the fact that the orbit in reality is elliptical. According to one of Kepler's laws of planetary motion, the empirical pillars on which Newton's theory stands, the earth moves with a speed such that the area swept out by a radius drawn from the sun to the earth increases at a constant rate. Since in an elliptical orbit the length of this radial arm varies, going from a minimum at the *perihelion* to a maximum at the *aphelion* (a relatively small change for the earth's orbit), this means that the angle swept out by the radial arm in a given time varies from point to point along the orbit. It follows that the length of the solar day varies throughout the year but is always in the neighborhood of four minutes longer than the sidereal day. This variation of the

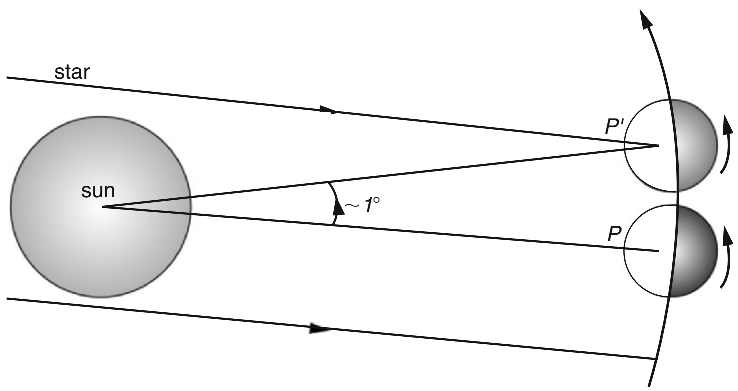


Figure 1.3 The motions of the earth and the difference between the lengths of the sidereal and solar days

solar day must not, of course, be confused with the seasonal variation of daylight hours, which has to do with the inclination of the earth's axis to the ecliptic plane.

## 1.6 The Precession of the Equinoxes

To complicate matters further, the nonspherical shape of the earth (which would be symmetric about its axis of rotation if we ignored tidal action) brings into play a torque, originating from the gravitational pull of the moon, tending to turn the axis of the earth towards a direction perpendicular to the plane of the moon's orbit. To those unfamiliar with gyroscopic motion, the effect of this torque is rather remarkable: instead of simply turning the axis directly from the old direction towards the new, it causes the axis to swing around, tracing out the surface of a geometric cone around the new direction as axis. This motion is familiar to anyone who has watched a spinning top; as a result of the torque tending to make it fall, its axis instead swings around in a vertical cone. This motion is called precession of the axis of spin. In the context of planetary motion, and in particular the earth's motion, this motion causes the points along the orbit of the earth where the seasons of the year occur to shift from year to year along the orbit. The reason for this is that the direction of the earth's axis determines the line of intersection of the earth's equatorial plane with the plane of its orbit. The two points around the orbit where this intersection occurs mark the vernal and autumnal *equinoxes*, which are conventionally taken to be the beginning of spring and autumn. Thus as the axis precesses, the equinoxes will also, and for this reason astronomers call this motion the precession of the equinoxes. Although the rate of precession is small, amounting to about one cycle in 26,000 years, nevertheless it says something about the constancy of the solar day, which you will remember varied from point to point along the earth's orbit.

The precession was first detected by one who is arguably one of the greatest astronomers of antiquity, Hipparchus, in the second century B.C. He made careful measurements of star positions, assigning to each star coordinates analogous to longitude and latitude. By comparing his observations with astronomical records dating back over 150 years before his time, he made the incredible discovery that the point in the night sky about which stars appear to rotate (because of the earth's rotation), that is, what is called the *celestial pole*, had definitely shifted. In view of how small the rate of precession is, amounting to no more than one minute of arc per year, this was no mean accomplishment.

## 1.7 The Sundial

The earliest devices for measuring the elapse of time *within* the span of a day were a natural derivative of the notion of time as being defined by the motions of celestial objects. In order to keep track of the sun's journey across the sky, the shadow clock was devised, which later developed into the sundial. In its most primitive form, it was simply a vertical straight pole, called a *gnomon*, whose shadow is cast upon a horizontal plate marked with lines corresponding to different subdivisions of the day. The principle was applied in many different forms: One of the earliest ancient Egyptian shadow clocks used the shadow of a horizontal bar placed in a north–south direction above a horizontal scale running east–west. These early shadow clocks did not indicate the time in hours, but rather in much larger subdivisions of the day. Through the centuries these clocks evolved into very sophisticated sundials, some of which even were designed to be portable.

The division of the day into 24 hours is traceable back to the ancient Sumerians, who inhabited the land that was known in classical times as Babylonia (Kramer, 1963). Their number system was sexagesimal in character, that is, based on 60, although the separate factors 6 and 10 do occur in combinations such as 6, 10, 60, 600, 3600. They actually had two distinct systems: an everyday mixed system and a pure sexagesimal system used exclusively in mathematical texts. The latter had the elements of a place-value system like our decimal system; however, the Sumerians lacked the concept and notation for zero; furthermore, their way of writing numbers did not indicate the absolute scale; that is, their representation of numbers was unique only to within multiplication by any power of 60. Nevertheless, the impact of the ancient Sumerian culture is evident in the way we subdivide the day into hours, minutes, and seconds; and the circle into degrees, minutes, and seconds of arc.

The Sumerian version of the shadow clock, like those of other ancient cultures, suffered from the same critical flaw: The shadow of a vertical shaft moves at a variable rate over the span of a day, and what is worse, the variability itself changes with the seasons and the latitude where the device is used. It would require a sophisticated knowledge of celestial mechanics to derive corrections to the observed readings for each day of the year and for different latitudes.

A radical improvement in the design of what came to be called sundials was made by Arab astronomers in the Middle Ages. This consisted in mounting the gnomon (which you will recall is the name given to the object producing the shadow) as nearly parallel as possible to the axis of rotation of the earth, that is, pointing toward the celestial pole, which is currently within 1° of the North Star (Polaris). This revolutionary change in design transformed the sundial into a serious instrument for the measurement of time. In order to appreciate the significance of this innovation, let us recall that the apparent daily motion of the sun, and indeed all celestial objects, is due simply to the earth's rotation about its axis; and therefore, to the extent that the earth's rotational motion is uniform, the sun's apparent angular position around that axis will also progress uniformly. It follows that the shadow cast by a shaft parallel to the axis onto a plane perpendicular to it will have an angular position that follows the sun, and it will therefore reproduce the rotation of the earth. Of course, it is only during about half of every rotation that the sun's rays will reach a given point on the earth's surface; however, unlike the shadow clock, the seasonal variation in the relative lengths of daytime and nighttime will in this case have no effect. Since the earth's rotation is very nearly constant, a circle drawn on the plane with the gnomon as center can be divided into 24 equal parts corresponding to the 24 hours of the day. In practical portable sundials, such as might have been used aboard ship at the time of Sir Francis Drake, provision must be made for setting the direction of the gnomon. This they could achieve by finding north with a magnetic compass and the latitude by means of a forerunner of the sextant.

The accuracy one can achieve with this type of sundial, while incomparably greater than the earlier primitive versions, nevertheless is limited by the extreme accuracy with which angular displacements of the shadow would have to be measured. Thus the shadow moves only 0.25° per minute; this implies that an error of 0.1° in angle measurement translates into an error of 24 seconds. This level of accuracy, though unimpressive by more recent standards, coupled with the fact that it provided an absolute reference with which to compare mechanical clocks, ensured the continued use of the sundial, in one form or another, from antiquity until the Enlightenment.

## 1.8 The Astrolabe

In this context we should include another astronomical instrument of ancient origin, also perfected by medieval Arab astronomers and instrument makers, called the astrolabe shown in Figure 1.4 (Priestley, 1964). It is a combination of an observational instrument and a computational aid enabling the determination of not only latitude, but also the time of day. It ultimately spread to Western Europe and was there in common use by navigators until the advent of the sextant in the eighteenth century.

The astrolabe consisted of a disk on whose rim was engraved a uniform scale with 24 divisions, surmounted by a plate engraved with a projection of the celestial sphere over which arcs of circles were inscribed. Pivoted concentrically were also a metal cutout star pattern called a *rete* and a metal pointer called the *rule*. On the back were concentric scales graduated in degrees, the signs of the zodiac, and the calendar months; another pointer was pivoted at the center. It would be out of place, and probably well beyond the interest of the reader, to devote much space to describing the intricacies of this instrument and how to get the most information out of it. Briefly, it may be said that it is assumed that the calendar month and day are known for the time it is to be used. The altitude of the sun is first observed using the degree scale on the back of the instrument while sighting the sun. From that side also one reads, for that date, the position of the sun along the zodiac. Using this information on the front side of the instrument, the star pattern is turned with respect to the projection of the celestial sphere until the sun's position in the zodiac agrees with its observed altitude. The time is read by appropriately setting one end of the rule and reading the other end on the 24 division scale. It is interesting that Geoffrey Chaucer, of *Canterbury Tales* fame, also wrote a *Treatise on the Astrolabe*

Figure 1.4 The astrolabe as depicted in Chaucer's *Treatise on the Astrolabe* (a) front side (b) back side

in 1391. Since it gave both latitude and time of day, the astrolabe was used by navigators well into the eighteenth century.

## 1.9 Water Clocks

Among the earliest non-astronomical devices for measuring time was the water clock, of which rudimentary examples have been found among ancient Egyptian artifacts dating back to 2000 B.C. It was essentially a conical stone vessel filled with water that escaped slowly through a small hole at the bottom provided for that purpose. A uniform scale was marked along the side of the vessel to enable the elapsed time to be gauged by how far the water level had fallen. The ancients were led through experience to the need for a conical shape in order to achieve an approximately uniform scale. It would have been fairly obvious that the level in a straight cylinder falls faster when nearly full than when nearly empty. However, that the shape should be conical, rather than spherical or some other shape, is less obvious and must have been arrived at rather through convention than careful observation. The hydrodynamic problem that the design presents is actually a fairly complicated one; a lot depends on whether or not the water passes through a channel-like opening, in which case the viscosity of the water plays an important role, making the rate of flow of water more or less proportional to the pressure, and therefore to the height of the surface above the opening. On the other hand, if the opening is such that the effect of viscosity is negligible, then we have ideal conditions where *Bernoulli's principle* should apply and the *kinetic energy* (that is, the *square* of the velocity) of the escaping water should be determined by the pressure, or depth below the surface.

Since the pressure in the water at the depth of the opening is proportional to the depth, irrespective of the shape of the container, it follows that if the flow rate is proportional to pressure, a constant rate of fall of the water level will be achieved if the area of the surface of the water is proportional to the depth of the opening. This is ideally satisfied by a cylinder whose axial cross section is approximately parabolic.

Rather than attempt to perfect the shape of the container, which we now appreciate is a good deal more complicated than would at first appear, the actual development of water clocks took a much more promising tack in achieving a constant rate of flow by providing, in the words of the plumbing profession, a "constant head," that is, a fixed water level above the hole. This advance is attributed to an Alexandrian by the name of Ktesibios (also given the Latinized spelling Ctesibius), a celebrated inventor of Ptolemaic Alexandria around 250 B.C. (de Camp, 1960). His accomplishments included other mechanical and hydraulic devices, such as a water pump and pipe organ. From his work evolved the Hellenistic type of water clock, called *clepsydra*, that was in common use throughout classical times. Such clocks were commonly used then to allot time to speakers in a debate: When the water ran out, it was time to stop. Successive speakers were assigned the first water, second water, and so on. This may have something to do with the expression "of the first water" as something of the finest quality. The essential design is illustrated schematically in Figure 1.5.

The constant pressure head is achieved simply by allowing an adequate continuous flow from some source into the vessel and preventing the level from rising above a fixed point by having an overflow outlet. The constant flow was collected in a graduated straight cylinder. The design often incorporated various time-display mechanisms that were actuated by the constant rise in the water level. In one instance a float supported a straight ratchet engaging a toothed wheel; to this was attached a pointer to indicate the time on a circular dial. In other designs a pointer was joined to the float by a vertical shaft, enabling the rise in the float to be read on a vertical scale drawn on the surface of a rotatable drum. By varying the scale at different points around the periphery of the drum, it was possible to accommodate the seasonal variations in the length of the hour, which was then defined as a certain fraction of the period from sunrise to sunset.

In China, water clocks are known to have existed at least from the sixth century of the Christian era; but their development took a more elaborate mechanical turn. In place of the time being measured by the continuous motion of a simple float rising in a linear fashion, the Chinese took things to a higher level of sophistication by introducing the idea of using the flow of water to control the rate of turning of a water wheel; not continuously, but in discrete steps, much like the crown wheel in the tower clock escapement mechanism, to be discussed in the next section. The water flowed at a constant rate into successive buckets mounted on short swivel

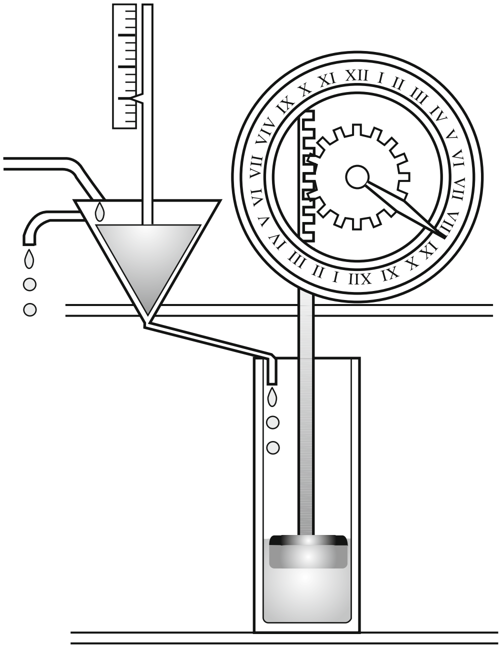


Figure 1.5 Schematic drawing of an ancient Greek water clock

arms between numerous equally spaced spokes of a wheel, free to turn in a vertical plane about a fixed axle. By a clever arrangement of balanced beams, levers, and connecting rods, the rotation of the water wheel was automatically stopped by blocking one of the spokes while a predetermined amount of water flowed into each bucket in succession. When the critical amount of water had been reached, the bucket arm was able to tilt against an accurate counterweight at the other end of a balance beam, in effect "weighing the contents of the bucket" before allowing the wheel to turn until the next spoke was engaged and the wheel stopped again for the next bucket to fill.

The accuracy achieved in a well-constructed clepsydra was comparable to the sundial, but of course their time scale was not absolute, in the sense that they had to be calibrated against a scale based on astronomical observations. A serious limitation of the water clock is its obvious nonportability; it is difficult to imagine how it could be made suitable for use aboard a ship on the high seas.

Another device that we should mention based on the flow of material through a small hole is, of course, the *hourglass*, the universal symbol of the fleeting nature of time. Sand was not the only substance used; the choice was directed towards greater reproducibility. Many granular solids are efficient absorbers of moisture, a property that clearly disqualifies them, since they would have a greater tendency to form clusters. The rate of flow through a constriction clearly depends on the grain or cluster size as well as friction between grains. Hourglasses provided a convenient way of determining a fixed interval of time, and sets of hourglasses were used aboard ships to mark the *watch*, the 4-hour spell of duty.

## 1.10 Tower Clocks

A great deal has been written about mechanical clocks and clockmaking, a testament to the enduring fascination with which the subject has been regarded through the ages. We will limit our discussion of this subject to a review of their design and performance from the vantage point of present-day horology.

In the analysis of the operation of mechanical clocks it is useful to separate the mechanism into three essential functional parts: first, the energy source, which has usually taken the form of a falling weight or a coiled spring; second, a mechanical system capable of inherently stable periodic motion to serve as regulator; third, a mechanism to derive and display the time in the desired units. The third part consists of gear trains and a dial. Of these the most critical and challenging is the regulating system, and the history of the advancement in mechanical clockmaking is the history of the development of this part of clock design.

The fundamental problems in regulator design reduce to two in number: first, an oscillatory system must be found whose period of oscillation is constant and insensitive to changes in the physical environment and operating conditions; second, a method must be found for the regulator to control the transmission of power from the energy source to the rest of the clockwork with the least possible reaction on the regulator. Some interaction is essential to sustain the oscillation of the regulator. However, this must be small compared to the oscillation energy of the reference system of the regulator. As with any mechanical system, there will always be frictional forces present, which in the absence of an adequate source of excitation energy will cause the oscillator to come to rest. Therefore, a small driving force must act on it in step with the motion to maintain a nearly constant level of excitation. But this requirement runs counter to the function of the regulator as a controller: Rather than controlling, it is being controlled. The ideal situation would be one in which the reference oscillator was free to execute its natural oscillations without any external perturbations acting on it.

To reconcile these opposing requirements and achieve the best possible outcome requires the following: First, the oscillator must have very low inherent friction, enabling it to continue to execute its motion with only a very weak driving force; and second, it must exert its control in a "trigger" fashion. This means that a small force exerted by the controlling element, acting for only a small fraction of the period of oscillation, must control a much larger force transmitted from the energy source to the gear train driving the clockwork. The mechanical means of achieving this is called the *escapement*.

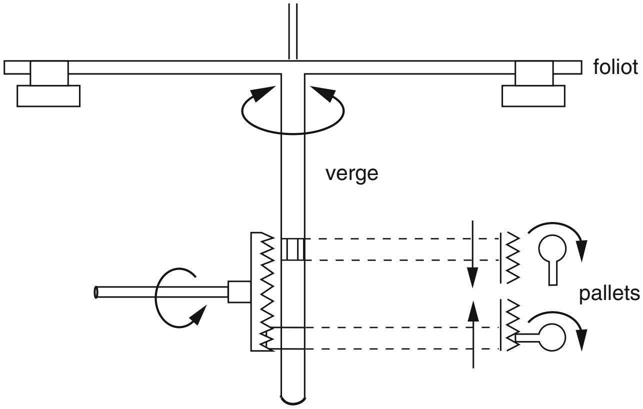


Figure 1.6 The foliot and verge escapement for tower clocks

An early version extant in the fourteenth century was widely used in tower clocks for cathedrals, public squares, etc. for almost three centuries. It is called the *verge* and *foliot* escapement and is shown in its basic form in Figure 1.6. A straight horizontal beam, the foliot, with equal weights balanced at its ends, is suspended at its middle. Rigidly attached to the foliot at its point of suspension is a vertical spindle, called the verge, to which are rigidly attached two small flat projections, called *pallets*; these engage at diametrically opposite points a vertical wheel (the scape or crown wheel) with cogs perpendicular to its face. The planes of the pallets are parallel to the axis of the verge, but are typically ninety degrees apart. This balanced foliot–verge system is capable of simple periodic angular motion about the verge as axis. The torsion in the suspension of the foliot provides the necessary restoring torque when the foliot is turned away from its equilibrium position. The action of the pallets as the foliot rotates back and forth is to momentarily block the cogwheel alternately by one pallet, then the other. The rotation of the cogwheel, which derives its torque from the energy source, is thereby regulated. The reaction back on the oscillating foliot occurs at the contact between the pallets and the cogs. As earlier pointed out, this reaction tends to sustain the oscillation of the foliot. If the suspension material is chosen to have good elastic properties with a particularly low internal friction and if the foliot is massive to increase the energy and the period of oscillation, this regulator can be expected to be relatively stable and insensitive to small perturbations, such as air drafts, noise, and vibration. In judging these mechanical clocks we should separate the principles upon which the design is based from the implementation of those principles, that is, the choice of materials and the level of precision in the manufacture of the clocks. If we consider the operating principles of the clocks we have been describing, we note that the foliot–verge system is really a type of torsion pendulum, and as such is capable of as great a constancy of oscillation period as are later developments, for example the pendulum. Its limitations arise principally from the design of its escapement; the force of reaction is too large and acts for too large a fraction of the period. The choice of material for the suspension is also critical; a fused quartz fiber suspension would have excellent elastic properties. Such quartz suspensions have been widely used in torsion balances since the seventeenth century, because of both their strength and elastic properties; the restoring torque they provide is linearly dependent on the rotation angle.

## 1.11 The Pendulum Clock

Two important advances were made in the seventeenth century: First came the pendulum as the regulator, and then, of equal importance, came the "deadbeat" anchor escapement. Let us consider these in turn.

The story of Galileo's timing the swings of a chandelier at the cathedral in Pisa using his pulse is well known. The discovery of the "isochronism" of the pendulum, that is, taking an equal time to complete a swing no matter how widely it swings, dates from 1583 when Galileo was a medical student. The story is usually repeated as an example of extraordinary resourcefulness in his desire to study the pendulum. This may be so, but it should be noted that he was at the time also interested in medicine and in particular the pulse rate as an indicator of fever (Drake, 1967). In fact, there is no published account by him at this time suggesting the use of the pendulum as a regulator for mechanical clocks. He did, however, use it to construct an instrument to conveniently measure the pulse rate in patients. It consisted essentially of a pendulum with variable length, which was adjusted to match the pulse rate of the patient. It was calibrated to read directly conditions such as "slow" or "feverish." It was not until a few months before his death, in 1642, that Galileo suggested the application of the pendulum to clocks. He had become blind in 1638 and was no longer able to put his idea into practice. He dictated a design to his son Vincenzo, who made drawings but did not actually complete a working model. The credit for actually incorporating a pendulum into the design of a clock around 1656 goes to Huygens, a name associated in the mind of every physics student with the wave theory of light.

The pendulum is essentially an object pivoted or suspended so that it swings freely. For purposes of analysis, we distinguish between a simple pendulum, which consists of a small object suspended by a thin string of negligible mass, and a compound pendulum, in which the mass distribution along the pendulum is not negligible. The essential characteristic of the pendulum, as Galileo noted, is that for small swings the period of oscillation, that is, the time to complete a swing in one direction and back to the starting point, is the same no matter how wide the swing, provided that it remains small. This property caught Galileo's attention because it appears to run counter to what might superficially be expected: After all, with a large swing, the pendulum bob has farther to travel, and it is indeed remarkable that its speed varies in just such a way that the oscillation period is always the same.

Actually, the same could be said of a beam suspended by a material with suitable elastic properties, such as fused quartz. They both display *simple harmonic motion*. But while the latter depends on the property of the suspension material, the pendulum has no such dependence on materials, which are generally subject to variation. However, the period of the pendulum does depend on its *radius of gyration*, which depends on the distribution of mass along the length of the pendulum. Since all materials expand and contract with the rise and fall of temperature, the constancy of the period is limited by any fluctuations in the temperature. We may attempt to overcome this limitation by taking one or all of the following steps: Choose a material that has extraordinarily low thermal expansion, such as the alloy *invar*; regulate the temperature to reduce its fluctuations; and use a composite pendulum incorporating two materials of differing expansion coefficients, such as brass and steel, in such a way that the expansion of one is compensated by the other.

Another important limitation of the pendulum as a reference oscillator is that its period depends on the strength of gravity, which varies from point to point on the earth's surface. This is because the earth is neither spherical nor homogeneous. As far back as 1672, it was established through pendulum measurements that the acceleration due to gravity is different for different geographical locations. This was explained by Newton by assuming a model of the earth as a uniform gravitating plastic body, which, by virtue of its spin, would bulge around the equator into an oblate spheroid. The value of the fractional difference between the earth's radii at the equator and poles was later computed in 1737 by the Frenchman Clairaut to be 1 in 299. The acceleration due to gravity is also dependent on altitude in a way that may be affected by local topography and geology. The differences in the times indicated by a pendulum clock at different geographical locations could be on the order of one minute per day. Another source of fluctuation in the period of a pendulum is air resistance, whose drag on the swinging pendulum depends on the density of the air and thus the atmospheric pressure.

The other major development, which came around 1670, was a much improved escapement: the *anchor* escapement, and later the *deadbeat anchor* introduced by Graham in 1715. Figure 1.7 shows the essential design.

Unlike the verge–foliot escapement, where the pallets engage diametrically opposite points on the scape wheel, the anchor escapement acts on a sector of a rachet wheel having radial teeth. This geometric difference allows the pallets to be separated by some distance along the rim of the scape wheel, and in consequence a smaller angular movement of the anchor about its axis is needed to engage and disengage the pallets and the wheel. This is advantageous to the performance of any regulator based on a mechanical oscillator, since its oscillation is simple harmonic only in the limit of small oscillations. But an even more important difference to note is that the pallets move at right angles to the direction of motion of the scape wheel teeth. This means that the force of interaction between the pallets and the teeth has little torque around the axis of the pallet mount and is therefore ineffective in disturbing the oscillation of the pendulum. Moreover, the pendulum is entirely free of the scape wheel for part of its oscillation, a first step toward

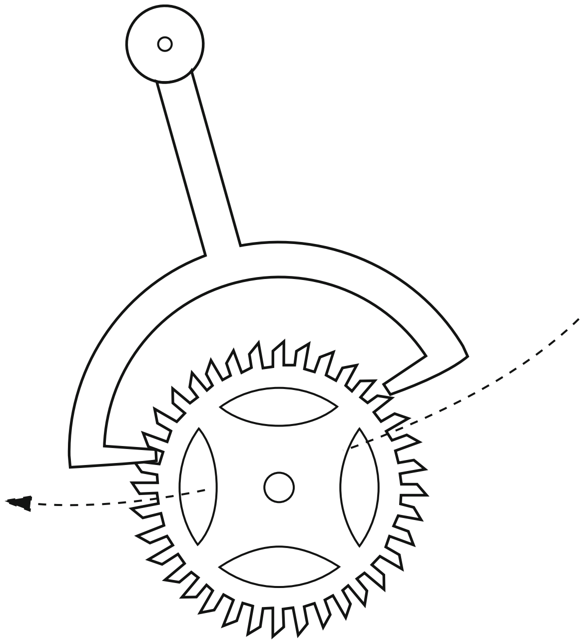

Figure 1.7 The anchor escapement

the ideal condition. The deadbeat design is so called because unlike the anchor escapement just described, there is no recoil of the scape wheel and the gear train behind it during a swing, or beat, of the pendulum. This was achieved by a careful contouring of the faces of the pallets and the teeth of the scape wheel. This refinement further improved the isolation of the pendulum and enhanced its performance as a regulator.

## 1.12 The Spring–Balance-Wheel Clock

For fixed installations, such as in observatories or clock towers, the pendulum-controlled clock became the most widely used, and by the end of the eighteenth century it had reached an accuracy sufficient to the demands of the day. However, there remained one area of need that the pendulum clock could not satisfy: shipboard and, in today's jargon, other mobile environments, where the clock may be subjected to erratic inertial forces. Moreover, any attempt to scale down the size of the clock to make it more portable would aggravate the problems of air resistance, friction, and curvature in the knife edge on which the pendulum is pivoted. A further disqualification for shipboard use is the variability of the period with geographical location, as previously described.

Although brave attempts were made to develop a pendulum clock that would be reliable in the field, it finally became clear that a new approach was required. This came in the form of the balance wheel and the spiral hairspring, which ultimately became universally used in all mechanical watches. The hairspring, or balance spring, was a spiral of fine resilient metal fixed at the outer end to the body of the watch and at the center of the spiral to the arbor of the balance wheel, which is delicately pivoted on jeweled bearings to reduce the rate of wear. The basic advantage of the spiral spring is that it provides a restoring torque on the balance wheel independent of gravity and permits a reduction in the strain (degree of internal deformation) in the material of the spring for a given rotation of the balance wheel. This is important if the restoring torque is to remain proportional to the angle of rotation of the balance wheel and lead to simple harmonic motion.

With almost every important advance in the world of ideas or in the practical world of devices one name has, through common usage, become associated as the one to whom all credit is due. However, we all know that almost always there were other thinkers and inventors who had made critical contributions to those advances. When the contest for recognition is between two equally prominent personalities, the controversy is resolved, if at all, along national lines. In the present instance there is no doubt that the Englishman Robert Hooke, of *Hooke's law* fame, had indeed proposed a spring–balance mechanism sufficiently accurate to determine longitude at sea. It seems that Hooke had ambitions of exploiting his ideas in an entrepreneurial spirit, not commonly avowed by physicists of his day. In any event, Hooke failed to form a syndicate, and he never actually "reduced his ideas to practice," as patent lawyers would say. On the other hand, Huygens, already credited with implementing the use of the pendulum as a regulator, did in fact have a clock constructed that was regulated by a balance wheel.

Much has been written about the British Admiralty's quest in the early eighteenth century to simplify the solution of an age-old problem in navigation: the determination of longitude at sea (Sobel, 1995). Unlike latitude, which can be deduced from straightforward observations, such as the altitude of the sun on the meridian (at noon) or the altitude of the star Polaris, longitude was computed by a rather complicated procedure devised by the astronomer Edmund Halley, better known for his comet. It had been recognized for some time that if a mariner at sea had a precise clock indicating Greenwich Mean Time (GMT), he could determine longitude by using it to find the time of local noon, for example. To spur interest in the development of such a shipboard clock, the British Admiralty established a Board of Longitude in 1714 that offered a reward of £20,000 to anyone who could determine longitude at sea with an error less than thirty miles. At the equator this corresponds to an error of about 1.7 minutes in time. For a voyage lasting one month this implies an error less than about 4 parts in $10^5$, beyond the capability of the existing clocks under shipboard conditions. A Yorkshireman named John Harrison, a far more gifted instrument maker than politician, perfected his first chronometer by 1735, an intricate piece of ingenious mechanical design to minimize friction, etc. Sadly, because of the novelty of his ideas and a prejudice in favor of the astronomical technique called *lunars*, the Board of Longitude was not much impressed and denied Harrison the award. Not until 1761, after his chronometers had been generally admitted to have merit, was the Admiralty willing to try one out on a voyage to Jamaica, during which it lost less than 2 minutes at a fairly constant rate. It was Captain Cook, another Yorkshireman, who by carrying Harrison's timepieces on his long voyages demonstrated finally that old Harrison's chronometers had indeed met the Admiralty requirements and had fully deserved the award.

The principal problem with the balance wheel is its susceptibility to thermal changes in dimensions and consequent changes in the period of oscillation. As with the pendulum there are three remedies; of these the most universal is compensation of expansions and contractions due to temperature fluctuations through the use of two metals in the form of a *bimetallic strip*.

The ultimate success of the balance wheel as a regulator in precision mechanical clocks was made possible by further progress in the design of the escapement, culminating in the *détente*, or chronometer spring, escapement. This brought the performance level to a height unmatched until the arrival of electronic timekeeping. This escapement approximates more closely than any other the ideal of allowing the regulator to oscillate freely except for a very short period of interaction with the scape wheel.

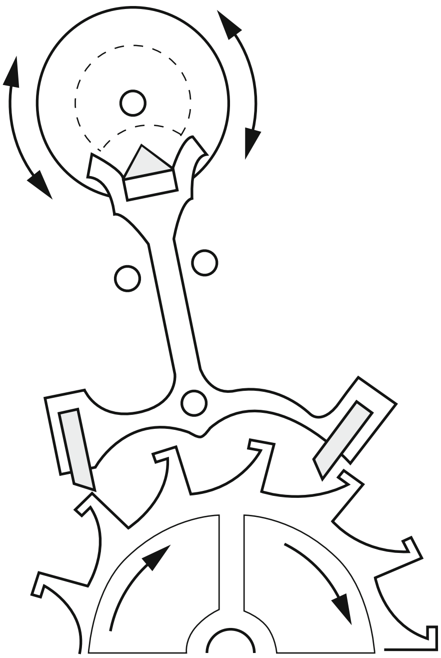

Figure 1.8 A typical escapement design in a high quality mechanical wristwatch

Over the succeeding centuries timepieces were progressively refined and made smaller; from the pocket watch to the dainty ladies' wristwatch. Figure 1.8 illustrates schematically the essential features of the escapement commonly used in high-quality wrist watches. Where it took Harrison literally years to painstakingly construct by hand a single clock, it ultimately became possible to mass produce them, thus making them universally affordable. But it is remarkable that from the point of view of accuracy, no purely mechanical clock has surpassed some of Harrison's later chronometers.