**Time Measurement for Flamsteed Astronomy Society**

When Bobby sent an e-mail asking for contributions to the History of Astronomy Group, I offered this talk which I was at that time preparing for the Institute of Physics. This subject covers both physics and astronomy, and I offered to turn down the physics and put a bit more detail in the astronomy if he thought it could be useful. He accepted, so here we are. There is a lot of history in it, though.

Well I have no idea what “time” is, and I am not going to attempt an explanation. Neither am I going to consider the relativistic effects on time of speed, acceleration or gravity. Everything I cover will be limited to one single inertial reference frame.I will take it that time is one of the four dimensions of relativistic space-time, although it clearly isn’t like the three space dimensions. But as a dimension, it can be measured and to measure something requires units of measurement.

In the case of time, the unit would be an interval of time which is reproducible. There are three natural periodic phenomena which everybody is familiar with, the day (rotational period of the Earth), the month (orbital period of the moon) and the year (orbital period of the Earth). Unfortunately they are not integer related, which gives calendar makers a problem. The shortest and most prominent interval is, of course, the day, and it is common to every culture and civilisation. So it differs in this way from the traditional units of length and mass, which were arbitrary and vary in different civilisations according, for example, to the length of a kings forearm or the mass of a particular stone or lump of metal. Everybody can agree the day in order to realise a unit of time.

Intervals and times less than a whole day can be estimated from the position of the Sun in the sky or measured with a sundial. Unfortunately this is only useful on clear days. Nighttime and cloudy skies need some other way. The first attempts to define a shorter and more convenient unit used the divisor 12, probably copying the 12 full moons in one year. In ancient China, the period from one sunrise to the next was divided into 12 “hours”, while the ancient Middle East cultures adopted 12 hours between sunrise and sunset, and 12 hours between sunset and sunrise. Of course this meant that the “hours” during the day differed from those during the day, and varied with the seasons. Primitive devices like water clocks and sand glasses were used to try to get equal hours, but not very successfully.

Intervals shorter than an hour were defined by Claudius Ptolemy who took a Babylonian system of division into 60 parts which had been adopted by the Greek astronomer Hipparchus for angular measure. Ptolemy divided the hour into 60 “partes minutae primae” and these in turn into 60 “partes minutae secondae”. - our minutes and seconds. Normal daily life didn’t need this accuracy, though, and people continued to be satisfied with half and quarter hours, even after the invention of mechanical clocks about the 10th Century.

There was an attempt in revolutionary France to use decimal division in all measurements, including time. The National Convention in 1793 officially adopted a 10-hour day, with 100 decimal minutes in each decimal hour and 100 decimal seconds in each decimal minute. It didn’t take off, though, and was abandoned two years later, ostensibly because of the cost of replacing all the nation’s clocks and because decimalisation of time would help only astronomers, not ordinary citizens. Ironically, it was repealed in the same law which introduced the metre as the unit of length.

John Flamsteed in 1677 was the first to accurately quantify the variation of the length of a day with the seasons, using the recently invented pendulum clock. He referred to it as “the equation of natural days” and used it to define “mean time”, constant and avoiding seasonal variations. You can often find this on sundials, usually in the form of a table of figures to correct the time indicated by the Sun’s shadow. But even if clocks were becoming more precise, they were not defining a unit of time. This fundamentally remained the day, and a clock had to be checked, and if necessary corrected, to observations of the Sun or stars. Accuracy was best achieved with the transit of stars, rather than the Sun, and the sidereal day had to be corrected to get the solar day (A sidereal day is 23 hours, 56 minutes, 4.0916 seconds). Clocks of this precision required accurate observations with telescopes rather than simple sundials and were carefully mounted at astronomical observatories.

The need for accurate time measurement was greatest in navigation at sea, where time difference can be used to find differences in longitude. This was effectively solved in 1766 with the publication of Maskelyne’s Nautical Almanac.

Using a sextant to measure the angular position of the Moon from identifiable stars, the Almanac enabled the navigator, wherever he might be in the world, to find what time it was at the observatory where the measurements for the Almanac had been made. Then, knowing local time at the position of the ship, he could calculate the longitude as 15° per hour of time difference. As Astronomer Royal, Maskelyne, of course, was making all these measurements at his observatory here in Greenwich, so this became the datum point and before long navigators on ships of every nationality were using the British Nautical Almanac to find their longitude east or west of the meridian of Greenwich.

This is the main reason why, at an international conference in 1884, the meridian through the centre of the transit telescope at Greenwich was recognised as the Prime Meridian and mean solar time on that meridian became Universal Time for the whole world.

Up until the middle of the 19th Century, though, ordinary life still reckoned time and set their clocks to the sundial. This meant that places separated east-west were running on different times. It didn’t matter at all to people who were only able to travel slowly by foot or horse, but when the railways arrived from about 1830 onwards they found a problem with timetabling.

A telegraph connection had been laid in 1851 from the Observatory at Greenwich to the railway at Lewisham as part of a project to connect the with the Paris Observatory. As the telegraph lines already existed throughout the railway network as a necessary part of running the service, an hourly signal originating at Greenwich could be received to regulate the clocks on display at all the principal railway stations in the country. “Railway Time” as it was called, soon gained popular acceptance as it was free from the variables which had previously existed. By 1855, 98% of public clocks were being set to GMT

But there were still ambiguities between local time and railway time. One of the main reasons for establishing a uniform time standard for the nation was the Licensing Act of 1872. This required by law, that public beer houses in towns had to close by midnight, and in country areas by 11 pm. Beer halls in the west of England didn’t want to close just because it was midnight in London, they wanted to keep on serving customers until it was midnight local time where they were - which could be 20 minutes or more later.

To sort out this and other problems, Parliament passed the Definition of Time Statute in August 1880 requiring that any legal expression of time referred to GMT in Great Britain and Dublin.

*In 1769, King George III had set up his own observatory in the grounds of his palace at Kew in order to observe the transit of Venus across the face of the Sun on 3 June that year. The observatory continued to be used, making observations of meridian crossings there. This meridian defined “Kings Time” which was adopted by Parliament and used to set all the clocks at the Palace of Westminster. The meridian is marked on the path at the River Thames today, close to Richmond Lock.*

*By my reckoning, the longitude of the Observatory is 0° 18’ 47.4” west of Greenwich, so the time difference between the two meridians would be 1 min 15 sec - not enough to be significant in those days. It was, of course, superseded in law by the Statute of 1880.*

After GMT had been accepted as Universal Time (UT) at the 1884 Washington Conference, all the other countries gradually signed up to it and adopted civil time through a series of time zones, which remain substantially the same today. The international conference on the metre in Paris in 1875 had concentrated on standardising the metre and the kilogram, and left the time unit, the second, to the astronomers. It was defined simply as 1/86400 part of a mean solar day - a definition which I remember having to learn by heart when I started doing physics at school in the early 1950s.

Early in the twentieth century radio time signals began to be broadcast by different nations all over the world to help navigators, but it was found that they could differ from each other by several seconds. France, which had kept its Bureau des Longitudes in existence long after the British had disbanded their Board of Longitude, undertook to coordinate these signals internationally, and the BIH was established within the Paris Observatory in 1912

By then it was known that electrical circuits could produce oscillations, as well as mechanical systems like pendulums and balance wheels, and that electronic oscillators had the potential for much higher frequency and greater precision. Using the piezo electric effect in a quartz crystal, the vibrational resonance was synchronised to an electrical oscillator and by the mid 1930s was achieving close to 1 ms per day. But these clocks still weren’t defining time: The clock still had to be calibrated to the astronomically observed rotation of the Earth just like the pendulum clocks.

The precision of quartz clocks was now good enough to show that the Earth rotation rate is variable, not constant as had always been assumed.

There are three types of variation in the rotation speed: secular, irregular and periodic.

The first , due to tidal friction, is a long term gradual slowing reckoned to be an increase in the length of the day by about 1.7 ms every 100 years over the last few centuries.

Irregular variation refers to changes from the average sped of rotation which continue for the order of 5 to 10 years. They may have an amplitude of 3 to 5 ms in the length of a day and can accumulate to 1 second or more of difference from clock time. These variations are thought to be due to tectonic changes in the mass distribution within the Earth.

Seasonal variations, probably due to the atmosphere winds and ocean currents.

But if the length of the day is varying, then the length of a second - defined as 1/86400 part of a mean solar day - must be varying too.. In 1927 Andre Danjon of the Paris Observatory proposed that time could be based on the motions of the planets rather than the rotation of the Earth, and that this would be more regular. This idea was taken up again 1948 when Gerald Clemence of the USNO proposed that time should be based on the position of the Earth in its orbit round Sun using the tables of Simon Newcomb which had been published in 1895. Clemence made it clear that his proposal was intended "for the convenience of astronomers and other scientists only" and that it was "logical to continue the use of mean solar time for civil purposes"

Simon Newcomb was born in Canada in 1835 and was a self-taught mathematician and polymath. As director of the USNO he was responsible for preparing the US Nautical Almanac each year. Now Maskelyne’s original Nautical Almanac of 1766 had been based on tables predicting the position of the Moon by Mayer, and since then such tables had been improved on by the observations and calculations of astronomers and mathematicians in several different countries. Newcomb was using the best tables of the Moon available at the time, due to Peter Andreas Hansen, which had been published in 1857

Newcomb found there were still significant errors in these tables and that in addition to the recent data used by Hansen, there were a lot of observations and measurements dating back to 1675 which Hansen had not used. Newcomb set out to find predictive equations for the positions of the planets fitting to all the data, and using the best calculation techniques was able to prepare ephemeris tables for the Sun and the four inner planets.

The Tables of the Sun listed the position of the Sun as a function of time. Clemence re-scaled the equations so that they gave the time as a function of the planet positions rather than the other way round and the year became the observed repeat for time measurement rather than the day.

*Just as an aside, it isn’t recognised today how important such tables were in the past. Now you can get triganometric and other mathematical functions instantly from your computer, pocket calculator or even mobile telephone. But even in the 1960s and 70s physicists needed their 6-figure log tables with sine, cosine and tangent functions listed as tables in order to do numerical calculations.*

In 1952 the IAU adopted Newcomb’s formulae and tables for astronomical purposes, and, together with the CIPM, set up a committee to look for solutions to the time problem. This committee made the following recommendations based on Newcomb’s tables.

*The mean longitude of the Sun should be defined by a polynomial in time, measured in Julian centuries from January 0, 1900, 12h UT (the epoch).*

*And the second should be defined as 1/31 556 925.975 of the length of the tropical year 1900*

This definition of Ephemeris Time (ET) was approved by the IAU in 1955 and recognised by the CGPM in 1960.

It was soon realised that although ET was defined from the longitude of the Sun, in practice it could be measured much more precisely from the indirect comparisons of the position of the Moon with lunar ephemerides. So specific versions of ET were determined from observations of the Moon, which could be used as secondary standards to allow a more precise timing of events than ET, but which might diverge from ET in the long term.

When the unit of time, the second, was defined by the rotation of the Earth, it had been possible to find UT to a few milliseconds by observation of stars crossing the meridian. With ET, observations over several years followed by adjustment calculations were needed to get the same accuracy.

Atomic Clocks

After the Second World War various attempts were made to use atomic and molecular resonances as frequency standards. The main problem was to find an atomic or molecular resonant frequency that was low enough to be measured electronically. Ammonia was tried in the USA, without success.

Working at the National Physical Laboratory in Teddington in 1955, Louis Essen and Jack Parry put together the first caesium beam frequency standard.

This used the hyperfine transition in caesium locked to a microwave signal. Of all the materials which might have been used, caesium has a distinct set of advantages:-

The only stable isotope is caesium 133 which, with an odd number of nucleons, has a nuclear magnetic moment.

As an alkali metal, caesium has a single electron in the outer orbital.

It has a low melting point, 28.5°C, so low thermal motion of the vapour, hence low Doppler spreading.

Essen and Parry used it as a frequency standard and locked it to a quartz crystal clock. They measured the frequency with respect to UT from the Greenwich Time Signal and got a value 9,192,631,830 ± 10 Hz

Now as a frequency standard, this was breaking entirely new ground. Up until then the frequency of the oscillators depended on physical parameters (the length of a pendulum or the size and cut of a quartz crystal) and had to be adjusted to match the rotation rate of the Earth, which fundamentally defined the ‘day’, and hence the second, even though this had been shown to be variable itself. The caesium clock gave a frequency which depended only on fundamental quantum constants, and would be exactly the same wherever and whenever it was measured. The only uncertainty was in how precisely it could be measured. The value of this as a time/frequency standard was recognised immediately, and the Director of NPL, Sir Edward Bullard, wrote to the IAU who were meeting in Dublin in 1955 suggesting that they hold off implementing Ephemeris Time and wait for more laboratories to build and get familiar with the caesium beams. He pointed out that astronomical observations over a period of four years would be needed to observe Ephemeris time to the accuracy that could be obtained by the first NPL caesium clock in a matter of minutes. Nevertheless, the astronomers went ahead with Ephemeris Time, reluctant to hand over the role of defining time to the physicists.

Now I mentioned before that ET was difficult to measure as a primary standard from the Sun because the rate of change was so slow with a one year repeat period. However the Moon appears to move much faster across the sky and could be used as a secondary standard for more precise measurements. But it’s difficult to measure the position of the Moon against the background of the fixed stars by capturing both on a single photographic plate, which was the best observational technique at that time.

The dynamic range of a photographic emulsion is limited, and an exposure to correctly record the Moon will not be long enough the capture the stars. On the other hand, a long exposure with a tracking telescope to record the stars will not only severely over-expose the Moon, but the Moon image will be smeared out by the relative motion of the Moon during the exposure.

It was to solve this problem that William Markowitz at the US Naval Observatory invented the “dual-rate Moon camera” in 1951. He arranged a plane-parallel disc of dark glass, just bigger than the image of the Moon, in position just above the photographic plate.

Tilting this plate displaces the image by a known amount, so by altering the angle of tilt during the exposure, the Moon image was tracked by the differential speed of the Moon relative to the stars. The Moon and stars were both recorded sharp and correctly exposed for the instant in time when the disc was perpendicular to the light beam.This Moon camera was the ideal instrument to use for calibrating the frequency of the caesium clock to ET. Markowitz with his telescopic camera attachment was in Washington and Essen with his caesium beam apparatus was in Teddington. Essen and Parry had conducted many comparisons with UT which showed that between 1955 and 1958 the period of Earth rotation to be increasing by 0.43 ms per year - considerably above the long term average. By using his Moon Camera, Markowitz was able to measure the relation between ET and UT and by comparing this with the caesium data for UT obtained by Essen and Parry over the same period, they were able to calibrate the caesium frequency as 9,192,631,770 ±20 Hz with time defined by the Ephemeris second.

By the late 1950s there were enough caesium clocks operating in different national laboratories that it became convenient to use them as secondary standard for Ephemeris Time. A new scale was introduced in 1958 called Atomic Time combined from these atomic clocks, which was set to UT on 1 January 1958. This definition of the second was adopted into the SI system in 1960

In 1967 this timescale was formally adopted as TAI. Responsibility for the definition of the second had been taken out of the hands of the astronomers and given to the national standards laboratories. The second was defined as 9 129 631 770 ticks of the caesium atomic clock. This timescale was administered by the BIH and was used as the basis for Coordinated Universal Time (UTC).

Now ET had been set exactly to noon UT on 1 January 1900, and the ET second had been defined by the IAU as 1/31 556 925.9747 of the [tropical year](https://en.wikipedia.org/wiki/Tropical_year) for 1900, or rather of what they had calculated the tropical year to have been based on Newcomb’s data from the 18th and 19th Centuries. But the Earth had been slowing down and so the internationally agreed second was actually too short.

In fact, it was effectively the same as what a mean solar second had been in about 1820. A year had to consist of 31,536,000 mean solar seconds but, by the mid 1950 it was nearer to 31,536,001 ET seconds because the Earth rotation is slowing down by 1.7 ms/day/century. Today this would be almost 31,536,002 ET seconds.Consequently, the international timescales, TAI and UTC, were to diverging, so in 1972 the decision was taken to introduce a “leap second” into Universal Time. This is analogous to a leap year day in the calendar and involved an extra second inserted when the difference increases to 0.6 s. The protocol allows for the leap second to be inserted or removed at the end of any month with preference to June and December. There was a 10 second jump in 1 January 1972 to account for the accumulated difference and since then there have been 16 occasions when a leap second has been introduced at the end of December, and 11 introduced at the end of June. This gives 27 leap seconds over 46 years - an average of 1 every 1.7 years.

Responsibility for introducing leap seconds was given in 1988 to the International Earth Rotation Service, which tracks slippage of the ITRF against the ICRF and issues the notice periodically in Bulletin C, which you can get on the Internet.

There will not be a leap second in June this year, and a decision on one in December will be taken in September this year.You are probably already aware of leap seconds and have read in newspaper articles that it is necessary to add a second because the Earth rotation has slowed down. Well, as you can see, this is not really true. The Earth rotation slows due to tidal friction from the Moon to increase the length of a day by about 1.7 ms over a century. With 365 days in a year, each day being 1.7 ms longer, this means that after 100 years the year has increased by only about 0.6 s. The real reason it is necessary to have so many leap seconds is because the value of the SI second, inherited from ET, is too short. Although the need for leap seconds could never be entirely eliminated, it would only have to be inserted about once per century if the TAI second had been pegged to UT1 in 1967, rather than to ET. If the SI second were true to today’s Earth rotation, a leap second would not be needed for more than 120 years rather than roughly every year. But it’s too late to re-define the second now as the whole SI system depends on it.

Now the reason for introducing the leap second into UTC was to make sure that the international time scale was close to UT1. This seemed a good idea at the time (1972), but with the development of faster computers and international communications, problems began to emerge.

Leap seconds cannot be anticipated much in advance and have to be implemented by hand as and when they occur. This is where errors can creep in. Also, they have to be inserted simultaneously in all segments of a system working worldwide. Just to give two examples, global navigational satellite systems (GNSS) require the atomic clocks in all the satellites to be synchronised to nanoseconds, and time-stamping on financial transactions must be to 100 microseconds (MiFID 2). Both of these would be grossly affected by an error of one whole second in synchronisation.

In 2013 there was an incident when Thompson Reuters news agency announced US trading figures 15 ms early, before they were available to non-subscribers, and in that 15 ms, £28M of trading deals were enacted. Unless time distribution is within MiFID specification you can imagine the law suits and disputes which could arise.

Various time distribution services have been established for financial trading. NPL has a time distribution hub in Docklands distributing UTC (UK), but many financial trading systems get their time from GNSS.

The GPS navigation system uses several atomic clocks in each satellite to generate a continuous time system known as GPS time which does not insert leap seconds. The offset from UTC is transmitted separately and can be used to generate UTC at the receiver. The epoch of GPS time is in January 1980, so it always lags TAI by 19 seconds and increasingly lags on UTC.

So we have a range of different time scales.In 2005 a proposal from the USA to abandon leap seconds from the international world standard was made to IERS. This has been discussed extensively at subsequent meetings and objections raised by agencies who want civil time to remain tied to UT1 so that the sun appears overhead at noon. The issue has repeatedly been deferred and sent to study groups, and is not due to be discussed again until the IERS meeting in 2023.

Now it’s obviously nice to have the clocks conforming to the Sun, but how accurately do we need civil time to match UT1? Well UT1 is the same as mean solar time on the Greenwich Meridian, zero degrees longitude. This was important in the days when navigation was by means of the Nautical Almanac, but do we really need this in the age of satellite navigation? If leap seconds were to be simply abandoned, as the 2005 proposal, the meridian on which mean solar time matched UTC would gradually, very very slowly, drift westwards. And it would continue drifting westwards. But would it matter if the sun were not overhead at noon? Would it really upset anyone if the meridian on which the Sun was overhear at 12:00 local mean time passed through, say, Birmingham instead of Greenwich?

Already, with the 1989 definition of the ITRF, the IRM runs measurably to the easy of the Airy meridian. This is equivalent to about 6 milliseconds, or the accumulated difference in Earth rotation over 10 years.

If the difference approaches an appreciable fraction of an hour (which would take thousands of years) we could simply move civil time to the next time zone to the west. Time zone boundaries are not fixed internationally, but are a matter for each nation to decide for itself, as you can see from this time zone map. And they are not all whole-hour changes between time zones. There are several time zones with half-hour differences and even one, Nepal, which choses a quarter-hour setting.

*Personally,* I think this would be the best solution and I look forward to it being implemented in 2023.

So my conclusion is that historically, when improvements in measuring time have conflicted with human appreciation of time, the physical measurements have always succeeded and astronomical customs regarding time units have followed the technology. This happened when crude time measurements like water clocks and candles replaced the unequal hours customs with equal hours, when pendulum clocks resulted in mean time rather than sundial time, and when the electric telegraph replaced local time with regulated time zones. I think that UTC using a quantum-based time unit, the caesium second, will be allowed to drift very slowly away from mean solar time on the Greenwich Meridian and we will not notice the difference in the end.

Just to finish up, here is the up-to-date chart of how the precision of time measurements had improved. The caesium beam immediately improved by two and a half orders of magnitude and subsequent improvements using cold atoms by a further five orders. And the latest optical clocks by another two and a half orders on this, giving an improvement of more than ten orders of magnitude during my lifetime.