ISSN 0970-5953

Volume 36. No. 2

April - June 2020

PHYSICS EDUCATION

www.physedu.in

Volume 36, Number 2

In this Issue

The laws of planetary motion, derived from those of a harmonic oscillator (following Arnold)

P. A. Horvathy[1] and Pengming Zhang[2]

[1]Institut Denis Poisson,
Tours University – Orléans University, UMR 7013 (France)
horvathy@lmpt.univ-tours.fr

[2]School of Physics and Astronomy,
Sun Yat-sen University, Zhuhai 519082, China
zhangpm5@mail.sysu.edu.cn

Submitted on 15-03-2020

Abstract

Kepler's laws of planetary motion are deduced from those of a harmonic oscillator following Arnold. Conversely the circular orbits through the Earth's center suggested by Galilei are consistent with an r^{-5} potential (as found before by Newton). Both the Kepler/oscillator correspondence and circular orbits are examples of dual potentials.

1 Introduction

Kepler's laws of planetary motion state that

1. K-I: *A planet moves on an ellipse, one of whose foci being occupied by the Sun*;
2. K-II: *The vector drawn from the Sun to the planet's position sweeps equal areas in equal times*
3. K-III: *The squares of the periods are as the cubes of the major axes of the ellipses.*

These laws can be deduced from the inverse-square force law and Newton's equations of motion; the usual proof requires higher mathematics, though [1].

Analogous statements can be demonstrated using elementary tools for a harmonic oscillator,

1. O-I: *The trajectory under a harmonic force is an ellipse, whose centre is the origin of the linear force*;
2. O-II: *the vector drawn from the centre of the ellipse to the position sweeps equal areas in equal times*;

3. O-III: *The periods are independent of the geometric shape of the trajectories.*

The similarities are manifest, but neither the differences can be overlooked.

This aim of note is intended to derive the Kepler laws from those of the oscillator by elaborating the rather concise indications given by Arnold [2].

The relation between Keplerian and harmonic motion is first established for circular trajectories. The second step is a purely geometric correspondence due to Zhukovsky [3] between two types of ellipses, namely those which appear for an oscillator and those for planetary motion, respectively. The third step is to extend this purely geometrical correspondence, to a dynamical one. This can be achieved by a rather subtle re-definition of time, viewed as a parameter along the trajectory.

The question asked by Newton in his *Principia* [4] was remarkably *different* from how planetary motion is studied in our time : instead of solving the dynamical equations for the inverse-square law as we do now, he asked, conversely: *With what form of a central force are the observed trajectories consistent ?*

Newton has actually found general answer for circular motion [5]; however his purely geometric argument is difficult to follow today. In sec.8 we solve this problem in a particular case considered by Galilei in his *"Dialogo..."* [6].

2 Circular motions

Let us consider the Kepler problem in the complex plane, with the Sun fixed at the origin. The motion of our planetary is determined by

$$\ddot{z} = -fM\frac{z}{|z|^3}, \tag{2.1}$$

where the "dot" means derivation w.r.t. newtonian time, $\dot{(\cdot)} = \frac{d}{dt}$. Our aim is to recover the usual Keplerian trajectories by solving Eq. (2.1).

We start our investigations with a (very) special case, namely with motion along a circle of radius $A = const$. The latter can be parameterised by the angle θ, $z(t) = A\,e^{i\theta(t)}$; then (2.1) requires

$$i\ddot{\theta} - \left(\dot{\theta}\right)^2 = -\frac{fM}{A^3}\,. \tag{2.2}$$

From the vanishing of the imaginary part we infer that the motion is uniform along the circle, $\ddot{\theta} = 0$, and from the real part we deduce the angular velocity,

$$\dot{\theta} = \sqrt{fM/A^3} \equiv \Omega. \tag{2.3}$$

It is important that the trajectory studied here can also be viewed as that of a harmonic oscillator, constrained to move on a circle. Again using a complex coordinate, w, the equation of motion of the planar oscillator are

$$w'' = -\Omega^2 w, \tag{2.4}$$

where Ω (assumed real) is the frequency of the oscillator and the "prime" denotes derivation w.r.t. "oscillator time", $(\,\cdot\,)' = \frac{d}{d\tau}$. Then $w(\tau) = B\,e^{i\gamma(\tau)}$ yields $i\gamma'' -$

$(\gamma')^2 = -\Omega^2$ which is (2.2), provided the angle θ is identified with γ, t with τ and Ω^2 with fM/A^3.

Below we extend this correspondence to general motions.

3 The planar oscillator

The motions of a planar oscillator are readily determined. Decomposing (2.4) into real and imaginary parts, we observe that they perform independent harmonic motion. With an appropriate choice of the parameters,

$$w(\tau) = a\cos\Omega\tau + ib\sin\Omega\tau. \tag{3.1}$$

Thus $w(\tau)$ describes an ellipse, whose centre is the origin and has major and minor axes a and b, respectively.

The areal velocity is half of the conserved angular momentum,

$$I_{osc} = |w|^2\,\frac{d\gamma}{d\tau} = \text{const}\,. \tag{3.2}$$

The period,

$$T_{osc} = \frac{2\pi}{\Omega}, \tag{3.3}$$

is independent of the geometrical data of the orbit, as stated.

In conclusion, we have proved the "Kepler" laws O-I–O-II–O-III of the oscillator-motion.

We record for further use the expression for the (conserved) energy in terms of the geometric data of the trajectory,

$$E_{osc} = \tfrac{1}{2}\left(|w'|^2 + \Omega^2|w|^2\right) = \tfrac{1}{2}\Omega^2(a^2+b^2) > 0. \tag{3.4}$$

4 Some (complex) geometry [2]

Let u denote a complex variable and let us consider the so-called "Zhukovsky-map" of the complex plane,

$$u \mapsto w = u + \frac{1}{u}\,. \tag{4.1}$$

Lemma 1. *The image of a circle of radius $\rho > 0$ whose centre is the origin of the u-plane is an* ellipse, *whose centre is the origin of the w-plane. The foci of the ellipse are at the points ± 2.*

Proof : if $u = \rho e^{i\phi}$, $0 \le \phi \le 2\pi$, then

$$w = \left(\rho + \frac{1}{\rho}\right)\cos\phi + i\left(\rho - \frac{1}{\rho}\right)\sin\phi, \tag{4.2}$$

which is the equation of an ellipse centered at $w = 0$ and has major and minor axes $a = \rho + \rho^{-1} \ge 2$ and $b = \rho - \rho^{-1} \ge 0$, respectively. The foci are at $\sqrt{a^2 - b^2} = \pm 2$ from the centre. The angles ϕ and $\gamma = \arg w$ are related as $\tan\gamma = \frac{\rho^2-1}{\rho^2+1}\tan\phi$.

When $\rho \to 1$ the ellipse degenerates to the segment $[-2, 2]$ on the real axis. If the radius of the u-circle is changed from $\rho \neq 1$

to ρ^{-1} then the image describes, by (4.2), the same ellipse reflected on the real axis. We can, therefore, chose $\rho > 1$, i.e., to restrict the mapping $u \mapsto w$ to the exterior of the unit circle.

Let us now consider, following Levi-Civita [7] the mapping of the w plane onto the z-plane

$$w \mapsto z = w^2. \tag{4.3}$$

Lemma 2. *The image under (4.3) of the previously constructed w-ellipse is an ellipse in the z-plane, whose left focus is the origin of the z-plane. Conversely, each ellipse of the z-plane with one focus at the origin is obtained (after a suitable rotation) as the square of an ellipse, whose centre is the origin of the w-plane.*

This follows from $z = w^2 = 2 + u^2 + u^{-2}$: putting $u = \rho e^{i\phi}$ yields,

$$z = 2 + \left(\rho^2 + \frac{1}{\rho^2}\right)\cos 2\phi + i\left(\rho^2 - \frac{1}{\rho^2}\right)\sin 2\phi, \tag{4.4}$$

cf. (4.2), which is again an ellipse, whose axes are $A = \rho^2 + \rho^{-2}$, and $B = \rho^2 - \rho^{-2}$, and is shifted by 2 units to the right. Its left focus is at $z = 0$.

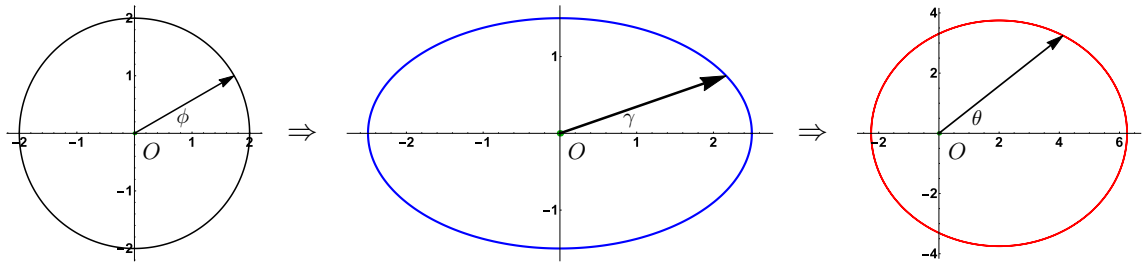
While u goes around the u-circle once, w describes the ellipse with centre at $w = 0$ also once. z however describes its ellipse *twice* : while w describes half of an ellipse, z describes the full image-ellipse. Thus ϕ can be restricted to $0 < \phi < \pi \Rightarrow 0 < \gamma < \pi$. The z-perihelion and aphelion points are, in particular, the images of the end points of the minor and resp. major axes of the w-figure, $\phi = \pi/2$ and $\phi = 0$, see Fig.1.

Let us note that while the Zhukovsky map (4.1) yields "true" w-ellipses but no circles ($a > b$), the Levi-Civita map (4.3) would work with no such restriction. The image of a w-circle $ae^{i\gamma}$ would be simply the z-circle $a^2e^{i2\gamma}$. The Zhukovsky map merely provides a convenient parametrization for the image.

Every z-ellipse centered at the origin can be obtained by a suitable rotation and dilation of the one considered here [2].

5 The oscillator - Kepler correspondence

The geometric construction of the previous section, and Eq. (4.3) in particular, swaps *oscillator trajectories* with *Keplerian* ones. One can wonder if the correspondence can be extended also to the *dynamics*. At first sight, the answer seems to be negative: differentiating $z = w^2$ w.r.t. τ the equation of motion (2.4) of w yields a rather complicated expression which is manifestly different from (2.1). This mismatch is also seen by expressing Kepler's areal velocity [i.e. half of the angular



u $\qquad$ $w = u + u^{-1}$ $\qquad$ $z = w^2$

Figure 1: *While* **u** *describes the* **circle** *of radius* $\rho \neq 1$ *of the complex plane,* $\mathbf{w = u + u^{-1}}$ *moves on an* **ellipse** *centered at the origin of the* **w***-plane, and* $\mathbf{z = w^2}$ *describes another* **ellipse** *one of whose foci is the origin of the* **z***-plane.*

momentum] derived using that of the oscillator,

$$I_K = |z|^2\theta' = 2|z|I_{osc} \tag{5.1}$$

[where we used $\theta = 2\gamma$], which *can not be a constant of the motion* therefore, unless $|z(t)|$ is a constant i.e., with the exception of circular trajectories. In other words, Kepler's 2nd law is not in general satisfied.

But Eq. (5.1) shows also the way to resolve the contradiction: *time*, or more precisely *the parameter used along the trajectories, should be redefined*. Let us indeed consider an arbitrary oscillator-motion $w(\tau)$, and define the new "time" along the trajectory as

$$t = \int |w(\tau)|^2 d\tau \quad \Longrightarrow \quad \frac{d}{dt} = \frac{1}{|w(\tau)|^2}\frac{d}{d\tau} \tag{5.2}$$

i.e., $\dot{(\,\cdot\,)} = |w|^{-2}(\,\cdot\,)'$. Then, using (2.4),

$$\ddot{z} = \frac{1}{w\bar{w}}\frac{d}{d\tau}\left(\frac{1}{w\bar{w}}\frac{d}{d\tau}w^2\right) = -2\left(|w'|^2 + \Omega^2|w|^2\right)\frac{1}{w\bar{w}^3} = -4E_{osc}\frac{w^2}{|w|^6}\,. \tag{5.3}$$

Here E_{osc} is the oscillator-energy in Eq. (3.4), which remains constant along the trajectory. Then, putting z for w^2 yields Eq. (2.1) of planetary motion with the identification

$$4E_{osc} = fM. \tag{5.4}$$

Hence, to each oscillator-trajectory is associated a Keplerian trajectory, whose gravitational force is four times the oscillator-energy, and vice versa. The inverse of (4.3-5.2)

$$w = \sqrt{z}, \quad \frac{d}{d\tau} = \sqrt{z\bar{z}}\,\frac{d}{dt} \quad \Longrightarrow \quad w'' = \frac{1}{2}\left(\frac{1}{2}|\dot{z}|^2 - \frac{fM}{|z|}\right) w, \tag{5.5}$$

which is the equation of motion of an oscillator, whose frequency (square) is proportional to the energy of the Keplerian trajectory,

$$w'' = -\Omega^2\, w, \quad \Omega^2 = -\frac{1}{2}E_K, \qquad E_K = \frac{1}{2}|\dot{z}|^2 - \frac{fM}{|z|} = -\frac{fM}{2A}, \tag{5.6}$$

where A denotes the major axis of the Keplerian ellipse. Let us stress that the transformation is defined along trajectories only: (5.5) associates a *different* oscillator-trajectory to each Keplerian orbit. For elliptic Kepler motions the frequency Ω is real, consistently with $E_K < 0$.

6 The Kepler laws

Now we derive the Kepler laws, KI-II-III, from those, OI-II-III, of the oscillator.

Firstly, the Keplerian orbits being images of the oscillator orbits satisfy KI.

Secondly,

$$I_K = |z|^2\frac{d\theta}{dt} = |w|^2\frac{d(2\gamma)}{d\tau} = 2I_{osc}, \tag{6.1}$$

so that the areal velocity of the planet, $\frac{1}{2}I_K$, is itself a constant of the motion; this is KII.

Finally, let denote T_K and T_{osc} the periods of the Keplerian resp. oscillator motions. From the conservation of the areal velocities, $\pi ab = \frac{1}{2}I_{osc}T_{osc}$, $\pi AB = \frac{1}{2}I_K T_K$, because the area of an ellipse is π-times the product of its two axes. Then, using $T_{osc} = 2\pi/\Omega$,

$$T_K = \pi\left(\frac{AB}{ab}\right)\frac{1}{\Omega}.$$

As seen in Sect. 4, $a = \rho + \rho^{-1}$, $b = \rho - \rho^{-1}$ and therefore $AB/ab = A$. But $a^2 + b^2 = 2(\rho^2 + \rho^{-2}) = 2A$. Then from the expression (3.4) for the energy of an oscillator-ellipse we infer, using (5.4), that a Keplerian trajectory with major axis A has frequency and period

$$\Omega = \sqrt{\frac{E_{osc}}{A}} = \sqrt{\frac{fM}{4A}} \quad \Rightarrow \quad T_K = 2\pi\sqrt{\frac{A^3}{fM}} \tag{6.2}$$

which plainly implies Kepler's III. law.

7 Scattered motions

As mentioned before, the formula for the (negative) total energy of planetary motion in (5.6) is only valid for bound (elliptic) motions; for unbound (parabolic or hyperbolic) motions, the energy is zero resp. positive. Our oscillator-planet correspondence can be extended to these motions, and in fact also to those in Rutherford's experiment with repulsive interaction.

Let us start with the geometry. We use again the sequence

$$u \mapsto w = u + u^{-1} \mapsto z = w^2;$$

the only difference with the previous case is that the variable u now describes, instead of a circle as on Fig.1, a *straight line through the origin* , as shown on Figs. 2-3. For $u = \rho\, e^{i\phi}$ w is formally again (4.2), but now it is $\rho \neq 0$ which varies (between $-\infty$ to ∞), while the angle is kept fixed, $\phi = \text{const}$. Then from

$$\left(\frac{\Re(w)}{\cos\phi}\right)^2 - \left(\frac{\Im(w)}{\sin\phi}\right)^2 = \left(\rho + \rho^{-1}\right)^2 - \left(\rho - \rho^{-1}\right)^2 = 4 \tag{7.1}$$

we infer that (for $\phi \neq 0,\ \pi$) we have now a *hyperbola* with real and imaginary axis

$$a = |2\cos\phi| \quad \text{and} \quad b = |2\sin\phi|\,, \tag{7.2}$$

respectively. The foci are at ± 2 because $a^2 + b^2 = 4$. Fixing ϕ between zero and $\pi/2$ and varying ρ from $-\infty \to \infty$, the full Zhukovsky hyperbola is obtained : when $0 < \phi < \pi/2$, then, for $\rho > 0$ we have $\Re(z) \geq 2$ and we get the right branch of the w-hyperbola. For $\rho < 0$ we get instead its left branch, $\Re(z) \leq -2$. If $\phi = 0$, w describes two half-lines ($\Re(z) \geq 2$ resp. $\Re(z) \leq -2$) of the real axis; for $\phi = \pi/2$ we get the imaginary axis.

Choosing ϕ between $\pi/2$ and π would yield once again the same figure, but in the opposite order. It is therefore enough to choose ϕ in the interval $0 \leq \phi \leq \pi/2$.

The square of the Zhukovsky hyperbola is again (4.4) which belongs again to a hyperbola shifted by 2 to the right. This follows from

$$\left(\frac{\Re(z) - 2}{\cos 2\phi}\right)^2 - \left(\frac{\Im(z)}{\sin 2\phi}\right)^2 = \left(\rho^2 + \rho^{-2}\right)^2 - \left(\rho^2 - \rho^{-2}\right)^2 = 4\,. \tag{7.3}$$

This expression is symmetric with respect to changing the sign of ρ, therefore we can restrict ourselves to $\rho > 0$. The real and imaginary axis are $A = |2\cos 2\phi|$ and $B = |2\sin 2\phi|$; its left focus is the origin.

The subtlety is that choosing the constant parameter ϕ between $\pi/4$ and $\pi/2$, we have $\Re(z) \leq 2$ so we get only the *left-branch* of the hyperbolae cf. Fig.2; fixing ϕ instead between zero to $\pi/4$ then $\Re(z) \geq 2$ so that z describes the *right-branch*, cf. Fig.3.

Turning to the dynamics, the oscillator-Kepler correspondence, (4.3-5.2) and (5.5) are formally the same as before, except that for positive planet-energy the oscillator-frequency (5.6) is imaginary, $\Omega^2 < 0$. Let us assume therefore that the linear force is repulsive (also called an inverted oscillator). Then

$$w'' = +|\Omega|^2 w \qquad \Longrightarrow \qquad w(\tau) = a\cosh|\Omega|\tau + ib\sinh|\Omega|\tau \tag{7.4}$$

cf. (3.1), which is a Zhukovsky-hyperbola with its centre at the origin.

The potential energy of the repulsive oscillator is negative, $-\frac{1}{2}|\Omega|^2|w|^2$. Therefore the total energy,

$$E_{osc} = \tfrac{1}{2}\left(|w'|^2 - |\Omega|^2|w|^2\right) = \tfrac{1}{2}|\Omega|^2(-a^2+b^2) \tag{7.5}$$

[cf. (3.4)] can be either positive or negative, depending on whether the imaginary or the real axis is longer.

Our investigations in Sec.5 are still valid, and still yield motion with inverse-square force law. When the oscillator-energy is positive, the Kepler problem with $fM = E_{osc} > 0$ is obtained, with a hyperbolic trajectory. $E_{osc} > 0$ means, moreover, that the imaginary axis is the longer one, $b > a$. By (7.2) this happens when $\pi/4 < \phi < \pi/2$ and then the image "Zhukovsky-ellipse" is the *left-branch*, – the one which turns *towards the Sun* as it should for attractive interaction. This is what happens for most non-periodic comets observed in the solar system.

What is the use of "oscillator" motions with negative energy, $E_{osc} < 0$? The equation of motion in this case,

$$\ddot{z} = (-E_{osc})\frac{z}{|z|^3} \tag{7.6}$$

cf. (5.3), describes the motion with a repulsive inverse-square force, as in the *Rutherford experiment*, when light α-particles with

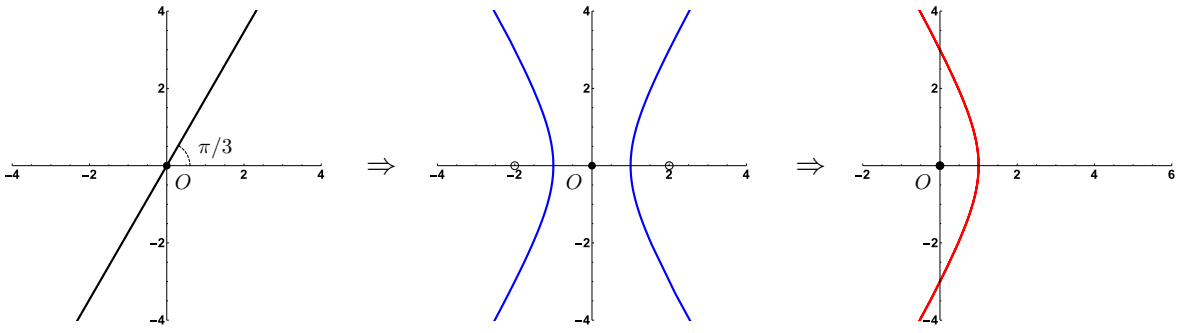


Figure 2: *The image of a* **straight line** *through the origin of the* **u***-plane with inclination* $\pi/4 < \phi < \pi/2$ *is a* **Zhukovsky hyperbola**, *whose centre is the origin of the* **w***-plane. Squaring the latter provides us with the* **left** *branch of a hyperbola in the* **z***-plane, whose inner focus is* $z = 0$.

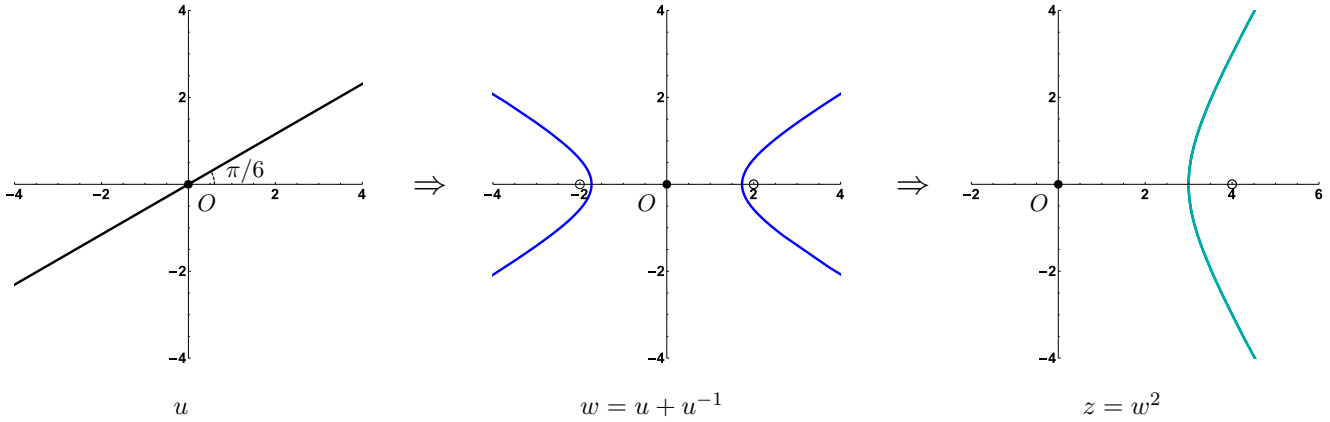


Figure 3: *The image of a* **straight line** *through the origin of the* **u***-plane with inclination* $0 < \phi < \pi/4$ *is a* **Zhukovsky hyperbola**, *whose centre is the origin of the* **w***-plane. Squaring the latter provides us with the* **right***-branch of a hyperbola in the z-plane in the* **z***-plane, whose outer focus is* $z = 0$.

charge q are scattered on a heavy atomic nucleus with charge of the same sign Q. The interaction is repulsive, described by (7.6), with the correspondence

$$4E_{osc} = -qQ. \tag{7.7}$$

However, by (7.5) having negative oscillator-energy implies the real axis being the longer one. But this happens when $0 < \phi < \pi/4$ – i.e., when squaring yields the *right-branch* of the z-hyperbola, shown on Fig.3. In this case the α-particle is *turned away* from the nucleus, consistently with the repulsive interaction..

Conversely, since the energy is always positive for the repulsive inverse-square force,

$$E_{Coulomb} = \frac{1}{2}|\dot{z}|^2 + \frac{qQ}{|z|} > 0 \tag{7.8}$$

the inverse transformation (5.5) associates to the Coulomb-Rutherford problem a repulsive linear system, whose (imaginary) frequency is determined by the Coulomb-energy (7.8).

Parabolic orbits have vanishing energy; their associated "oscillator-frequency" is therefore $\Omega = 0$ by (5.6) – i.e., w moves freely along a straight line. If the motion does not go through the origin, then one can achieve with a suitable dilation and rotation that the trajectory be $w(\tau) = i + \tau$. Then

$$z = w^2 = (\tau^2 - 1) + 2\tau i \tag{7.9}$$

is the equation of a "horizontally lying" parabola, as seen from $y^2 = 4x + 4$. The vertex of the parabola is at $x = (-1)$ and its focus at the origin $z = 0$. If our straight line does go through the origin, its image degenerates into a half-line.

8 Galilei and Newton

Galileo Galilei, in his "Dialogo…" [6] suggests that a body dropped from a tower (say the Leaning Tower of Pisa) on rotating Earth would follow a circular trajectory which, (if it was not stopped), would pass through the center of the Earth, cf. Fig.4.

What would be the law of gravitation consistent with Galilei's statement ?

Consider a body (of unit mass) dropped from C and assume that it follows a half-circle which passes through A, the Earth' center. Its diameter is thus the radius of the earth, $2\rho = AC = R$, see (Fig.5). The Earth attracts the body by a force $\boldsymbol{F}$ directed, for symmetry reasons, towards the center, A, of the Earth; assume that $F = |\boldsymbol{F}|$ is proportional to some power of the distance,

$$F = |\boldsymbol{F}| = \gamma\, r^{\mu}, \tag{8.1}$$

where γ and μ are constants to be determined. Decompose $\boldsymbol{F}$ into components : $\boldsymbol{F}_n$, the normal component, is directed towards the center, E, of the trajectory and determines the centripetal acceleration, while the tangential component, $\boldsymbol{F}_t$, determines the tangential acceleration,

$$\frac{v^2}{\rho} = F_n, \qquad a = F_t, \tag{8.2}$$

where v is the velocity at T and $a = \dot{v}$. Denoting the angle CAT by ϕ and the arc

length CT by s, we have $F_n = F\cos\phi$, $F_t = F\sin\phi$, $s = \rho\, 2\phi = R\phi \;\Rightarrow\; v = \dot{s} = R\dot{\phi}\,, a = R\ddot{\phi}$. Thus we have to solve

$$(\dot{\phi})^2 = \frac{F}{2R}\cos\phi, \qquad \ddot{\phi} = \frac{F}{R}\sin\phi\,. \quad (8.3)$$

Using (8.1) and $r = AT = 2\rho\cos\phi = R\cos\phi$ yields,

$$\dot{\phi} = \sqrt{\frac{\gamma}{2}}\, R^{\frac{\mu-1}{2}}\cos^{\frac{\mu+1}{2}}\phi \;\Rightarrow\; \ddot{\phi} = -\left(\frac{\mu+1}{2}\right)\sqrt{\frac{\gamma}{2}}\, R^{\frac{\mu-1}{2}}\cos^{\frac{\mu-1}{2}}\sin\phi\cdot\dot{\phi}\,. \quad (8.4)$$

Then comparing with (8.3) leaves us with an identity, *provided*,

$$\boxed{\mu = -5 \quad \text{i.e.} \quad F(r) = \gamma\, r^{-5}\,.} \quad (8.5)$$

The body's initial velocity is that of the Earth, $\dot{\phi}(0) = \omega = 2\pi/day$. By (8.4) $\gamma = 2R^6\omega^2$.

In fact, the problem can be further generalized. Let us assume that a point particle moves along a known planar trajectory $r = r(\phi)$ under a central force, which is directed towards the origin. The force depends on the distance and the azimuthal angle, $\mathbf{F} = \mathbf{F}(r,\phi) = F(r,\phi)\hat{\mathbf{r}}$. What force law can be consistent with the trajectory?

Angular momentum conservation implies that the trajectory lies in the plane and that $r^2\dot{\phi} = h = const$. The radial component of the equation of motion is

$$m(\ddot{r} - r\dot{\phi}^2) = F(r,\phi), \quad (8.6)$$

implying

$$\boxed{F(r,\phi) = m\left(\ddot{r} - \frac{h^2}{r^3}\right) = -\frac{mh^2}{r^2}\left(\frac{d^2(\frac{1}{r})}{d\phi^2} + \frac{1}{r}\right).} \quad (8.7)$$

For the Kepler ellipsis

$$r = \frac{p}{1+\epsilon\cos\phi} \;\Rightarrow\; F(r,\phi) = -\frac{mh^2}{p}\frac{1}{r^2}. \quad (8.8)$$

For circular motion, taking the force centrum as the origin, (8.7) implies

$$F(r,\phi) = -8mh^2R^2\frac{1}{r^5}, \quad (8.9)$$

consistently with (8.5).

9 Outlook and some history

The correspondence between a harmonic oscillator and the Kepler problem, suggested by Arnold [2] following Levi-Civita [7] and Bohlin [8] allowed us to deduce the Kepler laws of planetary motion from those, simpler, of an oscillator. The correspondence can be extended to scattered motions exemplified by non-periodic comets and Rutherford scattering.

A key element is the redefinition (5.2) of the parameter along each trajectory, suggested by the comparison (6.1) of the respective areal velocities.

Kepler, in his *"Astronomia Nova"* [9] deduced his First Law from Tycho's observations of Mars (and then boldly extended to all planets). Modern textbooks derive in turn the laws of planetary motion from Newton's inverse-square law [1]. The question can however be raised also conversely: *What form of a central force can be consistent with observed trajectories ?*

As an example, one can wonder with what force law would Galilei's (naive) picture in his "Dialogo", Fig.4, consistent ? A simple calculation shows that such a trajectory requires a force proportional to the *minus fifth power* of the distance, (8.5). The more general question : *"Which central force law does allow circular motion ?"* was *answered* by Newton [4], vol I. Proposition VII. Problem II. using elementary geometry.

At the end of the 17th century, the consistency of the inverse-square law with circular motion around the Earth' center was known widely ; Newton mentions Hooke, Wren, Huygens. However the consistency of Kepler's *elliptic* trajectories with the inverse-square law was proved first by Newton [4]. More generally, he proved that the Law of Universal Gravitation allows for conic sections. His contemporaries may have been wondering if these mathematical possibilities do actually exist in nature. The answer was given by Newton himself: collecting the observational data from all around the Earth, he proved that the Great Comet of 1680 followed a *parabolic* trajectory [4] [11]. Astronomers have been discovering weakly hyperbolic comets since the mid-1800s.

Arnold argues that both examples mentioned above are just particular cases of *dual potentials* [2], Thm 3 p. 97 : let us assume that we have a central force whose strength is proportional to the distance raised to the power a. Then the mapping $w \to z = w^\alpha$ carries its trajectories to those in a central field whose strength is proportional to the distance raised to the power A, where

$$(a+3)(A+3) = 4, \qquad \alpha = \frac{a+3}{2}. \quad (9.1)$$

The potentials r^a and and r^A are called dual. For example, the harmonic and the Newtonian forces are dual to each other : they correspond to $a = 1$ and $A = -2$ respectively ; then $\alpha = 2$ as above. The r^{-5} potential is *self-dual* : it corresponds to $a = -5 = A$. $\alpha = -1$ i.e. $z = 1/w$ is the inversion.

We just mention, in conclusion, that the duality relation mentioned here can be generalized using the theory of conformal mappings [2, 12, 13, 14, 15, 16, 17], which can be used to large variety of problems which include, apart of planetary motion, also quantum properties and even general relativity. However their study goes beyond our scope here.

References

[1] H. Goldstein, *Classical mechanics*, 2nd ed. Addison-Wesley (1980).

[2] V. I. Arnold: *Huygens & Barrow, Newton & Hook*. Birkhäuser (1990).

[3] http://en.wikipedia.org /wiki/Nikolay_Yegorovich_Zhukovsky

[4] I. S. Newton, *Philosophia Naturalis Pricipia Mathematica*. London: Royal Society of London (1686), translated by A. Motte as *Sir Isaac Newton's Mathematical Principles of Natural Philosphy and his System of the World* (1729). Translation revised by F. Cajori, Berkeley: University of California Press (1946).

[5] Newton has proved also that elliptic trajectories can be consistent with the inverse-square law.

[6] Galileo Galilei, "Dialogo sopra i due massimi sistemi del mondo," Firenze, (1632).

[7] T. Levi-Civita: "Sur la résolution qualitative du problème resteint des trois corps," Acta Math. **30**, 305-327 (1906), *Opere Matematiche*, Vol. **2**, p. 419. Bologna (1956).

[8] M. K. Bohlin, "Note sur le problème des deux corps et sur une intégration nouvelle dans le problème des trois corps," Bull. Astron. **28**, 113 (1911).

[9] Johannes Kepler : *Astronomia Nova* ... Heidelberg, (1609).

[10] https://en.wikipedia.org /wiki/GreatCometof1680#/

[11] According to present knowledge [10], the 1680 comet follows an elliptic trajectory with period $>$ 10000 years.

[12] A. K. Grant and J. L. Rosner, "Classical orbits in power law potentials," Am. J. Phys. **62** (1994) no.4, 310.

[13] A. Nersessian, V. Ter-Antonian and M. M. Tsulaia, "A Note on quantum Bohlin transformation," Mod. Phys. Lett. A **11** (1996) 1605, [hep-th/9604197].

[14] Y. Grandati, A. Berard and H. Mohrbach, "Bohlin-Arnold-Vassiliev's duality and conserved quantities," arXiv:0803.2610 [math-ph].

[15] D. Kothawala, "Duality of force laws and Conformal transformations," Am. J. Phys. **79** (2011) 6, [arXiv:1010.2238 [physics.class-ph]].

[16] G. W. Gibbons and M. Vyska, "The Application of Weierstrass elliptic functions to Schwarzschild Null Geodesics," Class. Quant. Grav. **29** (2012) 065016, [arXiv:1110.6508 [gr-qc]].

[17] S. Chen, G. W. Gibbons and Y. Yang, "Explicit Integration of Friedmann's Equation with Nonlinear Equations of State," JCAP **1505** (2015) 020, [arXiv:1502.05042 [gr-qc]]; "Friedmann-Lemaitre Cosmologies via Roulettes and Other Analytic Methods," JCAP **1510** (2015) 056, [arXiv:1508.06750 [gr-qc]].

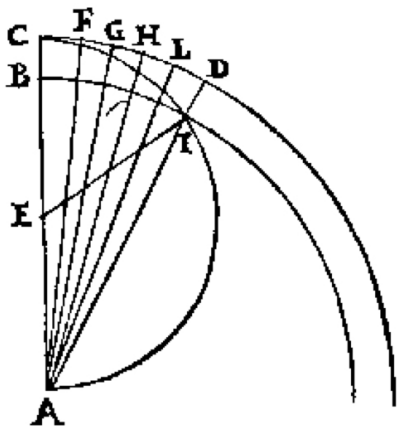


Figure 4: *Figure copied from Galilei's "Dialogo…"* [6] : who suggested that *a body dropped from C would follow a circular trajectory which passes through the center A of the earth.*

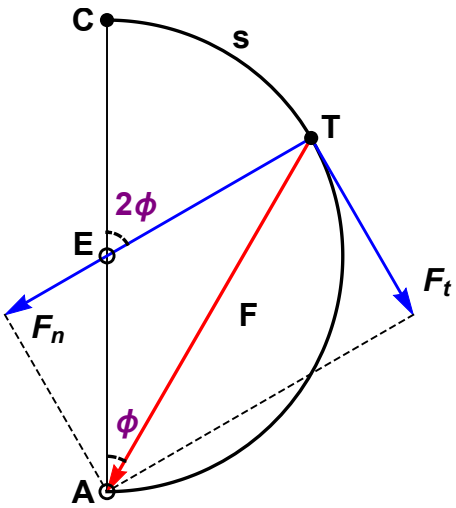


Figure 5: The force is directed towards A, the center of the Earth.

Determining planetary positions in the sky for ± 50 years to an accuracy of $\lesssim 1^\circ$ with a calculator

Tanmay Singal[1] and Ashok K. Singal[2]

[1]Department of Physics and Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China.
tanmaysingal@gmail.com

[2]Astronomy and Astrophysics Division, Physical Research Laboratory Navrangpura, Ahmedabad 380 009, India.
ashokkumar.singal@gmail.com

Submitted on 03-12-2019

Abstract

In this paper, we describe a very simple method to calculate the positions of the planets in the sky. The technique used enables us to calculate planetary positions to an accuracy of $\lesssim 1^\circ$ for ± 50 years from the starting epoch. Moreover, this involves very simple calculations and can be done using a calculator. All we need are the initial specifications of planetary orbits for some standard epoch and the time periods of their revolutions.

1 Introduction

The night-sky fascinates people. To be able to locate a planet in the night sky is something that thrills people. Since the planets move with respect to the background stars and continuously change their positions in the sky, locating them in the sky, especially when seen from Earth that continuously shifts its position around Sun, could appear be a non-trivial task. It is a general notion that calculating the planetary positions is a very tedious task, involving a lot of complicated mathematical equations and computer work. However, to be able to locate planets in the sky one does not really need very accurate positions. After all, Kepler's laws, which describe planetary orbits reasonably well, are mathematically simple. Hence, one could use Kepler's law to predict planetary positions in which mutual influence of planets is not considered. Thereby an accuracy of $\lesssim 1^\circ$ in planetary positions would be achieved.

In this paper, we employ a very sim-

ple method to calculate the positions of the planets. The technique we use enables us to calculate planetary positions to an accuracy of $\lesssim 1^\circ$ for ± 50 years from the starting epoch. Moreover, this involves very simple calculations and can be done using a calculator. All we need are the initial specifications of planetary orbits for some standard epoch and their time periods of revolution. Although accurate planetary positions could be obtained easily from the internet, yet it is very instructive and much more satisfying to be able to calculate these ourselves, starting from basic principles and using a simple procedure.

Our first step would be to calculate the positions of all the planets (including that of the Earth) in their orbits around the Sun. We initially consider the planets to revolve around the Sun in uniform circular motions. Knowing their original positions for the starting epoch, we calculate their approximate positions for the intended epoch. As a consequence of this approximation there will be some errors since the actual orbits are elliptical and in an elliptical orbit the angular speed of the planet is not uniform and to an extent varies. Therefore, to get more accurate positions, we need to make appropriate corrections, which are derived in Appendix A. These corrections account for the elliptical motion.

Knowing the positions of the planets around the Sun, we can then use simple coordinate geometry to transform their position with respect to an observer on Earth. Our task becomes simple since the orbits of all planets more or less lie in the same plane, viz. the ecliptic plane.

Actually for locating planets in sky, even manual calculations could suffice, however, to correct for the orbital ellipticity, numerical values of some trigonometric functions may be needed, otherwise one needs tabulated values of corrections [**?**]. Also, in order to transform the planet position to a geocentric perspective, a plot of the relative positions of Sun, Earth and the planet on a graph sheet or a chart, using a scale and a protractor, may be required. However, both these tasks could be performed with the help of a scientific calculator. Further, if one wants to calculate Moon's position too (even though one could locate Moon in sky easily, without its position calculation), required for knowing the New Moon or the Full Moon dates, or the dates of possible solar or lunar eclipses in a specific year, computations could be done on a scientific calculator.

In this paper, we calculate the motion of naked-eye planets only, although the procedure can be applied equally well for the remaining planets also.

2 Celestial co-ordinates

All celestial bodies in the sky, including stars, planets, Sun, Moon and other objects, appear to lie on the surface of a giant sphere called the Celestial Sphere. Due to Earth's eastward rotation around its axis, the ce-

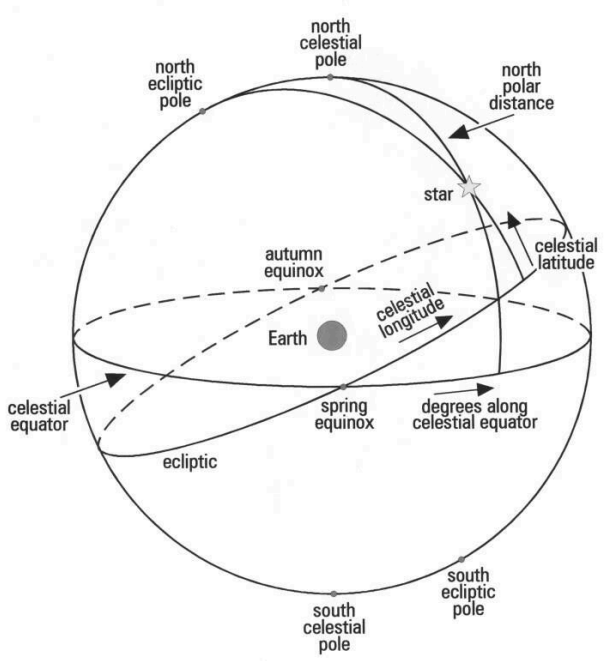


Figure 1: Celestial sphere showing the ecliptic co-ordinate system [1].

lestial sphere appears to rotate westward around Earth in 24 hours. Infinitely extending the plane of Earth's equator into space it appears to intersect the celestial sphere to form a circle, which is called the Celestial Equator.

As Earth moves around Sun - as seen from the Earth - Sun changes its position with respect to the background stars. The path that Sun takes on the celestial sphere is called the "Ecliptic". The familiar Zodiac constellations are just divisions of the ecliptic into twelve parts. Since all other planets revolve around Sun in nearly the same plane, they also appear to move on the ecliptic.

The celestial equator is inclined to the ecliptic by 23.5°. The points of intersections of these two circles on the celestial sphere are called the "Vernal Equinox" and the "Autumnal Equinox". The Vernal Equinox, also known as the Spring Equinox, is the point on the celestial sphere that the Sun passes through around 21st of March every year.

In astronomy, an epoch is a moment in time for which celestial co-ordinates or orbital elements are specified, while a celestial co-ordinate system is a co-ordinate system for mapping positions in the sky. There are different celestial co-ordinate systems, each using a co-ordinate grid projected on the celestial sphere. The co-ordinate systems differ only in their choice of the fundamental plane, which divides the sky into two equal hemispheres along a great circle . Each co-ordinate system is named for its choice of fundamental plane.

The ecliptic co-ordinate system is a celestial co-ordinate system that uses the ecliptic for its fundamental plane. The longitudinal angle is called the ecliptic longitude or celestial longitude (denoted by λ), measured eastwards from 0° to 360° from the vernal equinox. The latitudinal angle is called the ecliptic latitude or celestial latitude (denoted by β), measured positive towards the north. This coordinate system is particularly useful for charting solar system objects.

The Earth's axis of rotation precesses around the ecliptic axis with a time period of about 25800 years. Due to this, the equinoxes shift westwards on the ecliptic. Due to the westward shift of the Vernal

Table 1: Mean longitude λ_0 on 01/04/2020, 00:00 UT

Planet	λ_i (°):	T (days)	ω_0 (°/day)	λ_0 (°)
Mercury	250.2	87.969	4.09235	277.2
Venus	181.2	224.701	1.60213	150.6
Earth	100.0	365.256	0.98561	189.6
Mars	355.2	686.980	0.52403	270.9
Jupiter	34.3	4332.59	0.08309	288.8
Saturn	50.1	10759.2	0.03346	297.6

Equinox, which is the origin of the ecliptic co-ordinate system, the ecliptic longitude of the celestial bodies increases by an amount $360/258 \sim 1.4^\circ$ per century or a degree in ~ 72 years, the life span of a human.

Most planets, dwarf planets, and many small solar system bodies have orbits with small inclinations to the ecliptic plane, and therefore their ecliptic latitude β is always small. Due to only small deviations of the orbital planes of the planets from the plane of the ecliptic, the ecliptic longitude alone may suffice to locate planets in the sky.

3 Calculating planetary positions

3.1 Heliocentric circular orbit

Here, we consider the planets to move around Sun in circular orbits with a uniform angular speed. The initial values of mean longitudes (λ_i) of the planets given in Table 1 are for 1st of January, 2000 A.D., 00:00 UT (adapted from [2]). In Table 1, we have also listed the period, T (days), of revolution of the planets [3]. Then, the mean angular speed is give by, $\omega_0 = 360/T$ (°/day). We denote the Mean Longitude of the planets in the imaginary circular orbit for subsequent dates as λ_0.

We now demonstrate how to calculate λ_0 for Mars on April 1, 2020.

The initial mean longitude, λ_i, of Mars on 01.01.2000 at 00:00 UT = 355.2°.

Number of days between 01.01.2000 and 01.04.2020 = 7396.

Mean angle traversed duration this period $= 0.52403 \times 7396 = 3,875.7^\circ$.

So, on 01.04.2020 at 00:00 UT, the mean longitudes is

$\lambda_0, = 355.2 + 3,875.7 = 4,230.9 = 360 \times 11 + 270.9^\circ$.

We take out 11 integer number of complete orbits to obtain $\lambda_0 = 270.9^\circ$.

In the same way, mean longitudes of all planets have been calculated in Table 1 for the same epoch.

3.2 Heliocentric elliptical orbit

Our next step is to correct for the elliptical shape of the orbit as for some of the plan-

Table 2: Corrected longitude λ on 01/04/2020, 00:00 UT

Planet	λ_0 (°)	λ_p (°)	θ_0 (°)	e	$\Delta\theta$ (°)	θ (°)	λ (°)
Mercury	277.2	77.5	199.7	0.2056	-6.0	193.7	271.5
Venus	150.6	131.6	19.0	0.0068	0.3	19.2	151.1
Earth	189.6	102.9	86.7	0.0167	1.9	88.6	191.8
Mars	270.9	336.1	294.8	0.0934	-10.2	284.6	261.0
Jupiter	288.8	14.3	274.5	0.0485	-5.6	269.0	283.6
Saturn	297.6	93.1	204.5	0.0555	-2.5	202.0	295.4

ets, depending upon the eccentricity (e), the corrections could be substantial. In a circular motion, the angular speed of the planet is constant, however, in an elliptical orbit, the angular speed of the planet varies with time. Due to this varying angular speed, actual longitude λ of the planet will be somewhat different from the mean longitude λ_0.

Before we could make corrections for the elliptical shape of the orbit we need to know the orientation of the ellipse within the ecliptic and that can be defined by the longitude (λ_p) of the perihelion. The distance of the planet from Sun varies in an elliptical orbit and perihelion is the point on the elliptical orbit that lies closest to Sun (the orbital point farthest from Sun is called aphelion). Longitudinal distance of the planet from the perihelion along the elliptical orbit is known as its "anomaly" (denoted by θ), while angular distance of mean position of planet with respect to the perihelion is called the "mean anomaly" (denoted by θ_0). As has been discussed in Appendix A, there is a one–to–one correspondence between θ and θ_0.

The correction $\Delta\theta$ to be added to θ_0 (Appendix A, Equation 4) is

$$\Delta\theta = 2\,e\sin\theta_0 + \frac{5}{4}\,e^2\sin 2\theta_0,$$

where e is the eccentricity of the ellipse.

In Table 2, we have listed values of the longitude of perihelion (λ_p) and eccentricity (e) for all planets, taken from [2].

Let's consider Mercury's position on 01/04/20 at 00:00 UT.

Mean longitude, $\lambda_0 = 277.2^\circ$

Perihelion Longitude, $\lambda_p = 77.5^\circ$

Mean anomaly, $\theta_0 = \lambda_0 - \lambda_p = 199.7^\circ$

The 1st order correction then is

$\Delta\theta_1 = 2e\sin\theta_0 = 2 \times 0.2056 \times \sin(199.7^\circ) = -0.13861$ rad $= -7.9^\circ$,

while the 2nd order correction is

$\Delta\theta_2 = \frac{5}{4}e^2\sin 2\theta_0 = 1.25 \times (0.2056)^2 \times \sin(39.4^\circ) = 0.03354$ rad $= 1.9^\circ$.

The total correction then is

$\Delta\theta = \Delta\theta_1 + \Delta\theta_2 = -7.9 + 1.9 = -6.0^\circ$

The anomaly, $\theta = \theta_0 + \Delta\theta = 199.7 - 6.0 = 193.7^\circ$

Precession of vernal equinox in 20.25 years $= 20.25 \times 360/25800 \approx 0.3^\circ$.

Table 3: Geocentric longitude λ_g on 01/04/2020, 00:00 UT

Planet	a (A.U.)	e	r (A.U.)	λ (°)	r_g (A.U.)	λ_g (°)
Mercury	0.387	0.2056	0.463	271.5	1.024	345.3
Venus	0.723	0.0068	0.718	151.1	0.653	57.6
Earth/Sun	1.00	0.0167	0.999	191.8	0.999	11.8
Mars	1.52	0.0934	1.472	261.0	1.457	300.9
Jupiter	5.20	0.0485	5.192	283.6	5.318	294.4
Saturn	9.55	0.0555	10.04	295.4	10.32	300.8

The longitude, $\lambda = \lambda_0 + \Delta\theta$ + precession of vernal equinox $= 277.2 - 6.0 + 0.3 = 271.5°$.

We can obtain corrections for the elliptical orbits of the remaining planets in the same way. The calculated anomaly (θ) and longitude (λ) are listed in Table 2. The residual errors in λ values are $\lesssim 1°$.

3.3 Geocentric perspective

Until now, we have calculated the longitudes, λ, of the planets on the celestial sphere centred on the Sun. We can also calculate radii r of their orbits around the sun, for that epoch, giving their positions in polar form (r, λ). To get the positions of planets on the celestial sphere centred on the Earth, we first convert the polar co-ordinates into rectangular form with origin on Sun, and after shifting the origin from Sun to Earth, we change them back into polar form.

For converting into a rectangular form, we have to decide upon the direction of the X and Y axes. We assume X to be in the positive direction along the line joining Sun to the Vernal Equinox, and Y to be perpendicular to X in the ecliptic plane in such a way that the longitude is a positive angle.

3.4 An example

As an example, this procedure is demonstrated for Mercury's position on 01/04/20 at 00:00 UT.

3.4.1 Heliocentric co-ordinates

Distance, r, of Mercury from Sun can be obtained from the anomaly θ as,

$$r = \frac{a(1-e^2)}{1+e\cos\theta} = 0.463\text{A.U.},$$

where $a = 0.387$ A.U. is the length of semi–major axis of its elliptical orbit. Distances to planets in the Solar system are conventionally measured in astronomical units (A.U.), the mean distance of Earth from Sun, which is nearly 1.5×10^8 km.

The heliocentric polar co-ordinates of Mercury thus are (0.463 A.U., 271.5°). Then we can get heliocentric rectangular co-ordinates of Mercury as,

$X_h = r\cos(\lambda) = 0.012$ A.U.

$Y_h = r\sin(\lambda) = -0.463$ A.U.

Similarly we get heliocentric rectangular co-ordinates of Earth as,

$X_0 = -0.978$ A.U.

$Y_0 = -0.204$ A.U.

3.4.2 Geocentric co-ordinates

Geocentric rectangular co-ordinates of Mercury then are

$X_g = X_h - X_0 = 0.990$ A.U.

$Y_g = Y_h - Y_0 = -0.259$ A.U.

Converting these into polar form, we get the geocentric distance and longitude as,

$r_g = \sqrt{(X_g^2 + Y_g^2)} = 1.024$ A.U.

$\lambda_g = \tan^{-1}(Y_g/X_g) = 345.3^\circ$.

In Table 3 we have given the calculated geocentric longitudes of various planets on 01.04.2020 at 00:00 UT. In the Earth/Sun row in Table 3, r_g, λ_g are the geocentric values for the Sun's position. The position of Sun on the celestial sphere, as seen from Earth, is in a direction exactly opposite to that of Earth as seen from the Sun. Therefore the geocentric longitude of Sun is the heliocentric longitude of Earth plus 180°.

We have ignored any perturbations on the motion of a planet due to the effect of other planets which may distort its elliptical path. We are able to get the accuracy of $\lesssim 1^\circ$ for long periods ($\pm$50 years) because most of the terms ignored in the heliocentric longitude calculations are periodic in nature and do not grow indefinitely with time (see e.g. [4]). The other parameters characterizing the elliptical orbit, like the longitude of the perihelion, semi–major axis and eccentricity etc. change so slowly with time that for the accuracy we are interested in, these can be considered constant for $\pm$50 years.

4 Locating planets in sky

After having computed the geocentric longitudes of the planets, we are now in a position to locate them in the sky. Any one familiar with the Zodiac constellations could locate a planet from its position in the constellation in which it lies. The ecliptic is divided into 12 Zodiac signs – Aries, Taurus, Gemini, Cancer, Leo, Virgo, Libra, Scorpio, Sagittarius, Capricorn, Aquarius, Pisces. The Vernal equinox, at zero ecliptic longitude, is the start of the first Zodiac sign and is also known as the First Point of Aries. But there is a caveat attached. Because of the precession the vernal equinox has shifted westward by almost the full width of a constellation in the last $\sim$ 2000 years since when the Zodiac signs and constellation were perhaps first identified. As a consequence, the First Point of Aries now lies in the constellation Pisces. For example, on 01/04/2020, geocentric longitude 294.4° of Jupiter implies it is in the 10th Zodiac sign Capricorn, but actually lies in the Sagittarius constellation, taking into account the shift by one constellation due to precession. There are further complications. The twelve constellations are not all of equal length of arc along the ecliptic longitude. Moreover there is another constellation, viz. Ophiuchus,

through which the ecliptic passes. However these complications are somewhat set aside by the fact that there are only about half a dozen stars in the Zodiac with an apparent brilliance comparable to the naked–eye planets, therefore with some familiarity of the night-sky, one could locate the planets easily from their geocentric longitude values. It further helps to remember that unlike stars, the planets, because of their large angular sizes, do not twinkle.

For a more precise location of a planet we can calculate its relative angular distance from the Sun along the ecliptic. The difference between the geocentric positions of a planet and Sun (Table 4) is called the elongation (ψ) of the planet and it tells us about planet's position in the sky with respect to that of the Sun. The longitude increases eastwards, therefore, if the longitude of the planet is greater than that of the Sun, then the planet lies to the east of the Sun. That means, in the morning the Sun will be rising before the planet rises but in the evening the planet will be setting after the sunset. So the planet will be visible in the evening sky in the west. On the other hand, if the geocentric longitude of the planet is smaller than that of the Sun, then it lies to the west of the Sun and will rise before the sunrise and will be visible in the morning in the eastern sky.

In Table 4, positions of planets with respect to Sun on 01.04.2020 are given for 00:00 UT, which corresponds to 05:30 IST (Indian Standard Time). For example, the geocentric longitude of Venus with respect to Sun is

Table 4: Elongations of the planets on 01/04/2020

Planet	$\lambda_g(^\circ)$	$\psi(^\circ)$	$\lambda_g(^\circ)$	$\psi(^\circ)$
	05:30 IST		(17:30 IST)	
Sun	11.8	-	12.3	-
Mercury	345.3	-26.4	346.0	-26.3
Venus	57.6	45.9	58.1	45.8
Mars	300.9	-70.9	301.3	-71.0
Jupiter	294.4	-77.4	294.4	-77.8
Saturn	300.8	-71.0	300.8	-71.4

$57.6 - 11.8 = 45.9^\circ$. Thus Venus has an eastern elongation 45.9° on 01.04.2020, 05:30 IST, and would be visible only in the evening, after the sunset, in the western sky.

As Earth completes a rotation in 24 hours, the westward motion of the sky is at a rate $360/24 = 15^\circ$/ hour. This rate is strictly true for the celestial equator, but we can use this as an approximate rotation rate even for the ecliptic, which is inclined at 23.5° to the equator. Therefore Mercury, on 01.04.2020, will rise about $26.4/15 \sim 1.5$ hours before the sunrise, and could be seen in the eastern sky in the morning.

It is, however, possible to determine the planetary positions for any other time of the day. For instance, in the case of Venus, which is visible in the evening hours on our chosen date of 1/4/2020, it might be preferable to calculate the position on that date for 12:00 UT, corresponding to 17:30 hr IST, locally an evening time. For this, we use the number of days in Step 1 as 7396.5. It should be noted that not only the longitude of each

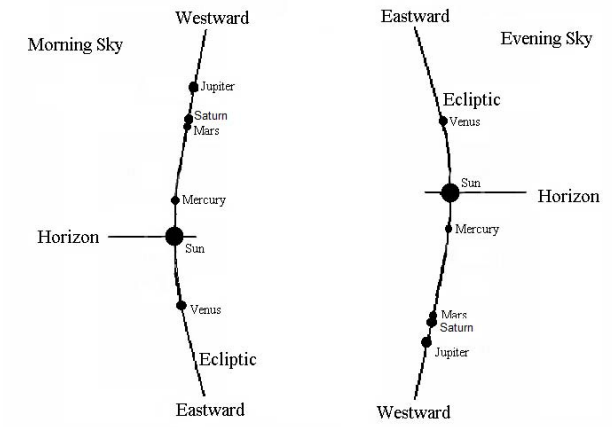


Figure 2: A schematic representation of the elongations of planets at the times of sunrise and sunset on April 1, 2020.

planet around Sun might change by a certain amount, even the longitude of Earth advances by $\sim 1^\circ$ in a day, thus affecting the elongation of even Jupiter and Saturn (Table 4), whose angular speeds are relatively small (Table 1). In Table 4, we have listed the elongations of all five naked-eye planets on 01-04-2020 at 00:00 hr UT (5:30 IST) and at 12:00 hr UT (17:30 IST). Venus with an eastern elongation $\sim 46^\circ$ on that evening, will be setting approximately three hours after the sunset. This way, one can easily locate the planets in the sky from their elongations.

Figure 2 is a schematic representation of the relative positions of various planets with respect to the Sun at the horizon, on the morning and evening of April 1, 2020.

5 Conclusions

It is generally thought that calculation of position of planets in the night sky is a difficult job, which can be accomplished only by complex scientific computations, using fast computers. Our motive here has been to dispel such a notion and bring out the fact that such complex and accurate computations are not always really necessary. One can calculate the position of planets using the method derived here and get the thrill of finding the planet at the predicted position in the night sky.

We have been able to obtain the position of planets within an accuracy of $\lesssim 1^\circ$., using a calculator. This method can be used to reckon planetary positions up to ± 50 years of the starting epoch.

Appendix A: Correction for the orbital ellipticity

We compute the correction for motion of a planet in an actual elliptical orbit from that in an imaginary circular orbit. Period of revolution in circular orbit is taken to be exactly the same as that in the elliptical orbit. The origin of the mean longitude in circular orbit is chosen such that λ_0 coincides with the longitude λ of the planet when it is at the perihelion in its elliptical orbit. For mathematical convenience, we take $t = 0$ at that instant. Then $\lambda_0(0) = \lambda(0)$. Let $\lambda_{\rm p}$ be the longitude of the perihelion of the planet's elliptical orbit. We subtract $\lambda_{\rm p}$ from λ_0 and λ to obtain what is called respectively, the mean anomaly θ_0 and the anomaly θ of the planet. Thus $\theta_0 = \lambda_0 - \lambda_{\rm p}$, and $\theta = \lambda - \lambda_{\rm p}$.

Then $\theta_0(0) = \theta(0)$

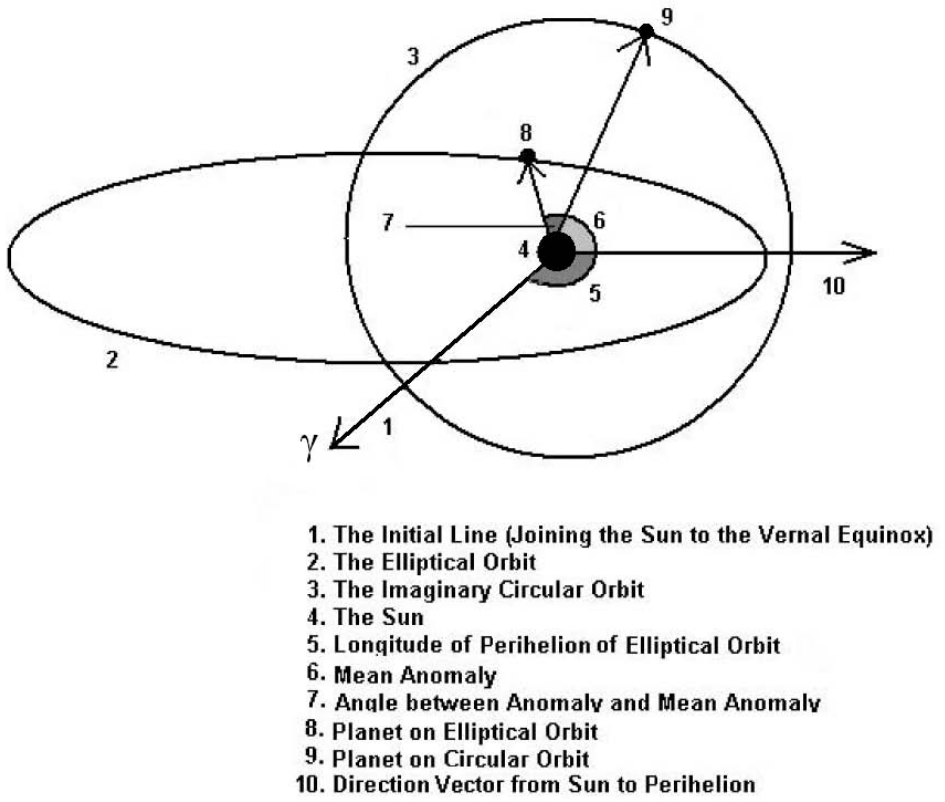


Figure 3: A schematic diagram of the planet in circular and elliptical motion.

In circular motion, the angular speed, ω_0, of the planet is constant. However, in the elliptical motion, the angular speed of the planet varies with time.

Let a time t has passed after $t = 0$. Then, the change in θ_0 of the planet is $\omega_0 t$ whereas the change in θ of the planet won't be the same because of the variation in the angular speed along elliptical trajectory.

Let $\Delta\theta(t) = \theta(t)$ - $\theta_0(t)$.

We know that $\theta_0(t)$ and $\theta(t)$ are periodic by the same time interval, T, as T is the time period of revolution in both the cases (elliptical and circular motion). Hence, all value of $\theta_0(t)$ and $\theta(t)$ repeat after a time period of T. Hence, $\theta_0(t)$ and $\theta(t)$ have a one–to–one relation. Hence, $\Delta\theta$ also repeats after time T. The uniform circular motion thus is a useful approximation because the error $\Delta\theta$ is periodic with time and does not keep accumulating with time to grow to very large values.

To find the correction, first consider an elliptical orbit of a planet around the Sun as shown in Figure 3. We use the equation of the ellipse in polar co-ordinates (r, θ) where θ is the anomaly. The equation of the ellipse then is

$$r = \frac{l}{1 + e\cos\theta} = \frac{a(1 - e^2)}{1 + e\cos\theta}, \qquad (1)$$

where $l = a(1-e^2)$ is the semi–latus rectum with a as the semi–major axis and e the eccentricity of the ellipse. The semi–minor axis of the ellipse is $b = a\sqrt{1-e^2}$.

Now, total area of the ellipse $A = \pi ab$ is swept in T, the time period of revolution. From Kepler's second law we know that the rate of area swept out by the position vector of planet (with respect to Sun) is a constant. Therefore the rate of area swept is

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{r^2}{2}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\pi ab}{T}.$$

Substituting from Equation (1), we get

$$\frac{2\pi}{T} = \frac{(1-e^2)^{\frac{3}{2}}}{(1+e\cos\theta)^2}\frac{\mathrm{d}\theta}{\mathrm{d}t}.$$

We notice that $2\pi/T$ is nothing but the mean angular speed ω_0. Therefore

$$\theta_0(t) = \int_0^t \omega_0\,\mathrm{dt} = \int_0^t \frac{(1-e^2)^{\frac{3}{2}}}{(1+e\cos\theta)^2}\,\mathrm{d}\theta. \quad (2)$$

We want to get the equation in the form, $\theta = \theta_0 + \Delta\theta$, so that by adding the longitude of the perihelion on both sides of the equation, we could get the relation between the correct longitude λ and the mean longitude λ_0.

A direct integration of Equation (2) may not be possible, but we can expand the integrand as a series and integrate only a few first most significant terms. A binomial series expansion is possible because the eccentricity of an ellipse, $e < 1$. During the expansion we drop terms having higher than e^2 factor.

$$\theta_0(t) = \int_0^t (1 - \frac{3}{2}e^2 + \cdots)(1 - 2e\cos\theta + 3e^2\cos^2\theta + \cdots)\,\mathrm{d}\theta$$

After integration we get

$$\theta_0 = \theta - 2e\sin\theta + \frac{3}{4}e^2\sin 2\theta + \cdots, \quad (3)$$

which can be written as

$$\Delta\theta = \theta - \theta_0 = 2e\sin\theta - \frac{3}{4}e^2\sin 2\theta + \cdots$$

However we want the right hand side of Equation (3) to be expressed in terms of θ_0. For that we can substitute $\theta = \theta_0 + \Delta\theta$ on the right hand side to get,

$$\Delta\theta = 2e\sin(\theta_0+\Delta\theta) - \frac{3}{4}e^2\sin[2(\theta_0+\Delta\theta)] + \cdots$$

Expanding in powers of $\Delta\theta$ and neglecting terms of order $e\,(\Delta\theta)^2$, $e^2\,\Delta\theta$ and higher we get

$$\Delta\theta\,(1 - 2e\cos\theta_0) = (2e\sin\theta_0 - \frac{3}{4}e^2\sin 2\theta_0),$$

or

$$\Delta\theta = \frac{(2e\sin\theta_0 - \frac{3}{4}e^2\sin 2\theta_0)}{(1-2e\cos\theta_0)}.$$

Again Expanding in powers of e and ignoring terms of order e^3 or higher, we get

$$\Delta\theta = 2e\sin\theta_0 + \frac{5}{4}e^2\sin 2\theta_0 \quad (4)$$

which is the required expression for the correction term.

Appendix B: Position of the Moon

Here we determine Moon's position for any given epoch, say, April 1, 2020, 00 UT, starting from the initial epoch 1st January 2000, 00 UT.

Moon moves in a *geocentric* orbit of mean eccentricity $e = 0.0549$ and a tropical revolution period of $T = 27.32158$ days, corresponding to a mean angular speed $\omega_0 = 13.17640^\circ/\text{day}$ with respect to the vernal equinox. From the initial value $\lambda_{gi} = 211.7^\circ$ on 1st of January, 2000 A.D., 00:00 UT, Moon's mean geocentric longitude is calculated as, $\lambda_{g0} = 211.7 + 13.17640 \times 7396 = 104.4^\circ$.

But before we correct for the ellipticity of the Moon's orbit, we need to consider that unlike in a planetary orbit where the position of the perihelion change so slowly with time that it can be considered constant for ± 50 years, in Moon's orbit the perigee rotates forward with respect to the vernal equinox with a period of 3231.4 days (~ 8.85 years), corresponding to an angular speed $0.11141^\circ/\text{day}$. With an initial value of 83.3°, the longitude of the perigee, on April 1, 2020, 00:00 UT is then given by, $\lambda_p = 83.3 + 0.11141 \times 7396 = 187.3^\circ$. From this we get Moon's mean anomaly for a circular motion as, $\theta_0 = \lambda_{g0} - \lambda_p = 277.1^\circ$.

We can now use Equation (4) in the same way as for the planets to get the correction for the ellipticity $\Delta\theta = -6.3^\circ$, which gives Moon's corrected geocentric longitude $\lambda_g = \lambda_{g0} + \Delta\theta = 98.1^\circ$. However, the value obtained thus is accurate only up to $\sim 2^\circ$. The reason being that there is an important perturbation term, known as Evection, which depends upon Moon's mean elongation ψ_0 as well as its mean anomaly θ_0, and has a value, $\Delta\theta_{ev} = 1.27^\circ \sin(2\psi_0 - \theta_0)$. Substituting for $\psi_0 = 104.4 - 9.6 = 94.8^\circ$, we get, $\Delta\theta_{ev} = 1.27 \sin(2 \times 94.8 - 277.1) = -1.3^\circ$. Thus a more accurate value of Moon's geocentric longitude on April 1, 2020, 00:00 UT is $\lambda_g = 98.1 - 1.3 = 96.8^\circ$, and the elongation ψ calculated thence is 85.0°.

There is another feature of the Moon's motion which has a direct bearing on the time of occurrences of solar and lunar eclipses. Moon's orbit is inclined to the ecliptic with a mean inclination of 5.16°, intersecting it at points known as the ascending and descending nodes. The ascending node is the one where Moon crosses the ecliptic in a northward direction. Gravitational pull of the Sun causes a precession of the axis of Moon's orbital plane around that of the ecliptic with a tropical period of 6798.4 days ($\sim$ 18.6 years). Consequently the nodes of the Moon's orbit have a retrograde motion of $\sim 0.05295^\circ/\text{day}$ around the ecliptic. With an initial value of 125.1° on 1st January 2000, 00 UT, the longitude of the ascending node for the epoch April 1, 2020, 00:00 UT is $125.1 - 0.05295 \times 7396 = -266.5^\circ$, implying $\psi_\Omega = 81.7^\circ$, while the descending node is 180° away at $\psi_\mho = -98.3^\circ$.

A solar eclipse can take place near the New Moon time, when Moon's elongation is $\sim 0^\circ$ and it could block sunlight to reach some parts of the Earth's surface, while a lunar eclipse can take place around the Full Moon time, when Moon's elongation is $\sim 180^\circ$ and Earth may be blocking sunlight from reaching the Moon. However, these

events could occur only when Sun's longitude is close to that of one of the nodes because it is then only Earth and Moon both get nearly aligned in a straight line with Sun, so that one of these celestial bodies could block sunlight to cause a shadow on the other celestial body.

Appendix C: The astronomical calendar

In all the examples presented above, the calculations were made using a scientific calculator. This procedure is appropriate for quickly getting planetary positions, to locate one or more planets in sky, occasionally, on some specific day. However if one wants to do many computations, say calculate positions for planets for many more days of the year, the process could become tedious and the chances of a numerical mistake occurring in manual calculations could become high. Since the process of computing planetary positions is a repetitive one, it could then be much more convenient to write a simple computer programme, using the algorithm described above, to carry out calculations. We have written such a programme to compute positions of all the planets as well as of the Moon for each day of a specified year and present the results in the form of an astronomical calendar which gives elongations of different planets for all days of the year.

The calendar allows us to locate the naked-eye planets in the sky for any time of the corresponding year. The horizontal axis displays the elongation in degrees, and is centred around the Sun, which by definition has a zero elongation. The vertical axis marks the day of the year. Thus to locate a planet on any given date of the year, we select that date on the vertical axis and then move in a horizontal direction till we find the planet. From the elongation of the planet we can easily locate it in the sky. Different paths of various planets as well as those of Sun, Moon and both nodes, can be identified from their conventional symbols, which are explained at the bottom of the figure. We can use the calendar to find what all planets are above the horizon at any time of the day. Suppose for a given date of the year we want to locate all planets visible in the sky at, say, dawn. The Sun, at 0° elongation, will at that time be just rising near the eastern horizon and the -180° elongation point in the calendar will be near the western horizon. The intermediate elongation points will be at in-between positions on the celestial hemisphere, e.g., the -90° elongation point will be close to the culmination point (the point nearest to the zenith). This way going along the horizontal direction from 0° to -180° at the chosen date, we will find the celestial position of all planets visible in the morning sky on that date. At dusk, with sun setting near the western horizon, the visible sky will stretch eastward from 0° to 180° elongation on the calendar. Similarly one can locate planets on the celestial sphere at other hours of the day. At

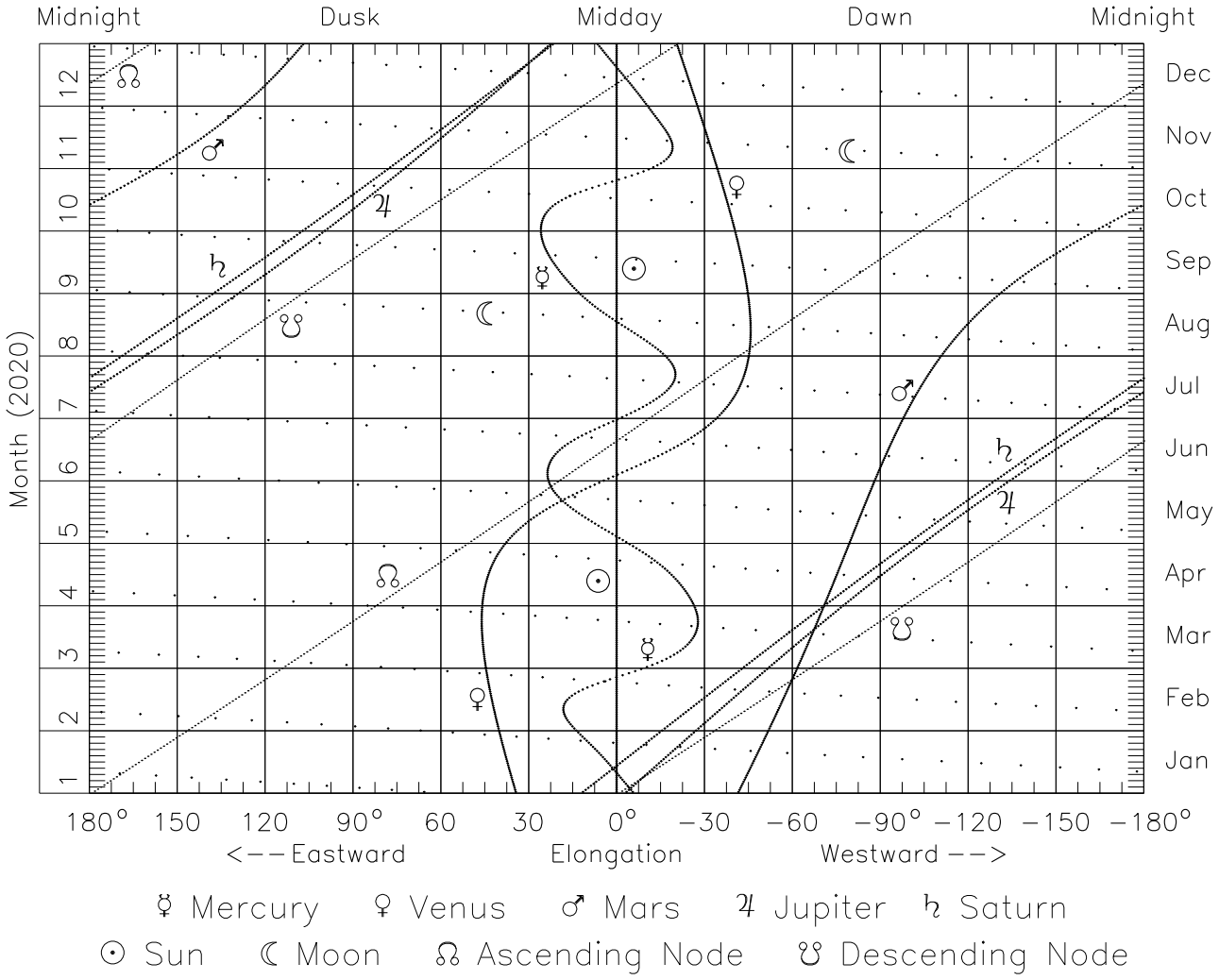


Figure 4: Astronomical calendar for locating planets in sky for the year 2020.

midnight, with culmination point being at 180° (which is the same as −180°), the sky towards west will stretch from 180° to 90° and that towards east will be from −180° to −90° elongation, while at 9 p.m., with the culmination point at 135°, the celestial hemisphere will stretch from west to east between 45° and 225° (equivalently −135°) elongations.

The slant dotted lines running across the calendar from west to east almost every month represent Moon's path. Their dates of intersection with the Sun's path at 0° elongation mean New Moon days, while intersections with the midnight lines (at ±180° elongation) imply Full Moon days, with the intermediate phases at the in-between dates. The faint lines in the calendar represent the relative positions of the ascending node and descending node of Moon's orbital plane. Intersection of the sun's path (at 0° elongation) or of the midnight lines (at ±180°) with

one of the lines of nodes, indicate the possibility of occurrence of an eclipse. In the neighbourhood of these intersection points, at the time of a New moon there might possibly occur a solar eclipse while at the Full moon time there is a possibility of lunar eclipse.

For many planetary phenomena, the astronomical calendar can act as a quick indicator. However, because of the limited resolution of the display in the calendar, for a better accuracy one might go back to the actual tabulated data from which the calendar has been generated. From our computed data for the longitudes of the Sun, Moon and the lines of nodes for the year 2020, we find that the possible dates for the solar eclipses in the year 2020 are 21st June and 14th December, which are the New Moon date close to the intersection point of the lines of nodes with Midday line (Sun's path at 0°). Similarly, possible lunar eclipses on Full Moon dates near the intersection points of the lines of nodes with Midnight lines (at $\pm 180^\circ$) would in 2020 fall on January 10, June 5, July 5 and November 30. However, as seen in Figure 4, these might only be Penumbral Lunar eclipses as the Full Moon positions do not fall very close to any of the intersection points.

From the astronomical calendar (Figure 4), we can expect some interesting astronomical vision in the year 2020. Jupiter and Saturn will be moving closer to each other in sky with the two being closest (within a degree – the accuracy limits of our calculations), an event known as a great conjunction, around 23rd of December. For about 10 days, near the end of March, when Jupiter and Saturn will be within $\sim 7^\circ$ of each other, Mars will be lying in between the two. Not only will there be three bright objects seen in close proximity of each other in sky, Mars, seen within a degree of Jupiter on 22nd of March will move to within a degree of Saturn by 1st of April. But this spectacular sight will be for the eyes of the early birds only, as this would be visible before sunrise, in the eastern sky.

References

[1] Singal, A. K., Phys. Ed. 34, 03, 02 (2018)

[2] http://www.hps.cam.ac.uk/starry/mathtechniqueslrg.jpg

[3] Fränz M. and Harper D., Planetary and Space Science 50, 217 (2002)

[4] Nicholson I., "Unfolding Our Universe", Cambridge University Press (1999)

[5] Simon J. L. et al, Astronomy Astrophysics 282, 663 (1994)

Junction Field Effect Transistor Characteristics: A new approach using ExpEYES-17

Durjoy Roy[1]
[1]Department of Electronic Science,
Rishi Bankim Chandra College,
Naihati, West Bengal 743165, India.
roy.durjoy@gmail.com

Submitted on 19-02-2020

Abstract

Teaching Laboratories in Physics and Electronic Science can be made more interesting using computer assisted data acquisition systems. An indigenous tool, named ExpEYES-17 is developed by the Inter-University Accelerator Center for this purpose, and the author has designed a new approach for a very classical hands-on laboratory experiment using the tool, and programming language python.

1 Introduction

Teaching Laboratories in Physics covers a plethora of experiments. In Electronic Science, there are some more experiments, tailor made for the subject. Conventionally the experiments are carried out using power supplies, analog or digital meters and oscilloscopes. Albeit use of classical modes of experiments have played a pivotal role in enhancing the skill of the student in handling them, a parallel approach of data acquisition method can also open new horizons in learning. The data acquisition methods provide a rapid acquisition of data in comparison with the classical methods, although the classical methods can provide more hands on experience. Combined together, classical method and data acquisition method can expose a graduate level students to best of both the worlds.

The *Inter-University Accelerator Center* (IUAC), New Delhi, is an autonomous research center of the University Grants Commission. ExpEYES-17, a microcontroller based data acquisition system, was designed by a research group in IUAC, and the hardware design is published under the CERN-OHL[1]. CERN-OHL is an Open Hardware Licensing method devised by the CERN (*Conseil Européen pour la Recherche Nucléaire*, or the European Organization for

Nuclear Research) under which one can design a hardware, and keep it open source. The software is distributed under the GNU General Public License, generally called the GNU-GPL. GNU, actually a synonym of an African animal called wildebeest, was a project started by Richard Stallman while he was in the Massachusetts Institute of Technology, where software's and operating systems were free, and mostly open source.

The whole Hardware and software, of ExpEYES-17, in essence, are open source, making the combination an excellent choice for laboratory experiments. The standard user interface of the software contains about fifty experiments to be performed. The author, is in the process of designing some more experiments that are not included in the software, and to make this setup more and more acceptable to the students and teachers of Physics and Electronic Science. In this Article, the author has designed a very popular experiment of graduate level, i.e., Drain characteristics and Transfer Characteristics of a JFET using the ExpEYES-17 hardware as the experiment and data aquisition platform, and writing necessary software code in python.

2 Introduction to ExpEYES-17

The IUAC has conceived the Phoenix Project in 2004. The name of the project was the abbreviation of '*P*hysics with *Ho*me-made *E*quipment & I*n*novat*i*ve E*x*periments'. ExpEYES-17 once again, is the abbreviation of '*Exp*eriments for *Y*oung *E*ngineers and *S*cientists'.

The kit is built around a PIC24EP64GP204 Micro-controller, along with other necessary hardware. A schematic of the complete hardware is available under the CERN-OHL at www.expeyes.in[2], and can be constructed anywhere.

2.1 Hardware of ExpEYES-17

Figure 1 shows the front end of the ExpEYES-17 kit. The kit has three separate connector blocks. One block acts as generator of various signals and voltages. The second block consists of inputs for measurements, and contains a digital storage oscilloscope. The Third Block is for I^2C modules. Here in this experiment, the author did not use the third block, and will limit the discussion within the used blocks only.

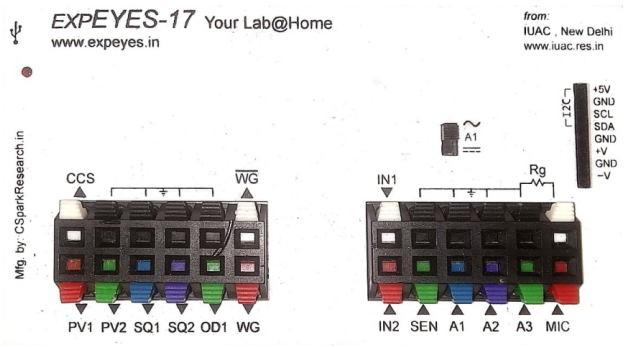


Figure 1: A view of the inputs & outputs of the ExpEYES-17 kit.

The ExpEYES-17 kit provides the user with input/outputs that works with real currents and voltages. It is not a simulator anyway, rather a hardware experiment

and data acquisition system. The kit connects to a Personal Computer (PC) running the ExpEYES-17 user software over Linux or Windows through a USB 2.0 Type-A to Type Mini-B connector. The USB 2.0 interface can transfer data at a speed up to 480 Mbit per second. Another important aspect of the interface is that the kit draws the power required to run, and to supply the components under experiment on the kit, from the PC itself through the USB interface as well. The USB 2.0 standard allows a maximum +5V DC supply voltage, and up to 500 mA DC current to be used by the ExpEYES-17 kit. Figure 2 shows the internal Printed Circuit Board along with the Input/Output Terminals of the ExpEYES-17 kit.

Figure 2: A view of the Internal Circuit Board of the ExpEYES-17 kit.

As the experiment will deal with the characteristics of a JFET, the Bias and input voltages to the JFET are to be taken from the Output (Signal Generator) Block of the ExpEYES kit, and the output of the FET will be fed to the Input (Oscilloscope & Sensors) Block of the kit.

2.2 Input (Signal Generator) Block to JFET:

Here we have two variable voltage sources PV1 and PV2, having supply voltage in the range of (0 to ±5 VDC) and (0 to ±3.3 VDC) respectively. The resolution of the voltage sources are 12 bits, which implies that about 2.5mV steps can be achieved when changing the voltages. Two Square wave generators of Maximum Amplitude of 5 Volts are available via outputs SQ1 and SQ2. One digital Output OD1, one constant current source CCS are also available at the block. Two outputs, WG and $\overline{WG}$ are also there to provide sine and triangular waveforms. The complement over WG signifies an 180° phase inversion. The highest frequency available for the waves in limited to 5 KHz. 80mV, 1V and 3V are the three levels of voltage that are available at WG. All the waves are synthesized. There are four numbers of ground terminals at the signal generator block. All the outputs are controllable using the GUI, or python commands.

In this article, we will only be using the variable DC voltage sources PV1 and PV2 for the experiment involving a JFET. Circuit description of every single signal generation is out of the scope of this article. As the ExpEYES-17 is powered by USB 2.0, it gets only +5VDC supply from the PC. The microcontroller used here operates at +3.3VDC, with a tolerance of ±0.3VDC. +3.3VDC is generated from the +5VDC USB 2.0 connection, and supplied to the microcontroller. The +5V and +3.3V DC Voltages

are then supplied to PV1 and PV2 respectively, through charge pumped DC-to-DC converters to generate the negative voltages as well, built around LM324 OP-AMPs. The value of the voltages are defined by Pulse-width Modulation from the micro controller.

The summary of the signal generators are given in table 1 .

Table 1: Summary of the outputs of the ExpEYES-17 which provides input to the JFET

O/P	Description
CCS	Constant Current Source. Provides about 1.1 mA under software controlled switch.
PV1	Variable voltage source from -5VDC to +5VDC, minimum step size 2.5 mV, controllable through software.
PV2	Variable voltage source from -3.3VDC to +3.3VDC, minimum step size 0.8 mV, controllable through software.
SQ1	Output swings from 0 to 5 volts and frequency can be varied from 4Hz to 5kHz. Duty cycle is programmable.
SQ2	Output swings from 0 to 5 volts and frequency can be varied from 4Hz to 5kHz. Duty cycle is programmable. Not usable when using WG.
OD1	The voltage at OD1 can be set to 0 or 5 volts, using software.
WG	Sine/Triangular Wave generator. Frequency can be varied from 5Hz to 5kHz. The peak value of the amplitude can be set to 3 volts, 1.0 volt or 80 mV. Shape of the output waveform is programmable. Using the GUI sine or triangular can be selected.
$\overline{WG}$	Inverted WG.

2.3 Output (Oscilloscope & Sensors) Block:

The Input block consists of inputs IN1 & IN2. IN1 can measure capacitances, and IN2 can count frequencies up to several Megahertz. SEN input is used to measure resistance. Two analog oscilloscopes channels A1, and A2 are available as inputs, with maximum input voltage ± 16 Volts, and maximum sampling rate of 1 Msps, and input impedance of 1 MΩ. A third oscilloscope channel for precision voltage measurement for voltages ranging within ±3.3 V, and with input impedance of 10 MΩ is available through A3 input. A fourth oscilloscope input, specially configured for use with condenser microphone is available through the MIC terminal. Four ground terminals are available here as well.

In this article, we will only be using the channel A2 for measuring the voltages. Channels A1, A2 and A3 work in almost similar fashion. The input signal at A2 faces a 1 MΩ resistance before getting attenuated by a TL082 OPAMP attenuator, and then, a 6S21 Programmable Gain Amplifier (PGA) is used to control the gain of the input, before feeding it to the micro-controller. Circuit description of every single input channels are beyond the scope of the article.

The summary of the inputs are given in table 2 below.

Table 2: Summary of the inputs of the ExpEYES-17 which senses the output of the JFET

I/P	Description
IN1	Capacitance meter: Capacitance connected between IN1 and Ground can be measured.
IN2	Frequency Counter: Capable of measuring frequencies up to several MHz
SEN	Resistive Sensor Input: meant for sensors like Light Dependent Resistor, Thermistor, Phototransistor etc.
A1	±16V Analog Input: Can measure voltage within the ±16 volts range. The input voltage range can be selected from .5V to 16V full-scale. Voltage at these terminals can be displayed as a function of time, giving the functionality of a low frequency oscilloscope. The maximum sampling rate is 1 Msps /channel.
A2	Same as A1
A3	±3.3V Analog Input: Can measure voltage within the ±3.3 volts range. Capable of displaying very small amplitude signals. The input impedance of A3 is 10 MΩ.
MIC	Microphone input: A condenser microphone can be connected to this terminal and the output can be captured.

The Input and the Output sections are interfaced with a set of push-to-release terminal blocks as shown in Figure 1 & 2, thus eliminating the need of breadboards to conduct simple experiments. However, breadboards can well be used with this kit, with hook-up wires.

2.4 Sofware of ExpEYES-17:

The software of ExpEYES-17 is designed in Python, and C. It consists of an oscilloscope display, occupying most of the interface. At the right side, control sliders are available for PV1, PV2, SQ1, and WG. Signal type can be selected from the dropdown menu

above the WG slider. On the menu bar, about 50 experiments of different levels of Physics Teaching Lab is arranged. 4 Oscilloscope channel controls, namely A1, A2, A3 and MIC are at the sides of the oscilloscope display, including the time base and trigger control. Measurements of Capacitor, Resistor and Frequency are available at the top right. The Interface is shown in Figure 3.

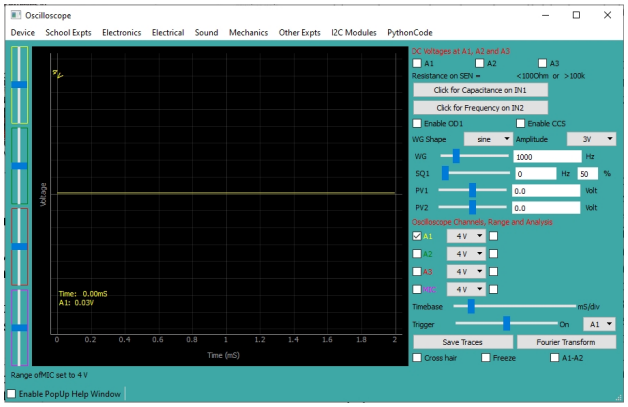


Figure 3: A view of the User Interface (Software) of the ExpEYES-17 kit.

The software is an interface that controls the hardware. It is not a simulator. Instead, it tells the microcontroller how to generate signals, and interpret the inputs, and send it back to the PC. It is also possible to directly use python codes to drive the kit by calling specific libraries that handle the hardware. In this article, the author has used the later approach. The User's Manual for the kit is available under GNU-GPL.[9]

3 Working Principle of a Junction Field Effect Transistor (JFET):

Junction Field Effect Transistors are semiconductor active devices that can control current under the influence of applied electric field. Heinrich Welker first patented JFET in 1940 and ever since, this device remained a popular choice of circuit designers.

In JFET, a region of extrinsic semiconductor, either n type, or p type, called the channel, forms an ohmic region through which a current can pass. Unlike a Bipolar Junction Transistor or BJT, there is no junction across the channel. There are two terminals connected via ohmic contacts at either end of the channel, namely the Source, and the Drain. The current that flows through the channel due to difference of voltages applied at the Source and the Drain V_{DS} is called the Drain Current I_D, and it depends upon the resistivity and the dimensions of the channel, and the V_{DS} as well. This article will limit its discussion within the n-channel JFETs where the channel is made of n-type semiconductor, having electrons as majority carriers, and holes as minority carriers.

In n-channel JFETs, narrow region(s) of heavily doped p-type semiconductors, generally known to be as p+ region(s), and is (are) developed over the side(s) of the channel, depending upon the geometry of the device. Figure 4 shows a schematic of cross sectional view of the geometry of an n-channel JFET. Terminal(s) connected at the

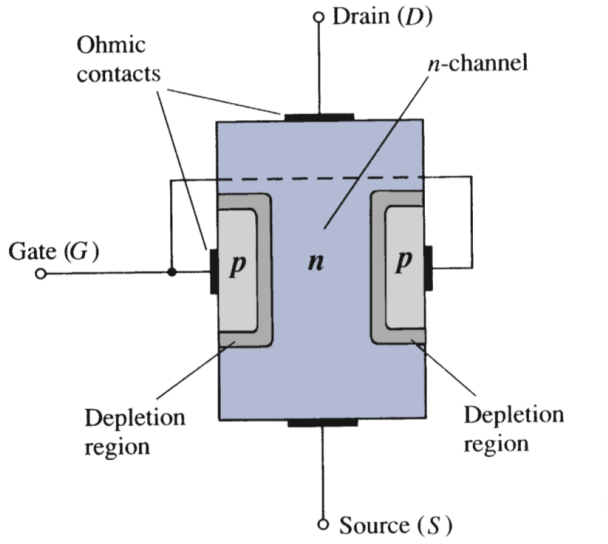


Figure 4: View of the cross section for a cylindrical symmetry geometry of an n-channel JFET (Published from R Boylestad & L. Nashelsky, Electronic Devices and Circuit Theory, 11e, Pearson 2014[3])

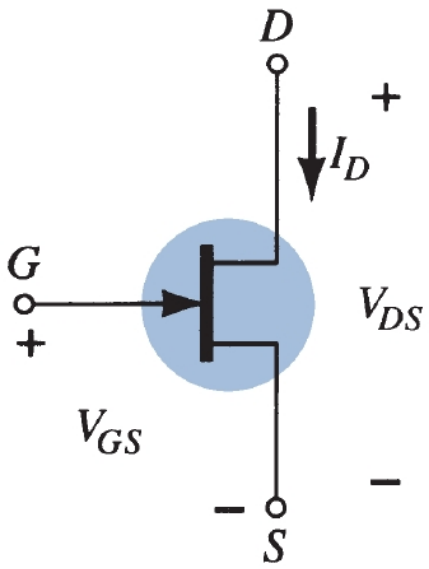


Figure 5: Symbol of a n-channel JFET, showing the terminals, the voltages, and their polarities, and the direction of the Drain Current (Published from R Boylestad & L. Nashelsky, Electronic Devices and Circuit Theory, 11e Pearson 2014[3])

p+ regions is (are) known as the Gate. Figure 5 shows the symbol of a n-Channel JFET, with the terminals, the voltages, and their polarities, and the direction of the Drain Current.

A reverse voltage applied at the Gate terminal with reference to the Source V_{GS}, creates a depletion region at the Channel-Gate p-n junction. The depletion region penetrates more in to the channel region than the gate region due to the difference of doping concentration between the Channel and the Gate. An increment of this reverse voltage V_{GS} results into widening of the depletion region, more into the channel, thereby reducing the effective channel width. As channel shrinks, the available number of majority careers drops, and so does the drain current I_D. No current passes through the Gate-Channel junction except an infinitesimally small reverse current, electric field builds up at the gate with increase of the V_{GS}, and therefore, the device is called a Junction Field Effect Transistor.

An interesting phenomenon observed in a JFET is channel pinch off. Figure 6 shows the model of JFET with biasing arrangement, given by William Shockley. Consider the case where $V_{GS} = 0$, i.e., no voltage is applied across the gate-channel p-n junction. Now, if V_{DS} is increased, the drain voltage V_D will be at higher positive voltage than the source voltage $V_S = 0$, while

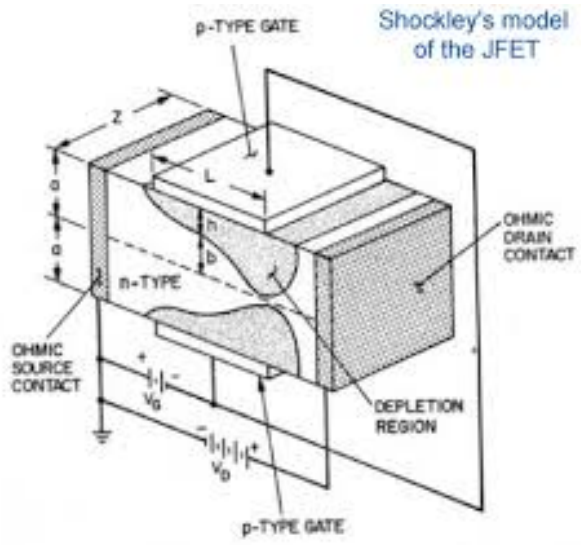


Figure 6: Shockley's model of n-channel JFET, showing the dimensions of the depletion region, and pinch off, along with the required bias voltages (Published from E. F. Schubert, JFETs and MESFETs (Lecture Notes), Rensselaer Polytechnic Institute, 2003)[4])

along the length L of the gate, the gate voltage V_G will remain equal to V_S = 0, considering the case where V_{GS} = 0, i.e., no voltage is applied across the gate-channel p-n junction. Increment of V_{DS} will increase the drain current I_D. As the drain current I_D is now only limited by the drain resistance R_D, the channel is resistive, and the voltage increases linearly along the channel. The increase of current here follows the Ohm's Law. Thus, the depletion region width h will increase with the increase of the V_{DS} near the drain side, and at some value of V_{DS}, it will almost block the channel. The value of the V_{DS} at which the channel is almost pinched off, while V_{GS} =0, is known as the pinch-off Voltage, or V_P. Post pinch off, drain current does not increase, and becomes almost constant. If the V_{DS} is increased well beyond pinch off, the device breaks down.

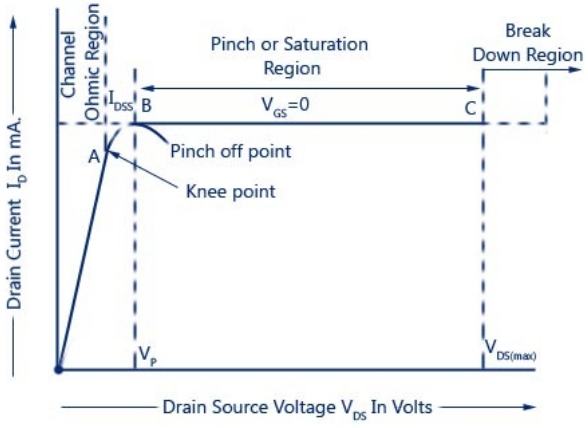


Figure 7: Typical plot of Drain Current vs. Drain Voltage for Shorted Gate, i.e., at V_{GS}=0 of an n-channel JFET (Published from circuitstoday.com[5])

A plot of the Drain Current vs the Drain Voltage for shorted Gate is shown in Fig 7. The Drain current attains its maximum value at pinch off, the value of the drain current at V_{DS}=V_P is referred to as the drain current at zero bias, or I_{DSS}. The current curve follows ohmic slope when V_{DS} is less than V_P, and flattens, or saturates, while V_{DS} is greater than V_P, and around the region where V_{DS}=V_P, it takes a knee-like shape while transitioning from ohmic to saturation region. With higher negative values of V_{GS}, the drain current saturates at lower pinch off voltages, and thus, lower values of IDs and the family of characteristics of V_{DS} vs ID for different V_{GS} are referred to as the drain characteristics of the JFET.

The transfer characteristics for a JFET can be determined experimentally, keeping drain-source voltage, V_{DS} constant and determining drain current, I_D for various values of gate-source voltage, V_{GS}. The transfer characteristic can also be derived from the drain characteristics by noting values of drain current, I_D corresponding to various values of gate-source voltage, V_{GS} for a constant drain-source voltage and plotting them. The transfer characteristics is a plot of I_D vs V_{GS}, and an important parameter of JFET, transconductance, or transferred conductance, the ratio that show the effect of the Gate voltage over the drain current, can be obtained from the transfer curve.

The current equation of a Common-Source n-Channel JFET is given by the Shockley equation

$$I_D = I_{DSS}\left[1 - \frac{V_{GS}}{V_P}\right]^2 \qquad (1)$$

From the above equation, it follows that the smaller the value of V_P, the greater the value of I_D. At V_{GS}=0, I_D has its maximum value.

The Dynamic Drain to Source Channel Resistance, is given by

$$r_d = \frac{\Delta V_{DS}}{\Delta I_D}\bigg|V_{GS} = constant \qquad (2)$$

Being dynamic in nature, this channel resistance depends on the region it is measured. It is lower in the ohmic or linear region, and higher at the channel pinch off or saturation region.

The static Drain to Source Channel Resistance is given by,

$$R_D = \frac{V_{DS}}{I_D}\bigg|V_{GS} = constant \qquad (3)$$

The transconductance g_m is given by

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}\bigg|V_{DS} = constant \qquad (4)$$

It is the main parameter that depicts the working of a Field Effect Transistor, which is a voltage controlled current source. It is the rate of change of the drain current with the change in the gate voltage. As the conductance is transferred from output to input, it is called Transferred Conductance, or transconductance in short.

Multiplying equations (2) and (4), we can get the amplification factor μ as

$$\mu = g_m \times r_d = \frac{\Delta I_D}{\Delta V_{GS}} \times \frac{\Delta V_{DS}}{\Delta I_D} = \frac{\Delta V_{DS}}{\Delta V_{GS}} \qquad (5)$$

μ is a dimensionless entity, and shows the voltage gain achieved by a JFET.

A typical plot showing the drain characteristics and the transfer characteristics of an n-channel JFET is shown in the Figure 8.

It may be noted that a p-channel JFET operates in the same way and have the similar characteristics as an n-channel JFET except that channel carriers are holes instead of electrons and the polarities of V_{GS} and V_{DS} are reversed.

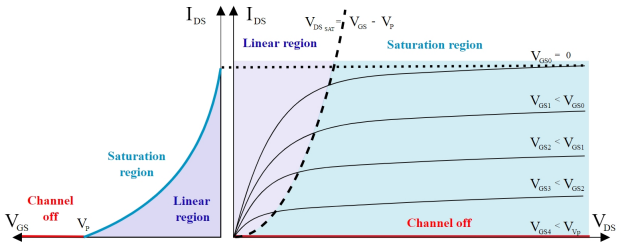


Figure 8: Typical Transfer and Drain Characteristics of a n-channel JFET (Published under creative commons license attributed to CC BY-SA 3.0[6])

4 Experiment Set-up:

A BFW10, a very well-known n-Channel JFET, available in TO-62 package is used for the experiment. The Drain terminal D of the JFET is connected to the variable voltage source PV1 via a resistance of 50 Ω, which is made of two 100 Ω resistances in parallel. ExpEYES-17 cannot measure current directly. This 50 Ω resistance is used to create a voltage drop so that the kit can sense the voltage drop, and find the current I_D by simply dividing the voltage drop by the known resistance. The Source terminal S of the BFW10 is connected directly to one of the ground terminals, and the Gate terminal of the JFET is directly connected to the variable voltage source PV2. Oscilloscope Channel A2 is connected to the intersection of the Drain and the 50Ω resistor to measure the Drain voltage.

The same experiment was repeated with a BFW11, having almost identical property with BFW10, but having lower values of parameters like drain voltage and currents.

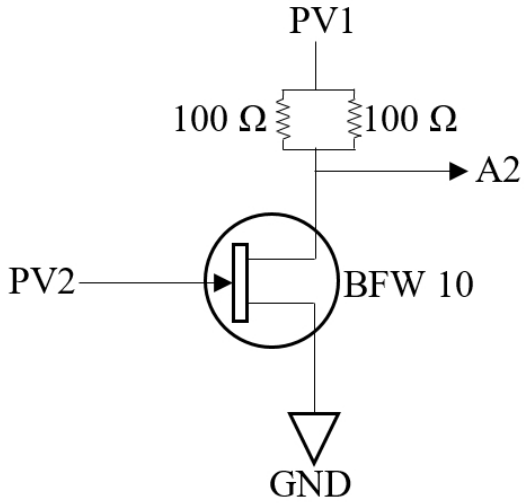


Figure 9: Circuit Schematic of the experiment setup

In this experiment, we did not use the ExpEYEs software. Instead, a python program is written and executed to directly control the voltages, and sense the output of the FET coming through the channel A2. However, there is an option in the software where one can use python code directly.

A schematic of the circuit is given in the Figure 9. Figure 10 shows the actual experiment setup on the ExpEYES-17 kit with a BFW10 n-channel JFET.

The above circuit is implemented on the ExpEYES-17 kit as shown in Figure 10. Breadboard is not used here, as the terminals of the kit were capable enough to hold the circuit.

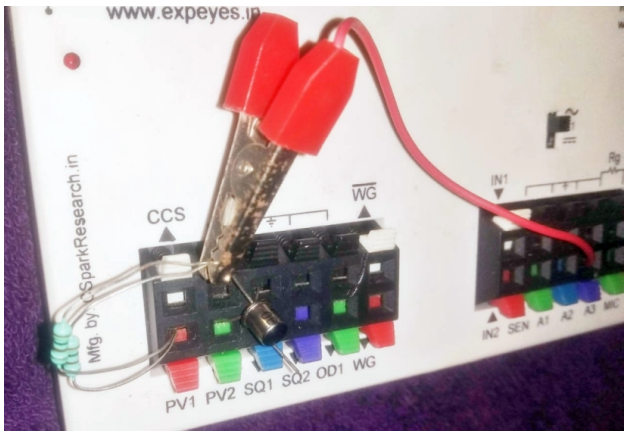


Figure 10: Actual Experiment set-up. The red crocodile clip is biting the resistors and the drain to connect the intersection to A2. The Gate is at PV2 (Green), and the Source is at Ground (Black). The other end of the resistors are connected to PV1 (Red). The Substrate lead is floating.

5 Python Coding:

The software of ExpEYES-17 is written in Python. The author has written the codes for drain characteristics and transfer characteristics. In both the program, `eyes17.eyes` file is called, in which, the controlling mechanism of the interface, and the methods of setting voltages, signals, and reading data are written.

5.1 Drain Characteristics:

In this program, PV2 is varied from 0 to – 3.3 volt to provide gate bias to the JFET. In a nested loop, PV1 is varied from 0 to 5 Volt to provide drain bias. The drain current I_D is found by dividing the voltage drop across the resistor by the expression `idr` $=$ `(PV2 − A2) * 1000/50`, where the factor 1000 is used to convert the current to milli amperes. The data for Drain Current and Drain to Source voltage are stored in two arrays `idra[]` and `vdsa[]` respectively, and plotted using `matplotlib`. The program is given below:

Listing 1: Python Code to Find & Plot n-Channel JFET Drain Characteristics

```
import eyes17.eyes
p = eyes17.eyes.open()
from pylab import *
import time
import matplotlib.pyplot as plt

plt.title('Drain Characteristiocs of n-Channel JFET BFW10')
plt.xlabel('Drain Voltage Vds (Volt)')
plt.ylabel('Drain Current Id (MilliAmp)')

pv2 = 0.0 #set the inital value of vss
ion() #Enable interactive plotting
```

```
while pv2 >= -3.3:
        vdsa = [] #Create array for Vds
        idra = [] #Create Array for Idr
        p.set_pv2(pv2)
        p.set_pv1(0)
        time.sleep(0.2)
        #set the intial value of vdd
        pv1 = 0.0
        while pv1 <= 5.0: #Check value of Vdd
                p.set_pv1(pv1)
                #Measure voltage vd to find current id and voltage vds
                a2=p.get_voltage('A2')
                #Find current id for rd=50 Ohm
                idr = (pv1-a2)*1000/50 #Find Drain Current Id in mA
                vds = a2 #find vds
                #append id & vds in the arrays
                idra.append(idr)
                vdsa.append(vds)
                print(a2,vds, idr)
                pv1 += 0.1
        plt.plot(vdsa,idra) #Plot the graph
        plt.grid(True) # Plot Grid Line
        pause(0.2)
        pv2 -= 0.55 #Next  Vgs
plt.show()
```

The above program also worked for a BFW11. The only change the author made was to edit the title of the plot.

5.2 Transfer Characteristics:

In this program, PV2 is varied from 0 to -3.3 volt to supply reverse bias to the Gate. The value of PV2 is taken as V_{GS}. The Drain current is fond using the same expression used in the previous program, as there is no change in the circuit setup. In a nested loop, the drain to source voltage is made constant at 4.5 volt by using A2 as a check parameter. The program is given below:

Listing 2: Python Code to Find & Plot n-Channel JFET Transfer Characteristics

```
import eyes17.eyes
p = eyes17.eyes.open()
from pylab import *
import time
import matplotlib.pyplot as plt

plt.title('Transfer Characteristics of n-Channel JFET BFW10')
plt.xlabel('Gate Voltage Vgs (Volt)')
plt.ylabel('Drain Current Id (MilliAmp)')

vgsa = [] #define array for Vgs
idra = [] #define array for Idr
ion() #Enable interactive plotting
pv1 = 5.0 #set the initial value of Vdd
pv2 = 0.0 #set the inital value of Vgg
p.set_pv2(0)
while pv2 >= -3.3: #Check The value of Vgs
        pv1 = 5.0 #Set the max value of Vdd first
        p.set_pv2(pv2)
        p.set_pv1(pv1)
        a2 = p.get_voltage('A2') #get the voltage at A2
        while a2 > 4.51:
                p.set_pv1(pv1)
                a2 = p.get_voltage('A2') #Get the latest value of VDS
                idr = (pv1-a2)*1000/50 #Find Id in mA for Rd= 50 Ohm
                vgs = pv2
                pv1 -= 0.01
                idra.append(idr) #Append Idr value to Array
                vgsa.append(vgs) #Append Vgs value to Array
                print(a2, vgs, idr)
                pv2 -= 0.1 #Next Value of Vgs
plt.plot(vgsa, idra)#Plot the Graph
plt.grid(True) #Enable Grid
plt.show()
```

The values of I_D and V_{GS} are stored in two arrays `idra[]` and `vgsa[]` after finding value of each set using the loops, and later, they are plotted using matplotlib.

The above program also worked for a BFW11. The only change the author made was to edit the title of the plot.

6 Results:

The results are obtained as plots as shown, in the Figures 11, 12 and 13 for BFW10, and Figures 14 & 15 for BFW11. It may be noted that these plots are directly obtained running the python codes given in the article. No separate method is applied to record the data and plot the curves.

Results for parameters like dynamic drain resistance, transconductance are calculated from the data available from the plots itself. In this article, calculations are done for BFW10 only.

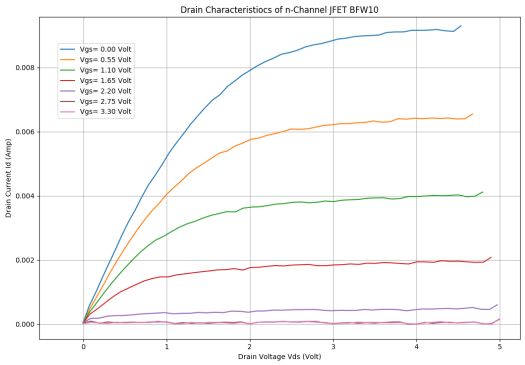


Figure 11: Drain Characteristics of BFW10 for varying V_{GS}

From figure 11, data for finding r_d using equation (2) is taken as in the table 3.

Table 3: Data for finding Drain Resistance

Value of V_{GS} in Volts	Value of V_{DS} in Volts	Value of ID in mA
- 1.1	4.20	3.915
- 1.1	4.40	3.935

From Table 3, we can find from equation (2), the value of r_d as

$$r_d = \frac{(4.40 - 4.20)}{(3.935 - 3.915) \times 10^{-3})} \Big| V_{GS} = -1.1V$$

$$r_d = \frac{(0.20)}{(0.02) \times 10^{-3})} = 10K\Omega \Big| V_{GS} = -1.1V$$

From Figure 13, Transfer Characteristics of BFW10, the I_{DSS} is found to be 8.84 mA, at V_{GS} = 0 Volts, which matches the value range mentioned in the datasheet of BFW10, i.e., 8-20 mA. The Pinch-off voltage V_P at which the drain currents becomes almost zero is at around 2.6 volt, which is also inline with the device parameters. From the plot of Figure 13, we get Table 4

Table 4: Data for finding Transconductance

Value of V_{DS} in Volts	Value of V_{GS} in Volts	Value of I_D in mA
4.5	1.0	4.24
4.5	1.5	2.37

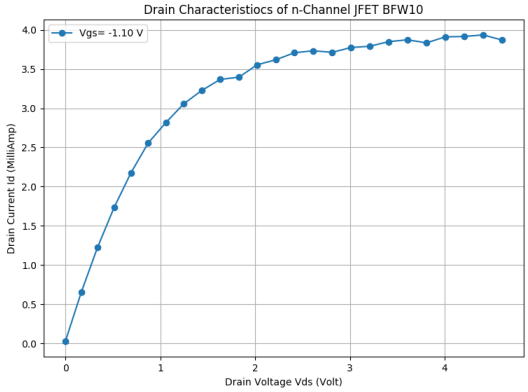


Figure 12: Drain Characteristics of BFW10 for V_{GS}= -1.1 Volt

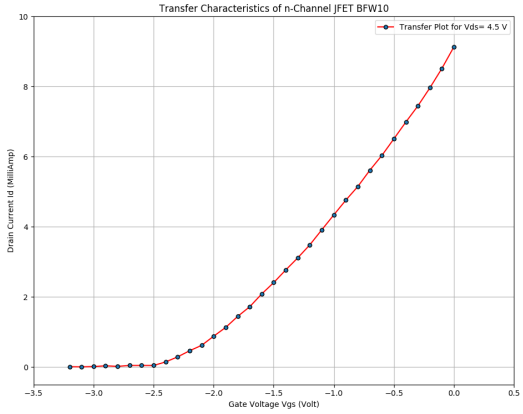


Figure 13: Transfer Characteristics of BFW10.

And from the Table 4, we can find from equation (3) the value of g_m.

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}\Bigg|V_{DS} = 4.5Volt$$

$$g_m = \frac{(1.87)\times 10^{-3}}{(0.5)}\Omega^{-1}\Bigg|V_{DS} = 4.5Volt$$

$$g_m = 3.74m\Omega^{-1}\Big|V_{DS} = 4.5Volt$$

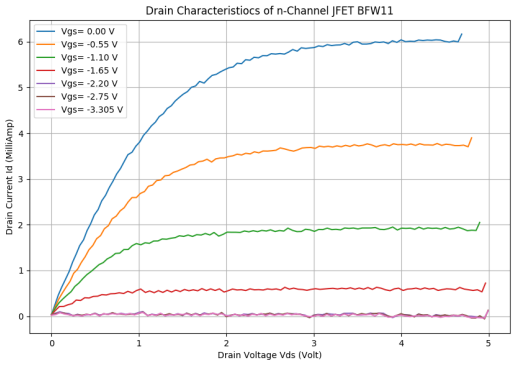


Figure 14: Drain Characteristics of BFW11 for varying V_{GS}.

Figures 14 & 15 show the Drain and Transfer Characteristics obtained using the program, where the BFW 10 is just replaced by the BFW11. It is evident from the plots that the I_{DSS} values of BFW11 is lesser than that of BFW10 at 6.02 mA. The Drain Characteristics of BFW11 are flatter than that of BFW10. From the small signal characteristics specifications of the two devices, it is found that at the same value of V_{DS} the values of the parameters like Zero Gate Volt-

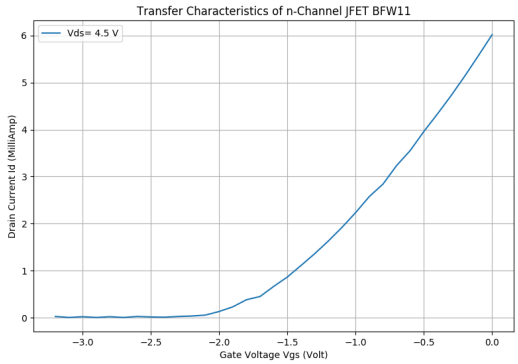


Figure 15: Transfer Characteristics of BFW11.

age Drain Current I_{DSS}, Gate-Source Cut-off voltage $V_{GS(off)}$, Output admittance Y_{OS} etc. for BFW11 is lesser than that of BFW10. These are in line with the differences observed in the characteristics here.

The more saturated nature of the output characteristics of BFW11 suggests smaller value of $Y_{OS} = \frac{\Delta I}{\Delta V}$ to be obtained from the Drain characteristics with reference to BFW10. This observation matches with the datasheet specifications. The values of the I_D in the transfer Characteristics of BFW11 reaches zero at around $V_{GS(off)}$ = - 2.3 Volt, while for BFW10, it is fount that $V_{GS(off)}$ = - 2.6 Volt. This observation remain in-line with the data sheet as well. The difference of values of I_{DSS} at $V_{GS} = 0$ for the devices under test also go with the specifications.

7 Discussions

The experiment, done in a conventional way, includes numerous steps like noting down the readings manually, and drawing the plots. Here, using the ExpEYES-17 kit, everything is done automatically and the students get the opportunity to get familiar with use of Python as a language, and the hardware as well. On the limitations issue, this kit can generate only up to 5 volt of voltages as it draws the require power through the USB connection, and the current generation is limited by the USB 2.0. The Drain Characteristics, thus does not completely reach the active region for JFETs like BFW 10, and thus, the drain resistance appears to be high. However, a very clear idea about automatic data acquisition along with the classical hands-on experiment can be obtained by the method described in this article.

Acknowledgments

The author thank all those who has developed and contributed in design of the kit, Especially, Dr. Ajith Kumar B. P., Satyanarayana V. V. V. of IUAC, and Jithin B. P. of Himachal Pradesh Open University. He is also indebted to his colleague Dr. Subhrajyoti Biswas of Rishi Bankim Chandra College.

References

[1] https://ohwr.org/project/cernohl/wikis/home

[2] www.expeyes.in

[3] R Boylestad & L. Nashelsky, Electronic Devices and Circuit Theory, 11th Ed. (2014) Pearson Education

[4] E. F. Schubert, JFETs and MESFETs (Lecture Notes) (2003), Rensselaer Polytechnic Institute

[5] http://www.circuitstoday.com/characteristics-of-jfets

[6] https://commons.wikimedia.org/w/index.php?curid=7042107

[7] D. Neamen & D. Biswas, Semiconductor Physics & Devices (1934), McGraw Hill

[8] Martin C. Brown Python: The Complete Reference, 4th Ed (2018), McGraw Hill

[9] Ajith Kumar B. P., & V.V.V Satyanarayana, The ExpEyes-17 User's Manual (2017), The Phoenix Project, Inter University Accelerator Center New Delhi

[10] Ajith Kumar B. P., Python for Education (2010), The Phoenix Project, Inter University Accelerator Center, New Delhi

[11] Ajith Kumar B. P. et. al., Innovative science experiments using Phoenix, (2009), Phys. Educ. 44 469

Development of Low-cost Photogate Assembly for Time Measurement

[1]Rohan D. Lahane, [2]Shirish R. Pathare, [2]Bhagyashri G. Latad, [1]Arvind Paranjpye

[1]Nehru Planetarium, Nehru Centre
Dr. Annie Besant Road, Worli. Mumbai - 400 018, India
rohan.lahane@gmail.com

[2]Homi Bhabha Centre for Science Education
V. N. Purav Marg, Mankhurd. Mumbai - 400088
shirish@hbcse.tifr.res.in

Submitted on 03-03-2020

Abstract

Students are introduced to the concepts such as force, motion in school syllabus. They perform experiments like sliding or rolling of an object on an inclined plane or find out the acceleration due to gravity using simple pendulum. The measurement of time in these experiments is done using a hand-held stopwatch. In this paper, we present an inexpensive and portable photogate assembly with a timer that can be used to measure time in different experiments in physics.

1 Introduction

In the physics laboratory, the time measurement in different experiments is done using a stopwatch. In order to achieve consistent observations of time, photogate assembly connected to a data-acquisition system is preferred [1]. These units are often expensive. In this paper, we present an inexpensive and compact photogate assembly that can be developed in a physics laboratory with already available components and instruments. The digital stopwatches with a least count of 0.01s are commonly used for time measurement. Over the period, the start/stop buttons of these stopwatches do not function due to mechanical defects. Such stopwatches can also be used to develop the assembly described in this paper.

2 Description of Assembly

The digital stopwatch needs one pulse to start and another pulse to stop. Generally this is achieved by manually pressing the

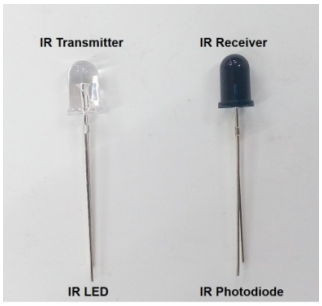


Figure 1: IR LED and Photodiode

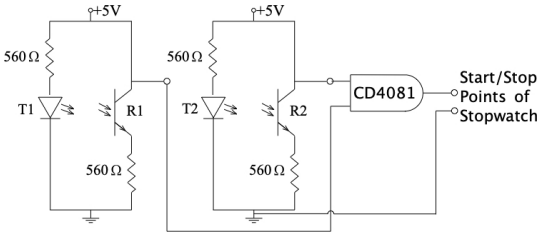


Figure 2: Circuit diagram of the assembly

start/stop button. Internally, with every press, two metal strips are made to touch each other, thereby either starting or stopping the stopwatch. Instead of doing it manually, we connect two pairs of infrared transmitter and receiver (Figure 1) to the contact points of these strips on the printed circuit board of the stopwatch through a CMOS AND gate, CD4081 (Figure 2). The detailed description of this assembly is given in two sections: 1. Photogates and 2. Stopwatch connections.

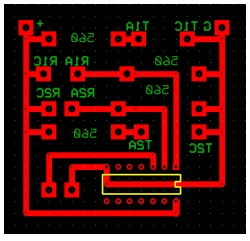


Figure 3: Printed Circuit Board

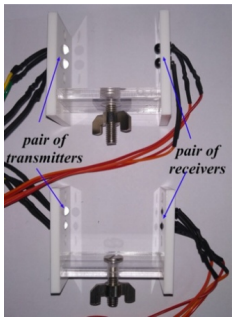


Figure 4: Photogates

1. Photogates: The photogate consists of infrared light emitting diodes (T1 and T2) and the phototransistors (R1 and R2) as transmitter and receiver. Each pair is biased using 560Ω resistor and a 5V supply derived from a power-bank. The output of each pair is fed to the CMOS AND gate from IC CD4081 (pin 1 and 2 of the IC). Figure 3 shows the printed circuit board made using ExpressPCB© software. The photogates were mounted in an U-shaped acrylic mount. The arms of the photogate assembly were fixed with two transmitters and two receivers to increase the area of intersection of the object passing through the photogates.(Figure 4).

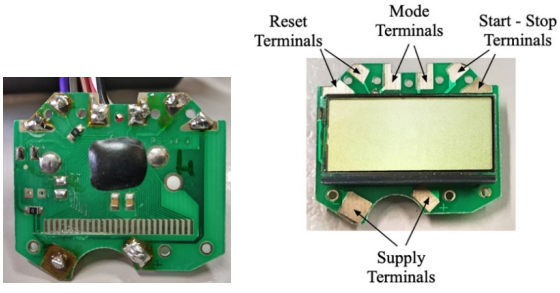


Figure 5: Soldering terminals

2. Stopwatch connections: The output of

Figure 6: Socket box

the IC and the common ground is connected to the start and stop terminals of the stopwatch. Figure 5 shows the printed circuit board of the stopwatch. The button cell of the stopwatch was removed and instead an external 1.5V AA pencil cell was used. Separate push switches were connected to the Mode and Reset terminals (Figure 5).

The printed circuit board and the 1.5V pencil cell are placed in a box. The modified stopwatch is mounted on the box. Sockets for connecting photogates and for 5V supply are fixed to the side-wall of the box. The ON/OFF switch, push buttons for MODE and RESET are fixed to the other side of the wall (Figure 6).

3 Experiment

We describe two experiments in which the photogate assembly can be used for time measurement. The first experiment is about the pendulum in which the period of oscillation is measured and the other experiment is based on study of rolling motion over an aluminum channel.

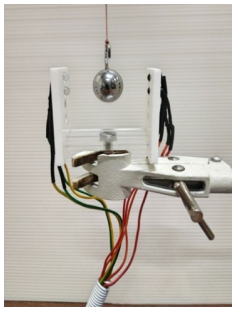


Figure 7: Photogate arrangements for pendulum experiments

1. Experiment 1: For measurement of period of oscillation of pendulum, only one of the two photogates assembly is needed. Figure 7 shows the arrangement of the apparatus. When the pendulum passes through the photogate the stopwatch will start. In its next passage through photogate the stopwatch will stop and we will get time for half oscillation. The period of one oscillation can be calculated from this observed time.

2. Experiment 2: In this experiment, students are introduced to the concept of average velocity and instantaneous velocity. The U channel used for the setup (Figure 8) is made inclined by using the SS bolt attached to it on one side. The pitch of the bolt is 1mm. Using trigonometric ratios, we can calculate the angles of inclination. A steel ball is rolled over the channel. By keeping one photogate at a fixed position from the release point and changing the position of

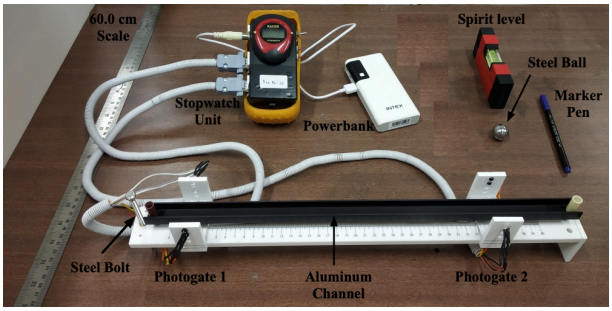


Figure 8: Inclined plane assembly

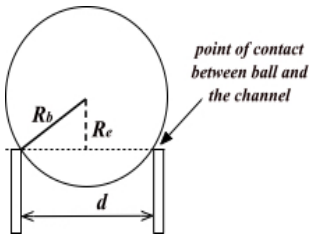


Figure 9: Rolling ball in contact with the U-channel

another photogate will yield the average velocity (Figure 10). The stopwatch will show the time that the ball takes to travel between two photogates.
If both the photogates are kept touching each other and then moved from the release point then this will yield approximately the instantaneous velocity (Figure 11).

Using the data for average velocity, we can get the acceleration of the ball at different angles of inclination. From the acceleration values for different inclination angles, the acceleration due to gravity can be determined by using the following formula [2].

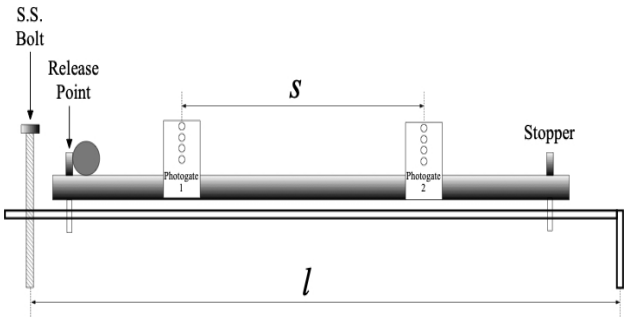


Figure 10: Schematic diagram for measuring the average velocity

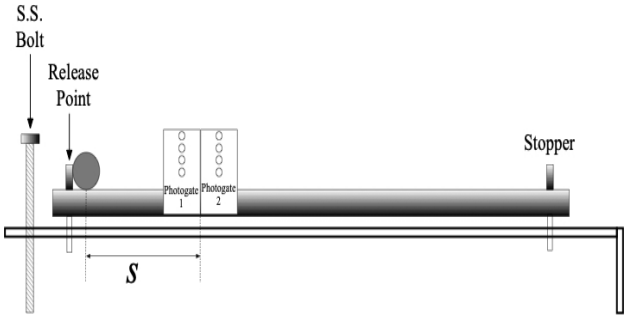


Figure 11: Schematic diagram for measuring the instantaneous velocity

$$a = g\left(\frac{sin\theta - \left(\frac{\rho}{R_e}\right)\left(\frac{R_b}{R_e}\right)cos\theta}{1 + \frac{2}{5}\left(\frac{R_b}{R_e}\right)^2}\right) \quad (1)$$

where, a = acceleration of the ball, g = acceleration due to gravity, ρ = coefficient of rolling friction, θ = the angle of inclination of the inclined plane, R_b = actual radius of the ball, R_e = effective radius of the ball (Figure 9).

4 Observations

In this section, we present the data collected for both the experiments. In experiment 1, we have tested our photogate assembly for consistent readings of time for simple pendulum and compound pendulum. We repeated the observations manually and with photogate for both the pendula.

In the second experiment, we collected the data for the experiment on inclined plane for average velocity measurement. We use equation (1) for calculating the acceleration due to gravity.

Experiment 1: Measurement of period of oscillation

We measured the period of oscillations for simple and compound pendula, manually as well as using the photogate assembly to check the consistency of the observations (Table 1). The last row of the table, ΔT, gives the worst case error in the measured period of oscillation.

Experiment 2: Rolling motion of steel ball

The acceleration of the ball at different inclination of inclined plane is determined by linearization of the kinematic equation of motion as

$$\frac{s}{t} = u + \frac{1}{2}at$$

The data was collected by measuring time t for various distances s at different angles of inclination. The slopes of the lines in the graph (Figure12) for different angles gives the acceleration a of the rolling ball (Table2).

Determination of acceleration due to gravity g:

Equation (1) can be linearized as

$$\frac{a}{cos\theta} = \left(\frac{g}{1+\frac{2}{5}\left(\frac{R_b}{R_e}\right)^2}\right) tan\theta - \left(\frac{g\left(\frac{\rho}{R_e}\right)\left(\frac{R_b}{R_e}\right)}{1+\frac{2}{5}\left(\frac{R_b}{R_e}\right)^2}\right) \quad (2)$$

From equation (2), we can plot the graph (Figure 13) between $\frac{a}{cos\theta}$ and $tan\theta$.

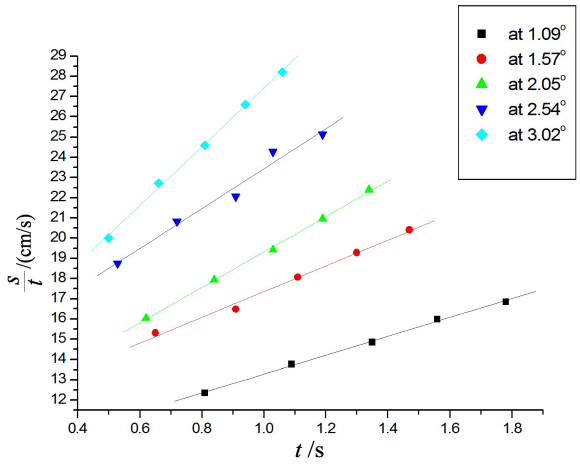


Figure 12: Graph of s/t against t for different inclination of Inclined plane

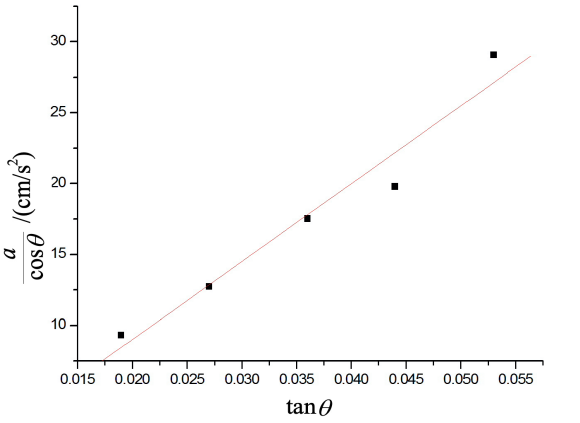


Figure 13: Graph of $a/cos\theta$ Vs $tan\theta$

Simple Pendulum			Compound Pendulum		
Time Period (/s)			Time period (/s)		
No.	Photogate	Manual	No.	Photogate	Manual
1	1.332	1.252	1	1.491	1.409
2	1.334	1.260	2	1.491	1.400
3	1.333	1.256	3	1.491	1.402
4	1.332	1.259	4	1.489	1.408
5	1.331	1.256	5	1.496	1.411
6	1.333	1.263	6	1.496	1.404
7	1.332	1.260	7	1.493	1.417
8	1.329	1.261	8	1.493	1.417
9	1.332	1.264	9	1.500	1.413
10	1.331	1.258	10	1.493	1.412
11	1.333	1.259	11	1.493	1.403
12	1.333	1.263	12	1.487	1.423
13	1.332	1.261	13	1.489	1.402
14	1.334	1.256	14	1.487	1.406
15	1.336	1.262	15	1.485	1.400
ΔT	0.003	0.006	ΔT	0.007	0.012

Table 1: Period of Oscillations for Simple pendulum and Compound pendulum

θ °	a $(/cm/s^2)$	$tan\theta$	$\frac{a}{cos\theta}$ $(/cm/s^2)$
1.09±0.02	9.324	0.019	9.326
1.57±0.02	12.738	0.027	12.743
2.05±0.02	17.516	0.036	17.527
2.54±0.02	19.772	0.044	19.791
3.02±0.02	29.026	0.053	29.066

Table 2: Acceleration at different angles

5 Calculations

$$1 + \frac{2}{5}\left(\frac{R_b}{R_e}\right)^2 = 1.7879$$

$$Slope = \frac{g}{1+\frac{2}{5}\left(\frac{R_b}{R_e}\right)^2} = 551.39 cm/s^2$$

$$g = 551.39 \times 1.7879 = 985.83 cm/s^2$$

$$g = 986 \pm 60 cm/s^2$$

6 Conclusion

The apparatus is very suitable in performing time measurements for different experiments where the measured time interval is more than 0.06s. The advantage of using this assembly is not only the low cost involved

in building it but also the ease of building it. Though the apparatus is low cost, it gives a good repeatability of observations.

Acknowledgments

The authors are thankful to Ms. Bhagyashri Todankar, Ms. Shivani Agre, Ms. Sharvani Shahapurkar, Mr. Avinash Jadhav and Mr. Sadashiv Shirodkar for their valuable inputs and help during the development stages. Authors also wish to thank the physics olympiad program of the Homi Bhabha Centre for Science Education, Mumbai. We acknowledge support from the olympiad program of HBCSE, funded by the Government of India, Department of Atomic Energy, under Project Identification No. RTI4001. Also, we like to put on record the help and support readily provided by our colleagues at the centers.

References

[1] Pathare S.R., Lahane R.D., Sawant S.S., Huli S. (2014) . Low-Cost Timer to Measure the Terminal Velocity of a Magnet Falling Through a Conducting Pipe, The Physics Teacher (AAPT), vol.52 (3), p.160-163.

[2] Doménech A., Doménech T., Cebrián J. (1986) Introduction to the Study of Rolling Friction. American Journal of Physics (AAPT), Vol 55, 231.

Visualising the concept of charging and discharging of capacitors using digital storage oscilloscope

I. Singh[1], B. Kaur[2], K Khun Khun[1], and G. Kaur[1]

[1] P.G. Department of Physics, Khalsa College,
Amritsar-143001, India.
iqbalsgh@yahoo.com,
kamal_minty@rediffmail.com,
gursharanghai@gmail.com,

[2] School of Mathematics,
Thapar Institute of Engineering and Technology,
Patiala-147004, India.
kbikramjeet@yahoo.com

Submitted on 16-03-2020

Abstract

This paper presents a unique idea to visualise the charging and discharging cycle for a set of capacitors connected in parallel in an experimental kit. The variation in voltage with time during the charging and discharging of capacitors is recorded as pulses and can be observed on the DSO screen. The time scale of the recorded data can be expanded for convenient and accurate measurement of pulse width. The method can also be used for the measurement of an unknown value of capacitance of a capacitor instead of manual counting of flashes of neon bulb in the conventional kit.

1 Introduction

Analogue cathode ray oscilloscope (CRO) is an important instrument found in physics laboratories for visualising waveforms and their amplitude, frequency measurements etc. The modified version of this instrument is available in the form of digital storage oscilloscope (DSO) which has an added advantage of auto controls that facilitates data recording and effective signal visualisation. The present paper is an attempt to visualise the process of charging and discharging of capacitors through visual display of voltage variation with time on the oscilloscope screen. The data for

the same can be recorded and stored for future use. The practical method to determine the capacitance of an unknown capacitor in the physics laboratory is normally demonstrated using an experimental kit having set of capacitors connected in parallel along with a neon bulb. The method involves the manual counting of the neon bulb flashes in stipulated time for determining the unknown value of capacitance [1, 2]. The conventional method is simple but fails to provide any visualisation of the concept of charging and discharging of a capacitor. The DSO is used to modify the existing version of experimental setup by connecting it across the capacitor combination.

2 Experimental setup

The experimental kit with capacitors C_1, C_2, C_3 and C_4 (of capacitance 0.11, 0.24, 0.31 and 0.48 μF respectively) are connected in parallel to a capacitor with unknown capacitance (Figure 1a-b). This set of capacitors is being charged by a dc source as shown in Figure 1c. A neon bulb connected in parallel to the circuit is used for counting the flashes of light given during the discharging of the capacitors [2]. A DSO Scientech make (model 401) or any other make, is connected across the capacitors (Figure 1c-d) to analyse the variation of voltage with time during the process of charging and discharging of capacitors. The DSO is also capable of capturing the screenshots and recording the data in a .csv file. The data can be plotted using Microsoft excel/Origin graphical software. A bench type multimeter (Agilent 401A) can also be used to store data in an interfaced computer.

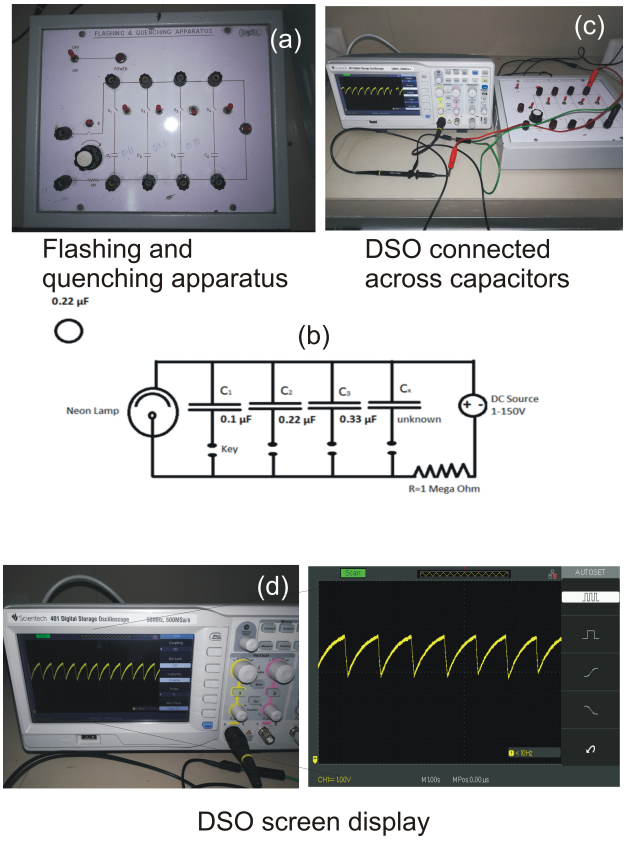


FIG. 1: Complete experimental setup with DSO.

3 Observations and discussion

A set of capacitors connected in parallel are charged through a battery which discharge subsequently by giving a flash of light on the neon bulb. The variation of voltage with time due to charging and discharging of capacitors results in the formation of pulses which can be observed on the DSO screen (Figure 1) and accurately counted from the

recorded data.

1. Figure 2 shows a typical charging/discharging cycle of the capacitor combination having capacitance of 0.96 µF. The complete cycle is marked as ABC. The points on the line joining A to B correspond to the voltage developed across the capacitors during charging. Once the capacitor combination is fully charged, it discharges along the line BC through the neon bulb by emitting a flash of light. This complete cycle of charging and discharging results in the formation of a pulse, the width of which depends on the net capacitance of the circuit. Number of such pulses are generated in a given time which can be counted on the DSO screen. The screenshot for the same is shown in Figure 1d.

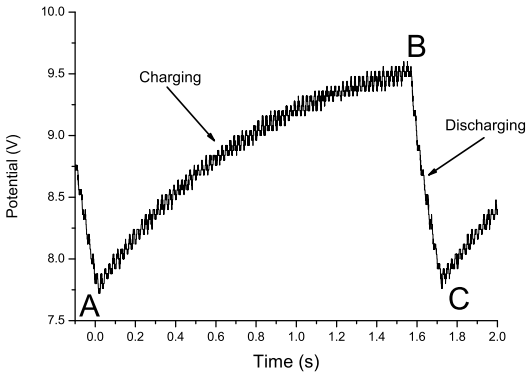


FIG. 2: Charging discharging cycle of 0.96 µF capacitor.

2. The net capacitance of the circuit can be varied by selecting different combinations of the capacitors. As the capacitance of the circuit increases the time taken to charge the capacitors will increase automatically and is shown in Figure 3. This effect is evident from the increase in width of the pulse. The screenshot of the data recorded in a .csv file is shown in Figure 4.

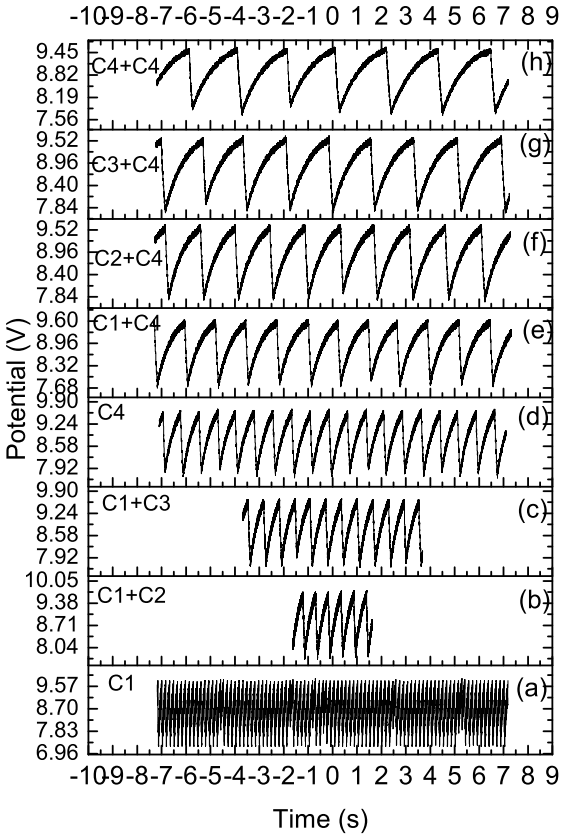
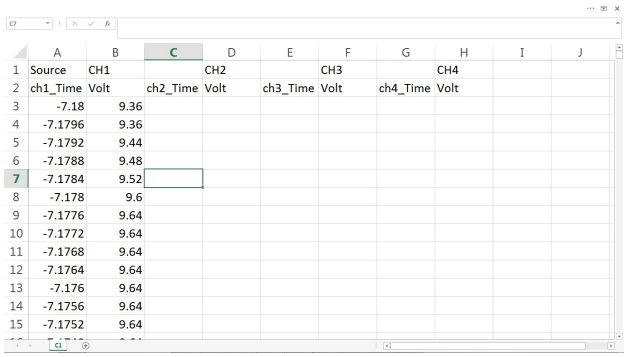


FIG. 3: Charging discharging cycles of various combination of capacitors connected in parallel having net capacitance (a) 0.11 µF, (b) 0.35 µF, (c) 0.42 µF, (d) 0.48 µF, (e) 0.59 µF, (f) 0.72 µF, (g) 0.79 µF and (h) 0.96 µF.

3. When the capacitance in the circuit is as low as 0.11 µF (Figure 3a), it is difficult

to count the flashes of the neon bulb accurately as required in the experimental procedure. However the DSO can be used to overcome this problem as it can count the exact number of pulses by expanding the time scale in the recorded data as shown in Figure 5. The



	A	B	C	D	E	F	G	H	I	J
1	Source	CH1		CH2		CH3		CH4		
2	ch1_Time	Volt	ch2_Time	Volt	ch3_Time	Volt	ch4_Time	Volt		
3	-7.18	9.36								
4	-7.1796	9.36								
5	-7.1792	9.44								
6	-7.1788	9.48								
7	-7.1784	9.52								
8	-7.178	9.6								
9	-7.1776	9.64								
10	-7.1772	9.64								
11	-7.1768	9.64								
12	-7.1764	9.64								
13	-7.176	9.64								
14	-7.1756	9.64								
15	-7.1752	9.64								

FIG. 4: Screenshot of the data recorded in an excel spreadsheet (.csv) for 0.11 µF. This is a JPG graphic.

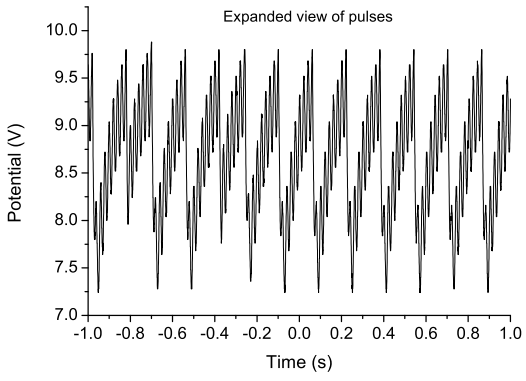


FIG. 5: Pulses due to charging and discharging of a capacitor with capacitance 0.11 µF on an expanded time scale.

above procedure to record the pulses using DSO can be implemented as an improved version of the existing flashing and quenching experimental setup used for the determination of unknown capacitance of a capacitor.

4 Conclusion

The paper presents the effective use of DSO to visualise and record the instantaneous variation of voltage across a set of capacitors connected in parallel during the charging and discharging process. In comparison to the analogue CROs, the inbuilt auto control functions of DSO makes it easier to record and analyse the signal data in any experiment.

5 Acknowledgment

Authors wants to acknowledge Dr. Mehal Singh, Principal, Khalsa College Amritsar, India for their encouragement and continuous support.

References

[1] T. Singh, G. Kaur, I Singh (2015). Practical Physics Vol-I 1sted ISBN-978-93-82956-68-6 (Vishal Publications, Jalandhar, India)

[2] C. L. Arora (2010). B.Sc. Practical Physics ISBN-9788121909099 (S Chand Publications, New Delhi, India)

Understanding Electrostatic Potential and Field Distributions by Simulation

Pragati Ashdhir, Aishwary Jai Singh and Chirag Tutlani

[1]Department of Physics, Hindu College, University of Delhi, Delhi-110007,India
pragatiashdhir@hinducollege.ac.in

aishwaryjai1999@gmail.com chirag161198@gmail.com

Submitted on 15-12-2019

Abstract

Electrostatic field and potential distribution in space due to different charge configurations are computed and simulated. The aim is to aid the teaching-learning of these concepts at the high school and introductory undergraduate level. The finite difference method is used to numerically solve Laplace and Poisson equations subject to boundary conditions appropriate for a given charge configuration. Charge configurations in the form of a single point charge, multiple point charges, a dipole, a quadrupole and a line charge have been discussed. Scilab, an open source computational software has been used for numerical simulations in the work.

1 Introduction

Electrostatics is the branch of Physics that deals with the phenomena arising due to stationary electric charges. The study of electrostatics has several real-world applications like the Van de Graff generator, photoconductor, xerography, laser printer, ink-jet printer and the electrostatic precipitators in air cleaners. The introductory concepts in electrostatics, namely electric field, potential and lines of force are rather abstract when introduced to the students at the high school level. A sound understanding of these concepts is essential for the future study of electromagnetism. The aim of the present work is to aid the teaching-learning of these concepts by means of simulation. The electrostatic field and potential distribution due to planar charge configurations in the form of single point charge, multiple point charge, a dipole, a quadrupole and line charge have been computed and simulated. A step by step simulation of the principle of superposition is done for a better understanding of the same. The electrostatic field representation in the form of the lines of force has been illustrated and discussed. The visualization of equipotential curves and surfaces

for a better understanding of the underlying physical concepts has also been attempted in the work. Scilab, an open source computational software has been used for the numerical simulations.

2 Theoretical Background

2.1 Coulomb's Law of Force:

The presence of electric charges is made known by the existence of attractive or repulsive forces between them. The interaction between electric charges at rest is described by Coulomb's law of electrostatics which states: Two stationary point electric charges repel or attract one another with a force proportional to product of the magnitude of charges and inversely proportional to the square of the distance between them. The statement can be written compactly in vector form as

$$\vec{F_{21}} = \frac{kq_1q_2}{r_{21}^2}\hat{r_{21}}, \tag{1}$$

where q_1 and q_2 are scalars giving the magnitude and polarity (positive or negative) of the respective charges, $\hat{r}_{21}$ is the unit vector in the direction from charge 1 to charge 2 and $\vec{F}_{21}$ is the force acting on charge 2 due to charge 1. k is the proportionality constant equal to $\frac{1}{4\pi\epsilon_o} = 9 \times 10^9 Nm^2C^{-2}$, where ϵ_o is the "permittivity" of free space. Equation (1) expresses the fact that like charges repel and unlike charges attract each other and that the electrostatic force is Newtonian, i.e., $\vec{F}_{21} = -\vec{F}_{12}$. The unit vector $\hat{r}_{21}$ is indicative of the fact that the force acts along the line joining the two charges. By the term 'point charge' is meant a charge whose spatial dimensions are negligibly small as compared to the distance of the charge from other charges in the given region of space.

2.2 Principle of Superposition:

For a system of many charges the force between any two charges is independent of the presence of other charges. This is the basis of Superposition Principle which is of fundamental importance in electromagnetism. According to this principle, the force on a charge in a given system of charges is the vector sum of Coulombic forces acting on it due to the other charges. Mathematically, it may be written as,

$$\vec{F_{ij}} = k\sum_{i\neq j}\frac{q_iq_j}{r_{ij}^2}\hat{r_{ij}}, \tag{2}$$

where the summation extends over all the charges $q_1, q_2, q_3, \ldots$ in the given system except the i^{th} charge on which the force is to be determined. A true understanding of Eqn.(2) gives insight into a great deal of the physics involved in understanding the electrostatic interactions and the versatile applications of the same.

2.3 Electrostatic Field:

The concept of electrostatic field is merely a way of describing a given system of fixed charges. It does so by giving the force in magnitude and direction that a unit positive charge would experience while exploring the region in vicinity of the charge dis-

tribution. In Eqn. (2) taking q_i as a unit positive charge, the electric field $\vec{E}$ due to the charge distribution at location of the point charge q_i can be written as

$$\vec{E}_i = k \sum_j \frac{q_j}{r_{ij}^2} \hat{r_{ij}}, \tag{3}$$

where $\hat{r_{ij}}$ is the unit vector of variable direction pointing from q_j towards the point in space where the unit test charge q_i is located, i.e., the point at which $\vec{E}_i$ is to be determined.It is also to be noted that electric field is a vector quantity and is always finite (or zero) at points in space that are away from the immediate vicinity of any of the charges causing the field. For points too close to the source charges, $\vec{E}_i$ tends to infinity since the distance r_{ij} in $1/r_{ij}^2$ tends to zero. In the study of electrostatics such points of singularities are thus needed to be avoided.

A vector plot can be used to visualize an electrostatic field in a given region of space. In this plot the magnitude and direction of $\vec{E}_i$ at various points in space is indicated by drawing little arrows near those points, making the arrow longer where the magnitude of $\vec{E}_i$ is larger and the head of the arrow pointing in the direction of the field. Such a representation of electric field is however not very useful if the range of magnitude of $\vec{E}_i$ is large in the given region of space because an accurate scaling of the magnitude of the field with the arrow length becomes difficult in such situations. The vectorial representation of electric field due to some of the charge configurations are shown later in the work.

2.4 Electrostatic Lines of Force

The visual depiction of electric field can also be done by drawing the lines of force or the field lines. These are simply curves in spatial domain whose tangents, at any point, lies in the direction of field at that point. Such curves are smooth and continuous except at singularities such as point charges or where field is zero. A field line plot does not directly give the magnitude of the field but does indicate the variation in the strength of field in a given region of space. The field lines converge (become closely spaced) in the region of strong field and diverge (become sparsely spaced) in the region where the field is weaker. These lines of force originate only on positive electric charges (source) and end only on negative electric charges (sink), elsewhere, they are continuous. The number of lines originating and terminating on charges is proportional to the magnitude of charges. Thus,the lines of force give a good qualitative description of the electrostatic field distribution. The lines of force have been plotted for some of the charge configurations depicting these features.

2.5 Electrostatic Potential

A deeper study of electrostatics in terms of forces and fields tends to become quite clumsy and cumbersome because of the vector nature of these physical quantities. The problem is overcome by an equivalent scalar formulation of the ideas related to the elec-

trostatic forces and fields in terms of a physical quantity called the Electrostatic Potential.

The electrostatic potential $V(x,y,z)$ associated with electrostatic field $\vec{E}(x,y,z)$ is defined as the work per unit charge required to be done to bring the unit positive test charge in from infinity to the point (x,y,z) in the field of a given charge configuration. It the is the conservative nature of static electric field that allows for an alternative way to obtain the $\vec{E}$ field apart from using Coulomb's law (Eqn.3). In this approach, if the electrostatic potential distribution $V(x,y,z)$ is known in a given region of space,then at a given point, say (x,y,z), the maximum rate of spatial variation of potential and the direction in which it takes place gives the corresponding electrostatic field at that point. The maximum rate of change of $V(x,y,z)$ at a given spatial point is called the gradient of the potential at that point. Mathematically, the relation between the electric field and the electric potential at a given point in space can be written as

$$\vec{E} = -\vec{\nabla}V. \tag{4}$$

Equation (4) in words states that electric field is negative of the gradient of potential. The minus sign is there because the electric field points from a region of positive potential towards a region of negative potential, whereas, the vector $\vec{\nabla}V$ is defined so that it points in the direction of increasing V. The $\vec{\nabla}$ in Eqn.4 is a vector differential operator which in terms of Cartesian coordinates is given by

$$\nabla \equiv \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k}.$$

The surfaces in space for which $V = constant$ are called Equipotential surfaces. The projection of these surfaces on a plane are the Equipotential curves. Such curves corresponding to certain charge configurations have been depicted in the subsequent part of the work.

2.6 Poisson's and Laplace's Equations

The potential formulation of electrostatics in which a vector problem is reduced to a scalar one is described by the Poisson's and Laplace's equations.The Poisson's equation in Cartesian coordinates is given by:

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\rho(x,y,z)}{\epsilon_o}, \tag{5}$$

where $\rho(x,y,z)$ is the charge density at a point (x,y,z) and ϵ_o is the permittivity of free space. For a region of space containing no charge, Eqn. (5) takes the form

$$\nabla^2 V = 0, \tag{6}$$

which is the Laplace's Equation. From Eqn.(5) and Eqn.(6) it can be seen that Laplace's Equation gives the way in which the potential must behave in a region containing no charge, and the Poisson's Equation does the same for a region in which there is a charge distribution. The operator $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ in the above equations is the Laplacian operator in Cartesian coordinates.

Further, it is to be noted that both the Poisson's and Laplace's equations are linear partial differential equations and hence the principle of superposition holds good for the solutions of these equations. If $V_i(i = 1, 2, \ldots n)$ are the solutions of Eqn.(5) or Eqn.(6) then

$$V = \sum_i V_i \tag{7}$$

is also the solution of the given equations. Thus, the superposition principle is also obeyed by electrostatic potential in a manner similar to electrostatic forces and fields.

3 Computational Background

3.1 Computing the Electrostaic Potential

The planar charge configurations considered in the present work can be assumed to lie in the $xy-$ plane. The corresponding Poisson's and Laplaces's equations in two dimensions can be written as

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = -\frac{\rho(x,y)}{\epsilon_o} \tag{8}$$

and

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \tag{9}$$

respectively. Using appropriate boundary conditions Eqn.(8) and Eqn.(9) have been solved using the finite difference method. The boundary conditions can be expressed as

$$V(x,y) = g(x,y), \tag{10}$$

where x and y variables have values corresponding to the boundaries of the region in which the solution to Eqn.(8) or Eqn.9 is sought.The region is divided into a square mesh or grid with equal step lengths in the x and y directions as

$$h = \Delta x = k = \Delta y = \frac{L}{n}, \tag{11}$$

where L is the length of the sides of the grid and there are $(n+1)$ grid points in both directions. Using the forward difference approximation, the discretized version of Eqn.(8) can be written as

$$V_{ij} = \frac{V_{i,j+1} + V_{i,j-1} + V_{i+1,j} + V_{i-1,j}}{4} + \frac{h^2}{\epsilon_o}\rho_{ij}. \tag{12}$$

and that of Eqn.(9) as

$$V_{ij} = \frac{V_{i,j+1} + V_{i,j-1} + V_{i+1,j} + V_{i-1,j}}{4}. \tag{13}$$

The indices i and j take up values $2, 3 \ldots n$. The discretization of the boundary conditions in Eqn.(10) can be expressed as

$$V_{i,1} = g_{i,1}, 2 \leq i \leq n, \tag{14}$$

$$V_{i,L} = g_{i,L}, 2 \leq i \leq n, \tag{15}$$

$$V_{1,j} = g_{1,j}, 2 \leq j \leq n \tag{16}$$

and

$$V_{L,j} = g_{L,j}, 2 \leq j \leq n. \tag{17}$$

With $(n+1)$ mesh points, the equations for V result in a system of $(n-1)^2$ linear equations in the $(n-1)^2$ unknown V_{ij}. Gauss Siedel iterative method has been used to solve these linear equations.

To simulate the fields for sources of charge far from spatial infinity, the value of potential is taken to be zero at all points

along the grid boundaries. By confining the source charge within a "grounded boundary" in this manner we "move" spatial infinity in to a finite location, i.e., the grid boundary acts like a virtual infinity for the given system of charges. Thus, the fields near the sources are made to approximate the fields for sources far from spatial infinity.

The CONTOUR2D and PLOT3D features of SCILAB have been used to plot the equipotential curves and to depict the "potential surfaces" respectively. In PLOT3D command the value of the computed potential is plotted along the z-axis as a function of the x and y coordinates of the respective grid points.

3.2 Computing the Electrostatic Field

To compute the electrostatic field at the grid points, Eqn.(4) is written in component form as

$$E_x = -\frac{\partial V(x,y)}{\partial x} \tag{18}$$

and

$$E_y = -\frac{\partial V(x,y)}{\partial y}. \tag{19}$$

The discretization of the above equations using the forward difference approximation yield

$$E_x(i,j) = -\left[\frac{V(i+1,j) - V(i,j)}{h}\right] \tag{20}$$

$$E_y(i,j) = -\left[\frac{V(i,j+1) - V(i,j)}{k}\right] \tag{21}$$

where $E_x(i,j)$ and $E_y(i,j)$ are the computed x and y components of the electrostatic field at the $(i,j)^{th}$ grid point. The CHAMP plot feature of SCILAB has been used to obtain the vector plot of the computed electrostatic field distribution.

3.3 Plotting the Electrostatic lines of force

The plotting of electrostatic lines of force is done for co-planar point charge configurations using the definition of electrostatic field given by Eqn.(3). If the coordinates of the test charge are (x,y) and that of the source charge are (x_s, y_s) then using Eqn.(3) we can write

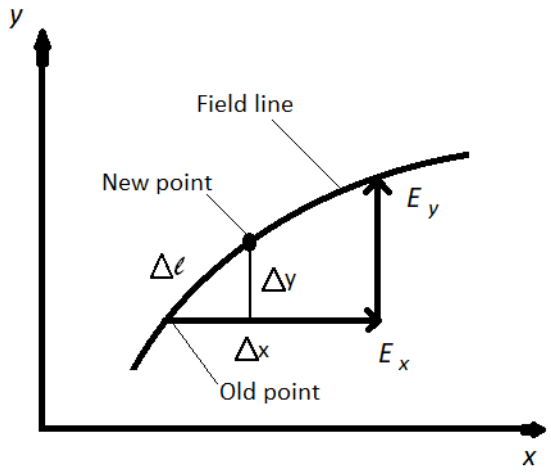


Figure 1: Plotting a field line

$$E_x = k\sum_s \frac{q_s(x - x_s)}{[(x-x_s)^2 + (y-y_s)^2]^{\frac{3}{2}}} \tag{22}$$

and

$$E_y = k\sum_s \frac{q_s(y - y_s)}{[(x-x_s)^2 + (y-y_s)^2]^{\frac{3}{2}}} \tag{23}$$

for the x and y components of the electrostatic field at the point (x,y). Next, if (x,y)

and $(x+\Delta x, y+\Delta y)$ are two closely spaced points on a field line as shown in figure 1, then the distance between the two points along the field line can be written as $\Delta l = \left[(\Delta x)^2 + (\Delta y)^2)\right]^{\frac{1}{2}}$. From Figure 1 it can be easily seen that

$$\frac{\Delta x}{\Delta l} = \frac{E_x}{E} = \frac{E_x}{\left[(E_x)^2 + (E_y)^2\right]^{\frac{1}{2}}} \quad (24)$$

or

$$\Delta x = \frac{(\Delta l)E_x}{\left[(E_x)^2 + (E_y)^2\right]^{\frac{1}{2}}} \quad (25)$$

and

$$\frac{\Delta y}{\Delta l} = \frac{E_y}{E} = \frac{E_y}{\left[(E_x)^2 + (E_y)^2\right]^{\frac{1}{2}}} \quad (26)$$

or

$$\Delta y = \frac{(\Delta l)E_y}{\left[(E_x)^2 + (E_y)^2\right]^{\frac{1}{2}}} \quad (27)$$

The algorithm used for plotting the lines of force is as follows:

Step 1: A starting point (x,y) is chosen on the field line to be plotted.

Step 2: E_x and E_y are computed at (x,y) using Eqn.(22) and Eqn.(23) respectively.

Step 3: A new point $(x^{'}, y^{'})$ is chosen on the field line such that $x^{'} = x + \Delta x$ and $y^{'} = y + \Delta y$, where the step sizes Δx and Δy are determined using Eqn.(25) and Eqn.(27) respectively.

Step 4: It is checked whether the new point is within the given range of coordinates, i.e., the point $(x^{'}, y^{'})$ lies within the grid boundaries.

Step 5: It is also checked that the new point $(x^{'}, y^{'})$ is not close to any of the point charges in the given configuration to avoid the singularities in $\vec{E}$.

Step 6: Steps 2 through 5 are repeated to generate new points until a line is completed within the given range of x and y.

Step 7: The generated points are plotted using PLOT2D feature of SCILAB.

Step 8: Another starting point is chosen and the algorithm is repeated to plot another line of force.

4 Simulated Fields and Potentials

4.1 Single Point Charge or Monopole

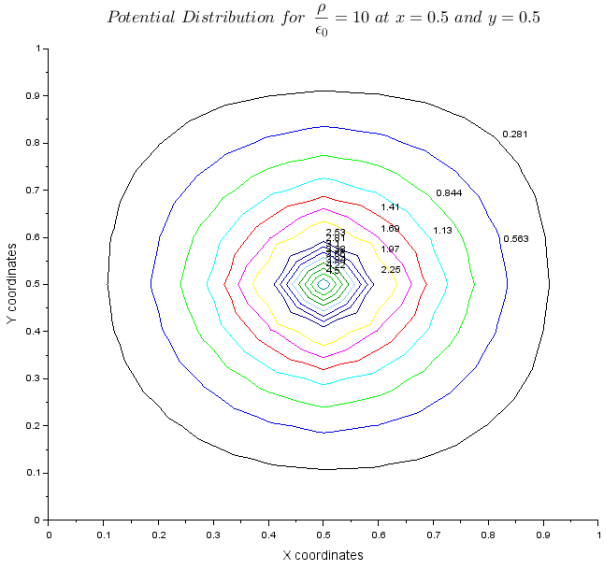


Figure 2: Contour plot of the potential distribution due to a monopole in the form of a positive point charge

To begin with, the electrostatic potential and field due to a positive point charge of strength $\frac{\rho}{\epsilon_o} = 10$ placed at the geometric center of a plate grounded on all sides have been simulated. Figure 2 is the contour plot of the potential distribution depicting the equipotential curves.Figure 3 is the zoomed version of figure 2 showing the details of potential distribution in the central region of the plate.

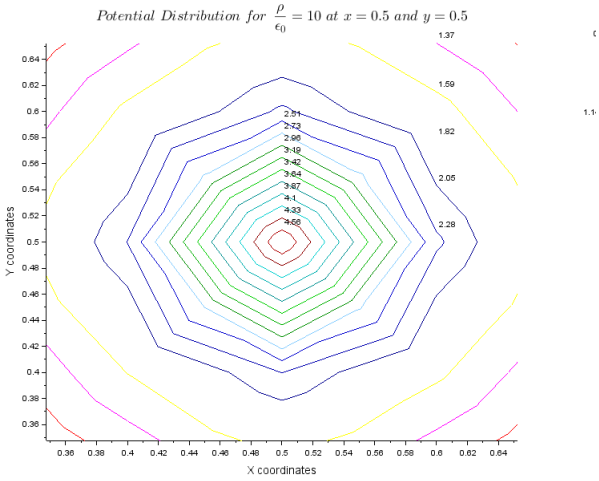


Figure 3: Zoomed Contour Plot of the potential distribution due to positive point charge

In figure 4 the computed potential is plotted along the z-axis as a function of x and y coordinates of the grid points. Such a plot can be regarded as a "Potential Surface". It aids in visualizing the electrostatic field as the negative gradient of electrostatic potential. The field $\vec{E}(x,y)$ at a given point (x,y) on the potential surface is a vector in the direction of steepest ascent having a magnitude equal to the slope of the surface at the given point measured in the given direction of steepest ascent of V(x,y). In figure 5 the lines of force and vector representation of the electrostatic field due to the given monopole are shown superimposed on each other. It can be seen that at the grid points that are common to both the plots, the field vectors are tangential to the lines of force. The lines of force are seen to spread out radially from the point charge. The $\vec{E}$ vectors are also directed radially outwards from the positive charge.From the plots in figure 2 and figure 5 it can be seen that the lines of force and equipotential curves are mutually perpendicular at all the points in the given region of space.

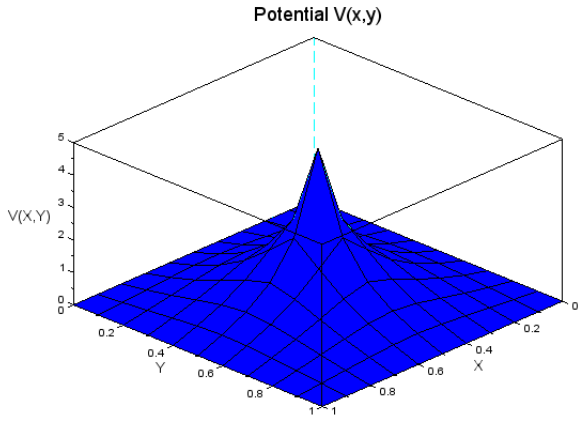


Figure 4: The Potential surface due to a monopole in the form of a positive point charge

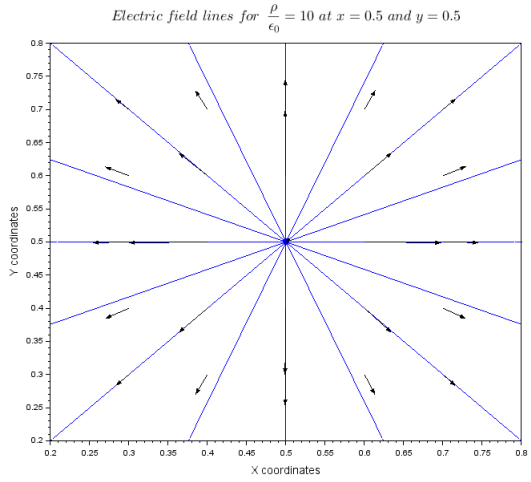


Figure 5: Lines of force and vectorial representation of the electrostatic field due to a positive monopole

4.2 Electric Dipole

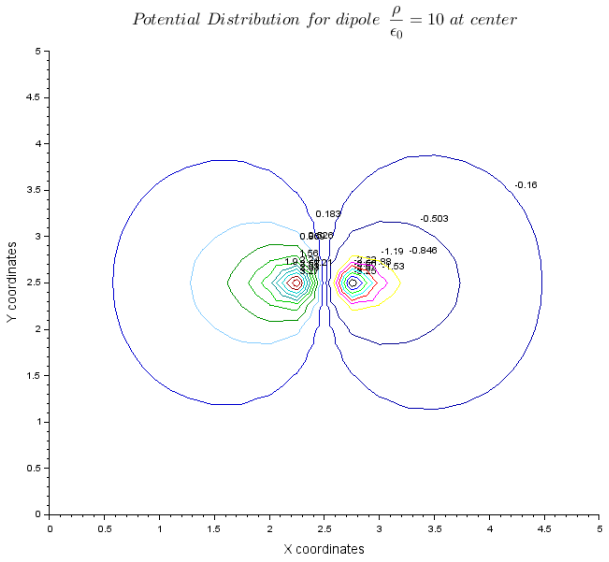


Figure 6: Concentric loops are equipotential curves representing the potential distribution due to a dipole

A system of equal and opposite point charges separated by a distance is defined to be an electric dipole. Electrostatic potential and electrostatic field due to a dipole of given strength placed at the geometric center of a plate grounded on all sides have been simulated. Each of the point charges constituting the dipole are taken to be of magnitude $\frac{\rho}{\epsilon_0} = 10$. The axis of the dipole is taken parallel to the x-axis. The contour plot of the electrostatic potential distribution due to the given dipole is shown in figure 6 and figure 7 gives the zoomed version of the same.

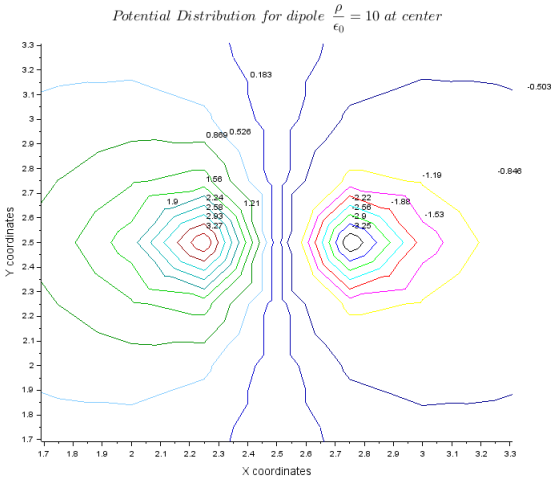


Figure 7: Zoomed Contour plot of the potential distribution due to a dipole

Figure 8 is the plot of the dipole potential surface. The negative dent in the surface corresponds to the negative point charge and the positive peak corresponds to the positive charge constituting the dipole. The lines of force and vector representation

of the dipolar electrostatic field is depicted in figure 9. The lines of force can be seen to originate from the positive charge and terminate at the negative charge. These plots also depict all the features related to potential curves, lines of force and the field vectors discussed earlier in the context of monopole charge.

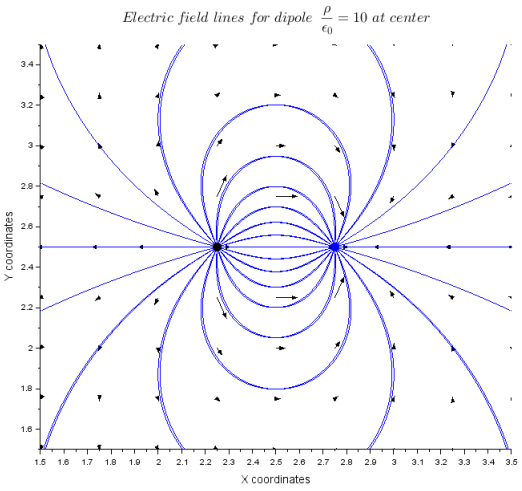


Figure 9: Lines of force and vectorial representation of electrostatic field distribution due to a dipole

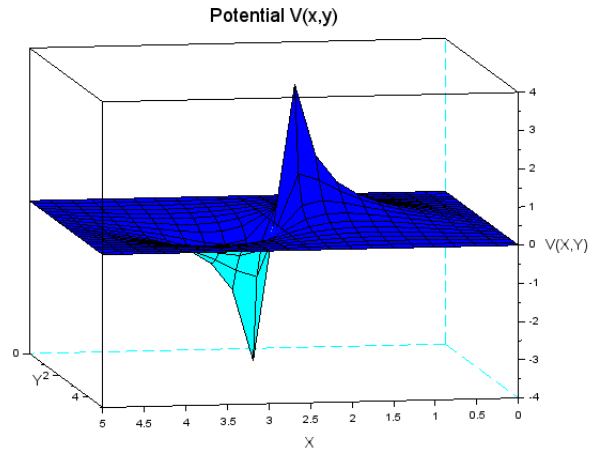


Figure 8: Potential surface due to a dipole

4.3 Electric Quadrupole

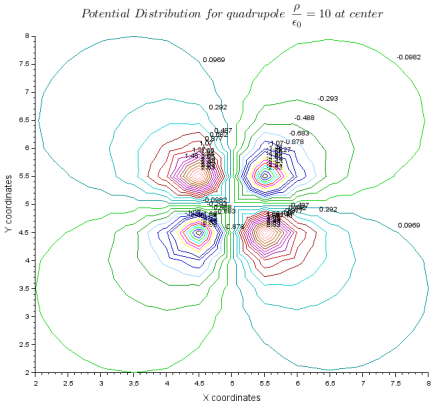


Figure 10: Contour plot of electrostatic potential distribution due to an electric quadrupole

An electric quadrupole is a system of two dipoles with equal and opposite dipole moments.Electrostatic potential and electro-

static field due to a quadrupole of given strength placed at the geometric center of a plate grounded on all sides have been simulated. The magnitude of each of the four charges forming the quadrupole is taken to be $\frac{\rho}{\epsilon_o} = 10$. Figure 10 shows the contour plot of electrostatic potential distribution due to the given quadrupole. Comparing figure 6 and figure 10, it can be seen that the quadrupole potential is a combination of potential distributions due to two dipoles forming the quadrupole. The zoomed version of quadrupolar potential distribution is shown in figure 11.

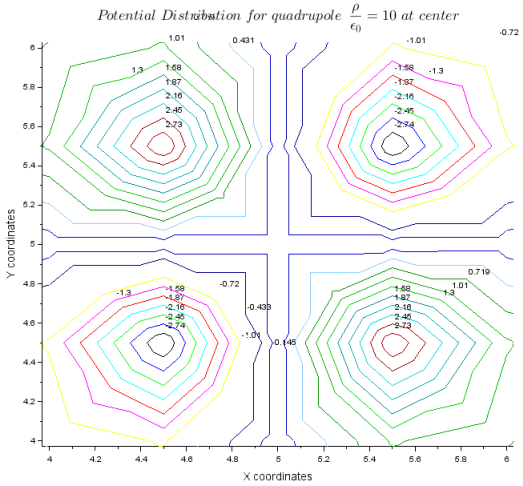


Figure 11: Zoomed Contour plot of the potential distribution due to a quadrupole

Figure 12 is the surface plot of the quadrupole potential distribution. It has two peaks corresponding to the two positive point charges and two negative dents corresponding to the two negative point charges forming the quadrupole. The lines of force and vector representation of electrostatic field due to the electric quadrupole is given in figure 13.

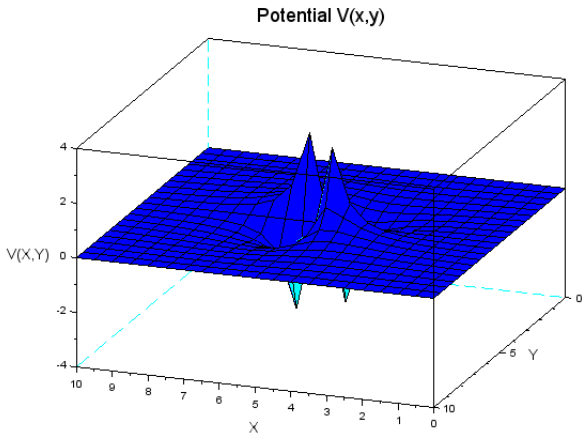


Figure 12: Surface plot of electrostatic quadrupole potential

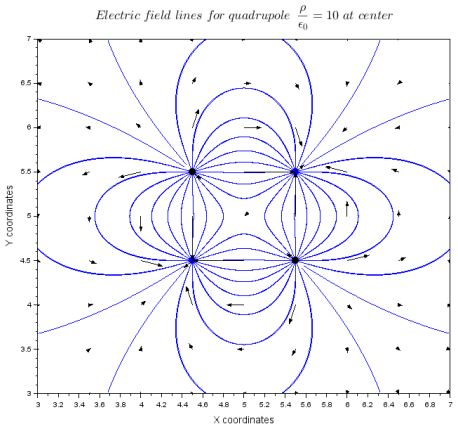


Figure 13: Lines of force and vectorial representation of electrostatic field due to an electric quadrupole. The electrostatic lines of force are seen to originate from the positive charges and sink at the negative.

4.4 Three Point Charges

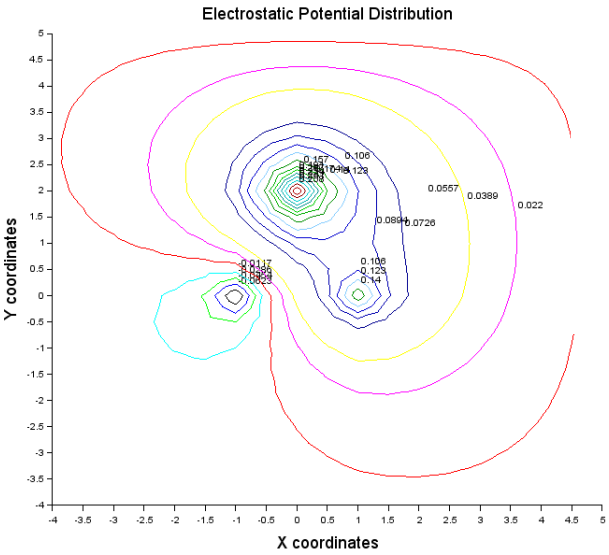


Figure 14: Contour plot of electrostatic potential due to three point charge configuration

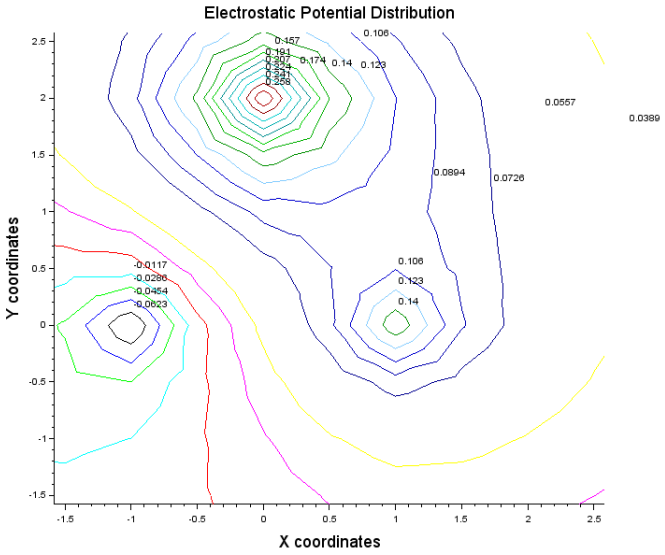


Figure 15: Zoomed Contour plot of electrostatic potential due to three point charge configuration

The charge configuration consists of three point charges with $\frac{\rho}{\epsilon_o}$ = 1,-1 and 2 units located at (1,0), (-1,0) and (0,2) respectively on a plate with dimensions of 10 × 10 units. The corresponding electrostatic potential and electrostatic field distributions are depicted in the figure 14, 15, 16 and 17. In figure 17 the lines of force can be seen to originate from the positive charges and terminate at the negative charge.

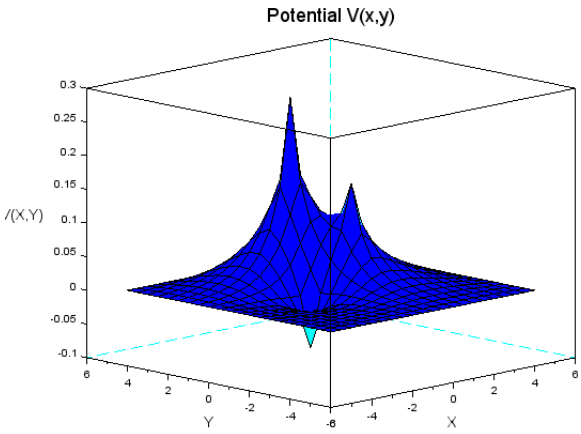


Figure 16: The surface plot of potential distribution due to three point charge configuration. The surface can be seen to have a negative dent with respect to the reference plane for the negative charge and positive peaks for positive charges, with the peak proportional to the magnitude of charge.

4.5 Five Point Charges

The charge configuration consists of five point charges with $\frac{\rho}{\epsilon_o=}1$ unit each located at (0,0), (1,1), (-1,1), (-1,-1) and (1,-1) respectively on a plate with dimensions of 5×5 units. The corresponding electrostatic potential and electrostatic field are depicted in the figures 18, 19, 20 and 21. In figure 18 it can be seen that as one moves outwards from the centre of the charge configuration, the contour curves begin to resemble the ones obtained for a single point charge configuration as shown in figure 2. This is indicative of the fact that at larger distances compared to the separation between the constituent charges, the given system of five point charges behaves like a monopole.

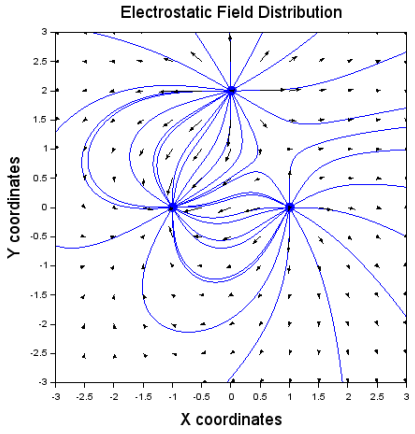


Figure 17: The lines of force and vectorial representation of the electrostatic field due to three point charge configuration

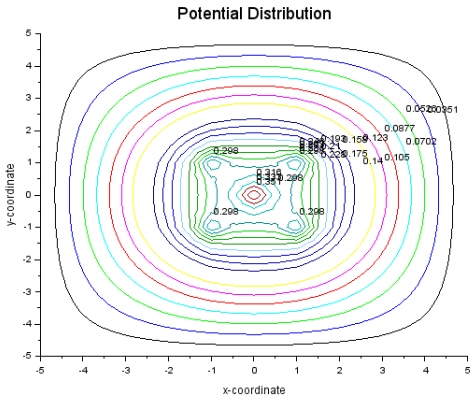


Figure 18: Contour plot of electrostatic potential distribution due to a five point charge configuration

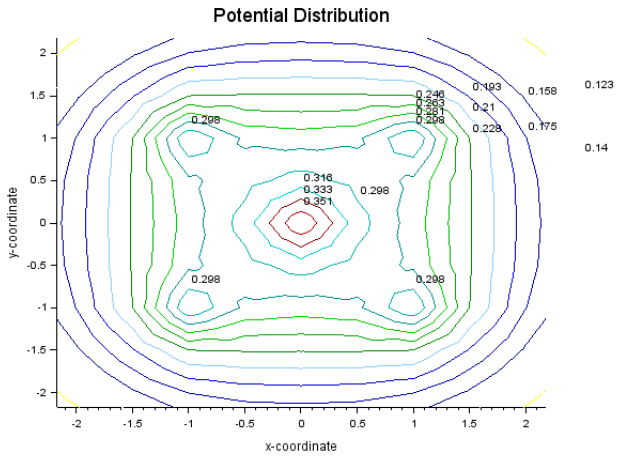


Figure 19: Zoomed contour plot of electrostatic potential distribution due to a five point charge configuration

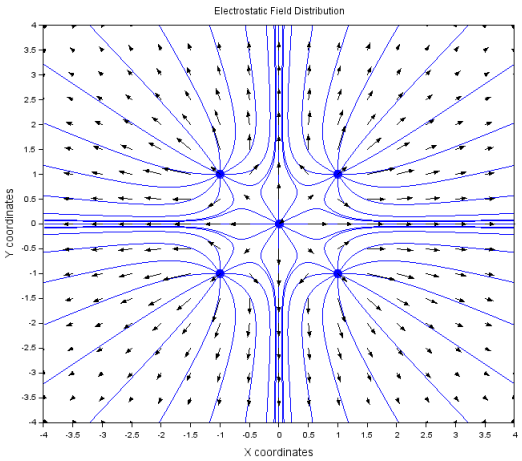


Figure 21: The lines of force and vectorial representation of the electrostatic field due to five point charge configuration. At the grid points common to both the plots, the field vectors can be seen to be tangential to the lines of force.

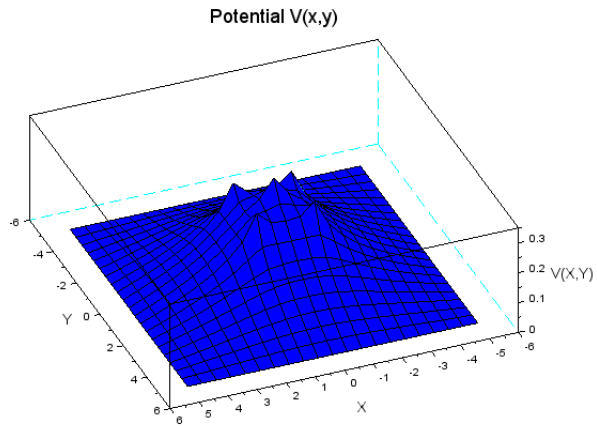


Figure 20: The surface plot of potential distribution of a five point charge configuration. The surface can be seen to have five positive peaks of equal heights corresponding to the five positive charges of equal magnitudes forming the charge configuration.

4.6 A Line Charge

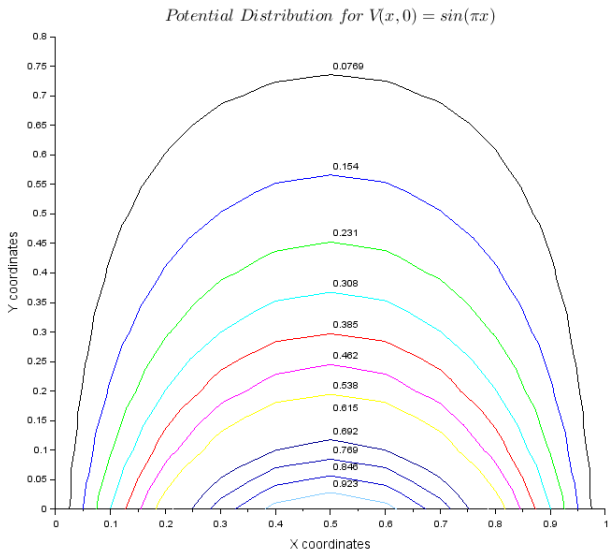


Figure 22: Contour curves depicting electrostatic potential distribution form semicircular loops about the line charge

The charge configurations discussed so far were all discrete point charge configurations. Poisson's equation subject to grounded boundary conditions was solved in each of these cases. In this section a continuous charge distribution in the form of a line charge is considered. The electrostatic potential and field distribution is computed for a sinusoidal line charge taken along one of the sides of a plate (the x- axis in the present simulation) with the other three sides of the plate taken to be grounded. Laplace's equation is solved with the given boundary conditions in this case. Figure 22 depicts the contour plot of the potential distribution due to the given line charge. The surface plot of the potential distribution is shown in figure 23. In both the figures, it can be seen that the potential exhibits a maxima at $x = 0.5$ and $y = 0$, where $V(x,0) = sin\frac{\pi}{2} = 1$.

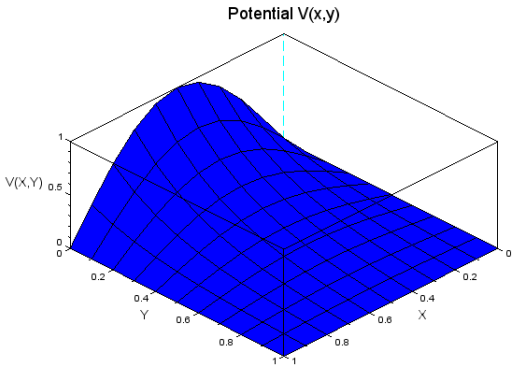


Figure 23: Surface plot of the electrostatic potential distribution due to a sinusoidal line charge

Figure 24 is the vector plot of the electrostatic field distribution due to the given line charge configuration. The lengths of arrows depict the variation in the strength of the field at the various grid points while the arrow heads point in the direction of the field at the given grid point.

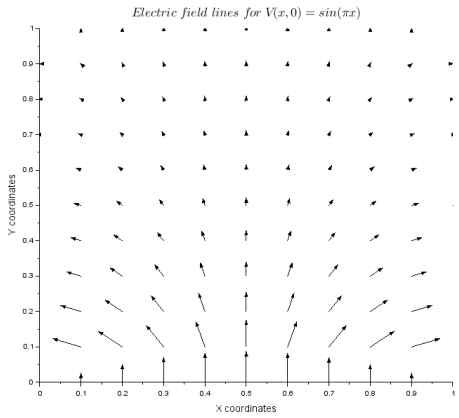


Figure 24: Vector plot of the electrostatic field due to a sinusoidal line charge distribution

4.7 Superposition Principle

For a given arrangement of point charges, the electrostatic field and potential at a given point in space due to one charge is independent of the presence of other charges in the configuration. This leads to the superposition of electrostatic fields or equivalently the superposition of electrostatic potentials due to the individual charges to give the resultant field or the equivalent potential of the configuration at a given point.

The superposition of electrostatic potential is simulated for a five point charge configuration. Five positive point charges each of magnitude $\frac{\rho}{\epsilon_o} = 1$ are taken to be located at the grid points with coordinates (-1,-1), (-1,1), (1,1), (1,-1) and (0,0). Figures 25 through 29 depict the contour plots of electrostatic potential distribution due to each of the above charges considered one at a time. When the potential distribution due to each single charge is superimposed one by one according to eqn.(7), the contour plots in figure 30 through figure 34 are obtained for the corresponding potential distributions.The contour plot of the potential distribution when all the five charges are taken together as a single charge configuration has aready been discussed earlier and is shown in figure 18. The similarity between the plots in figure 34 and figure 18 illustrate the superposition principle for electrostatic potential.

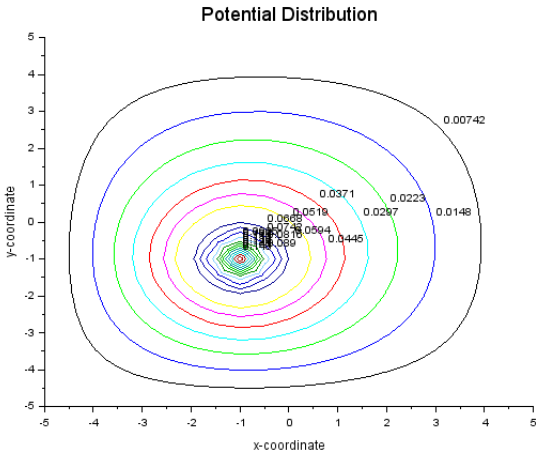


Figure 25: Contour Plot of Potential Distribution due to charge at (-1,-1)

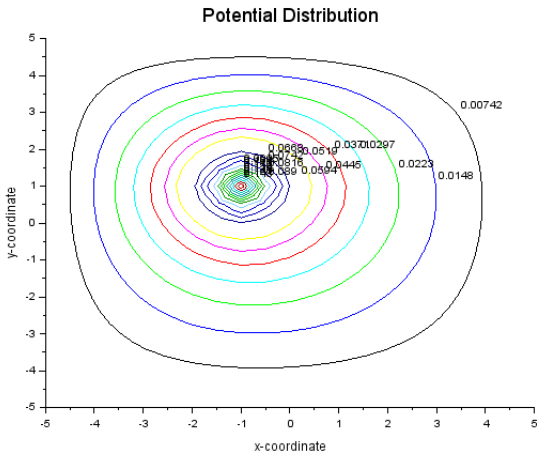


Figure 26: Contour Plot of Potential Distribution due to charge at (-1,1)

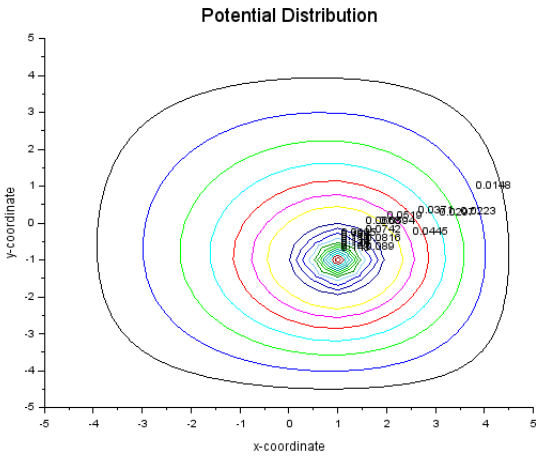


Figure 28: Contour Plot of Potential Distribution due to charge at (1,-1)

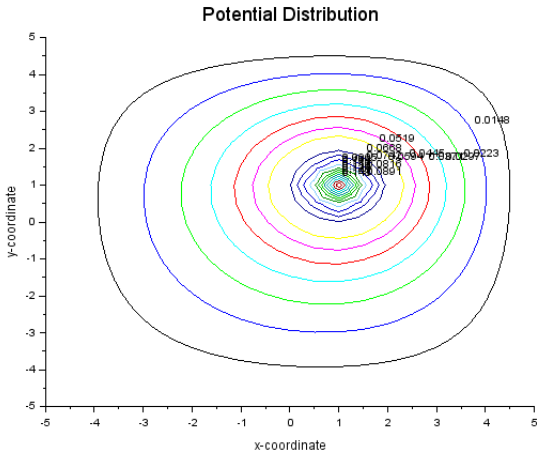


Figure 27: Contour Plot of Potential Distribution due to charge at (1,1)

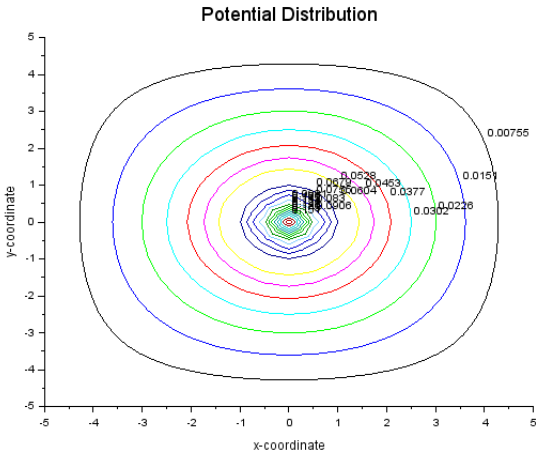


Figure 29: Contour Plot of Potential Distribution due to charge at (0,0)

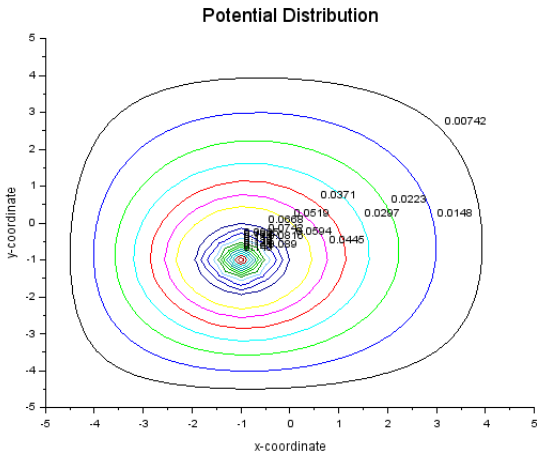


Figure 30: Contour plot of potential distribution when the first charge is placed at (-1,-1)

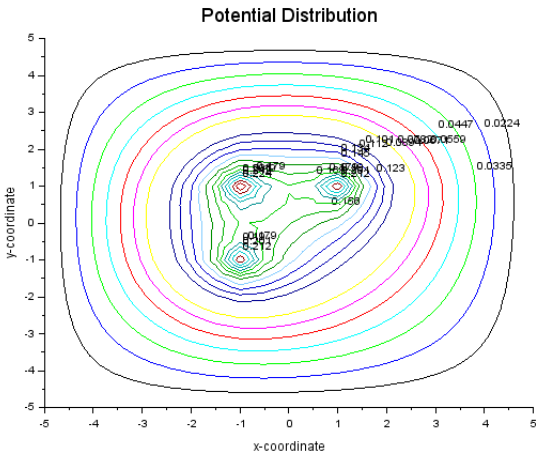


Figure 32: Contour plot of potential distribution when the third charge is added at (1,1)

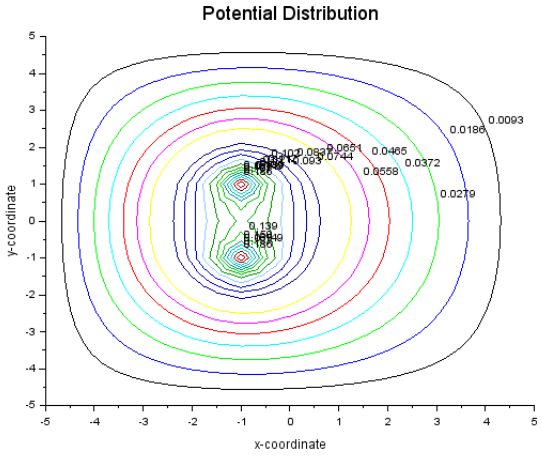


Figure 31: Contour plot of potential distribution when the second charge is added at (-1,1)

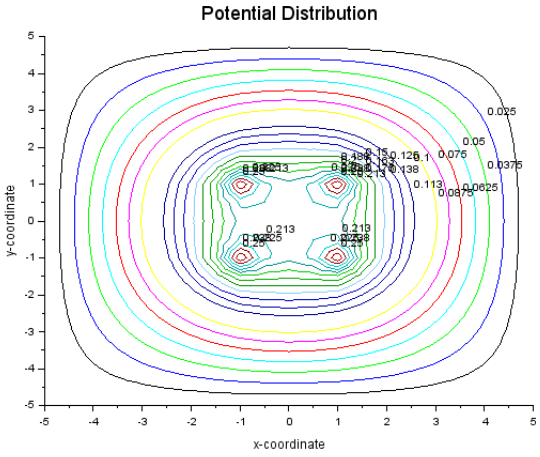


Figure 33: Contour plot of potential distribution when the fourth charge is added at (1,-1)

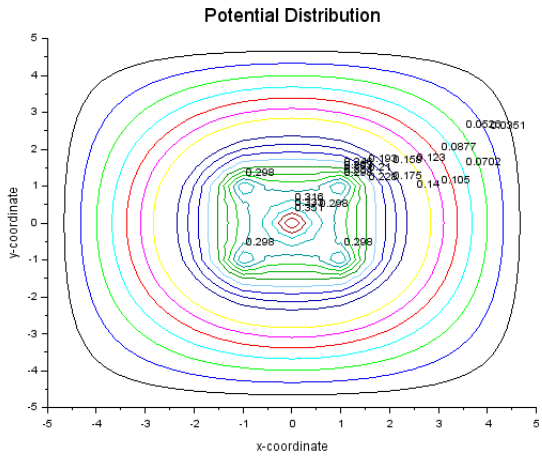


Figure 34: Contour plot of potential distribution when the fifth charge is added at (0,0)

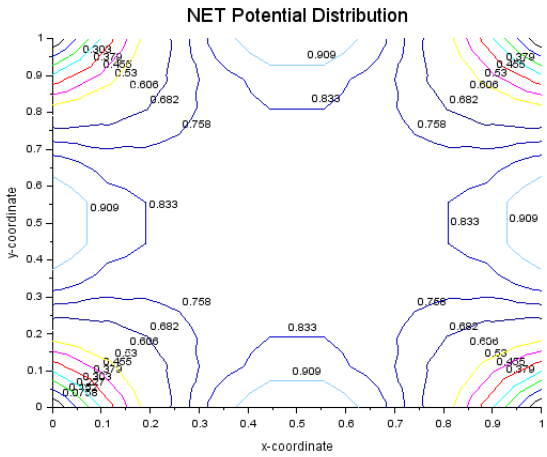


Figure 36: Contour plot of net potential distribution due to superposition of the four line charges, one on each boundary of the plate

In figures 35 and 36 the superposition principle for electrostatic potential distribution due to four sinusoidal line charges on the boundaries of the plate is illustrated. The centre of the plate can be seen to have zero potential as a consequence of the superposition principle.

5 Conclusion

The simulation of electrostatic fields and potential distribution done in the present work can be used as visual aid for an effective teaching of these concepts at the high school level where the students are not expected to do rigorous analytical treatment of such problems. At the introductory undergraduate level, the computational background discussed in the work can be used to integrate computational methods with the analytical methods usually adopted for solving problems in traditional classroom scenario for an enhanced teaching-learning experience of the fundamentals of electrostatics. The work can be easily extended to simulate some other fundamentals of electrostatics like the fringing of lines of force in a parallel plate capacitor, electrostatic shield-

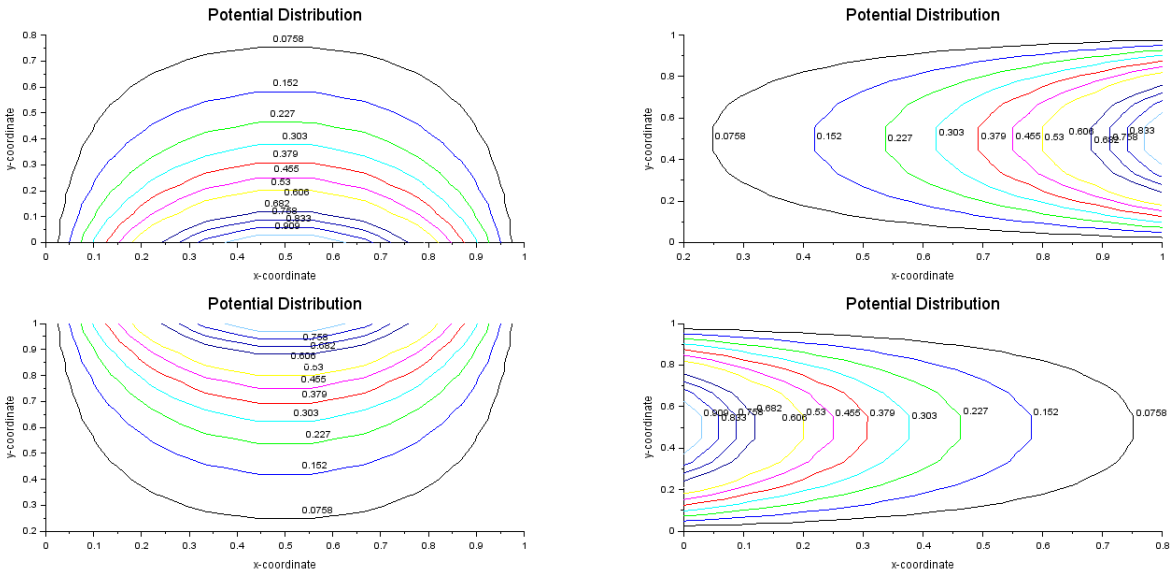


Figure 35: Contour plots of potential distribution when line charges are taken at the plate boundaries one at at a time; at y=0(top left), x=1(top right), y=1(bottom left) and x=0(bottom right)

ing etc. The extension of the work can also be done to compute and simulate electrostatic field and potential distribution due to charge configurations with spherical or cylindrical symmetry which find more practical applications but the analytical treatment of the same often becomes cumbersome.

6 REFERENCES

1. Matthew N.O. Sadiku, Principles of Electromagnetics, Oxford University Press.

2. Arthur F. Kip, Fundamentals of Electricity and Magnetism, McGraw-Hill Series in Fundamentals of Physics.

3. Edward M. Purcell, David J. Morin, *Electricity and Magnetism, Berkley Physics Course, Volume 2*, Tata McGraw - Hill Publishing Company Limited.

4. Joel Franklin, *Computational Methods for Physics*, Cambridge University Press.

5. SCILAB, version 6.0.1

Gamma Spectra of Non-Enriched Thorium Sources using PIN Photodiode and PMT based Detectors

Jithin Bhagavathi[1], Swapna Gora[1], V.V.V. Satyanarayana[2], O.S.K.S Sastri[1] and B.P Ajith[2]

[1]Central University of Himachal Pradesh, Bharat.
sastri.osks@gmail.com

[2]Inter University Accelerator Centre, New Delhi, Bharat.
ajith@iuac.res.in

Submitted on 14-12-2019

Abstract

Gamma spectra of non-enriched Thorium Nitrate $ThNO_3$ salt is studied using an indigenously developed, low-cost portable gamma spectrometer, based on CsI scintillator mated to a PIN photodiode. The results of our study are compared with commercially available detectors based on NaI scintillators coupled to PMT detectors, that come with a certain cost and are bulky. Even though the intensity of various peaks are relatively less prominent in the spectrum obtained with the photodiode detector as compared to the PMT based one under same experimental conditions, longer duration for data collection and analysis using log plots were able to aid in reliable identification of all the dominant peaks in the source. The experiment gives good insights into nuclear radiation processes and various interactions that take place within the detectors. Finally, we repeated the experiment with a sample of Monazite from the sands of Kerala and confirmed the presence of Thorium deposits in it, thus proving the effectiveness of our portable gamma spectrometer for studying environmental radiation.

keywords: Gamma spectroscopy, Non-enriched sources, ThNO3 salt, Monazite

1 Introduction

In a previous paper [1], we have discussed the alpha spectrum of ^{212}Bi obtained from electrolysis of $ThNO_3$ salt, a non-enriched source available from chemical suppliers. The same salt can be used in powder form to obtain the gammas emmited by it. A low cost Gamma spectrometer has been developed by us [2] which uses a CsI scintillator mated to a PIN photodiode via an index

matching optical grease interface. The design details are discussed in a separate paper communicated recently. The electronics developed for our Alpha spectrometer [3] has been suitably modified, and the software CN-Spec has been upgraded to obtain the required spectral analysis along with added features for visualising surface plots for coincidence experiments [4]. The entire design has been prepared for commercial use by CSpark Research (India) [11] as GammaSpec1K. This work is in continuation with our efforts to developing Nuclear physics lab based on PER strategies for advanced UG experiments[6] using non-enriched sources and affordable equipment. In the presentation of this paper, we shall be focussing on studying gamma spectra of non-enriched Thorium samples. The first step before performing the experiment is to model the study, and this consists of arriving at the learning goals by enquiring into what we wish to accomplish. This is followed by understanding the theoretical considerations regarding the Thorium source, deduced from the experimental evidences documented in standard databases such as ENSDF[7]. Then, the physical system is modeled in terms of possible interactions of gamma radiation in the detection medium. All this has been elaborated in the second section. In the third section, we focus on the experimental design strategy and model the measurement apparatus elucidating the principle of operation, the capabilities of the detection units and data acquisition. This is followed by the analysis of spectra and discussion of various aspects that characterize them in section 4. Finally, we extend our study to validate the presence of Thorium deposits in Monazite found in the sands of Kerala and draws conclusions.

2 Modeling

In model construction, as a first step, we need to answer the following simple questions?

1. Which real world phenomenon do we wish to study?
   Ans: We wish to study the gamma radiation emitted by Monazite sand commonly found in beaches of the Indian peninsula. It is known from literature, that the radioactive deposits in the sands contain deposits of Thorium and trace amounts of Uranium. We have with us in our laboratory, a non-enriched source of Thorium in the form of $ThNO_3$ salt, so it will be appropriate to first conduct the experiment on this source, and later apply the obtained knowledge to understand the radiation emitted by the Monazite sand.

2. Which aspects of the real-world system are included in the model?
   Ans: One of the naturally occurring radioactive series is that of Thorium, whose data is well documented in ENSDF. It goes through a series of alpha and beta decays during which the

daughter nucleus might be in one of its excited states that undergoes gamma decays. We focus only on obtaining the energies of these gamma emissions in our model.

3. What is being ignored?
   Ans: We are ignoring the detection of alpha and beta. Also, we are not interested in the intensities of the gamma emission that are quoted to be less than 10%, and also energies which are below 100 keV, keeping in mind the capabilities of our gamma spectrometer.

4. What principles are needed to describe the phenomenon?
   Ans: The interaction processes that gamma undergoes during the deposition of its energy within the detector medium needs to be understood. That is, the principles of compton effect, pair production and photo-peak are to be described while analyzing the obtained spectra.

5. What parameters are needed in the model?
   Ans: If the experiment is being introduced at the UG level, we could just focus on determining the Q values for alpha and beta for the various daughter nuclei of Thorium and try to make them appreciate the theoretical reasoning behind the observed series. At PG level, one could also include the total angular momentum I and parity pi values in the data and emphasize on the selection rules that are resulting in the various decays being observed.

6. What approximations or idealizations are made ?
   The approximations result from the limitations that are imposed due to the capabilities of the detector. For example, the FWHM of the spectrometer naturally dictates the resolution of the gammas that can be observed.

Further, since the gamma spectrum that is being studied is that of Thorium and we have already utilised a non-enriched source of ThNO3 salt for studying alpha spectra[1], we have extended the experiment to also include the study of spectra obtained from this salt using both the gamma spectrometer that we have developed[2] and the existing one in the lab based on NaI coupled to PMT. This would help in comparative analysis. Based on all of the above considerations, we have designed the lab specific learning goals as:

1. understanding the nuclear radiation processes that are resulting in the gamma emissions of Th,

2. comprehension of the various interactions that are responsible for energy deposition in the detectors

3. a broad overview of the measurement processes, and comparative analysis of spectra obtained from two different types of detection techniques.

4. application to studying samples that contribute to environmental radiation

2.1 Modeling the Source

The source is a non-enriched radioactive sample composed of thorium nitrate ($Th(NO_3).5H_2O$) in powder form which has ^{232}Th and its various daughter products in secular equilibrium via decays associated with alpha and beta emission. De-excitation of daughter products from these decays are accompanied by emissions of gamma rays. Alternatively, we have also used a sample of Monazite sand obtained from the beaches of kerala that contains Thorium deposits and some trace amounts of Uranium as well. Our focus is on:

1. Identifying the various gammas present in the two samples.

2. Appreciate the potential of the non-enriched, naturally radioactive Monazite sand as a safer alternative.

3. Comparing the performances of CsI mated to PIN photodiode to that of NaI scintillator attached with a Photo Multiplier Tube(PMT).

The learners' goals are to (i) model the gamma decays theoretically and (ii) obtain the quantitative predictions of observable phenomenon, in this case, the gamma energies that are emitted by Thorium samples.
The first objective is achieved by making the students observe the level diagrams of ^{232}Th and its entire decay chain from ENSDF website. The alpha and beta decays that the various nuclei undergo are governed by the respective selection rules. The intensities to the ground and excited states are given along with the respective Hindrance factor HF in case of alpha and log ft values in case of beta decays. Whenever a given parent nucleus decays into one of the excited states of the daughter, it is followed by gamma emission so as to reach the ground state. All these adopted gamma decays from various experiments performed by different groups across the world are given in the level structure diagrams of the nuclei and the students are asked to choose the levels resulting in gammas which are having intensity greater than 10% for the purpose of study. The resultant flow diagram is shown in Fig 1.

2.2 Modeling of Physical System

Gamma rays have high penetrating ability as compared to alpha and beta particles, and it is not possible to directly measure their energy with a semiconductor based detector such as a photodiode. Therefore, typically a scintillation crystal is used, which absorbs the gamma ray energy and emits a proportional number of scintillation photons whose wavelengths lie in the visible region. The difference in absorption probabilities for a semiconductor and scintillation detector is very evident in the log plot in Figure 2, and one can see that a simple silicon detector has a next to nothing chance of absorbing gamma rays with energies beyond 200keV.

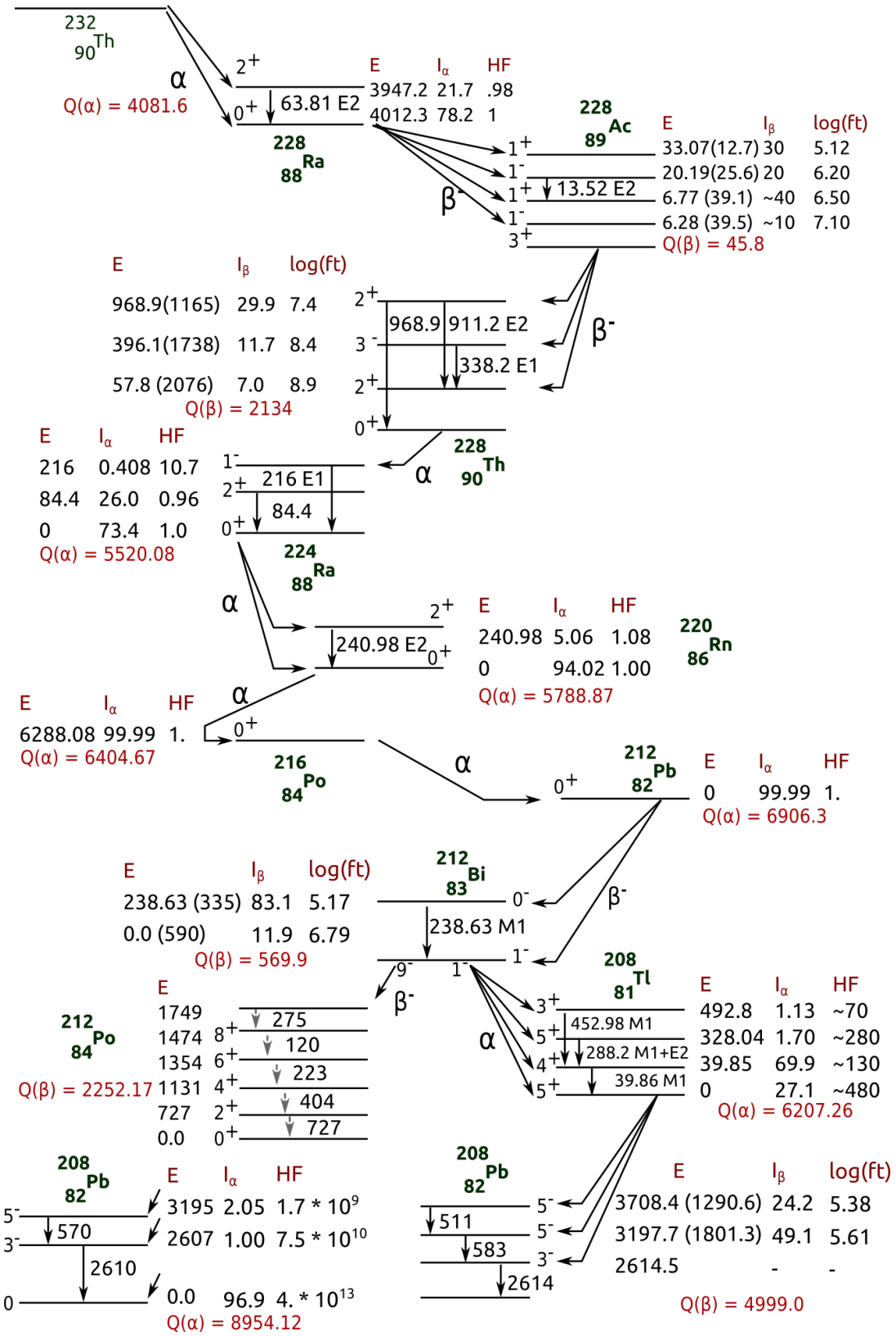


Figure 1: The decay chain of ^{232}Th highlighting the level structures of daughters that result in $>$ 10% gamma intensities.

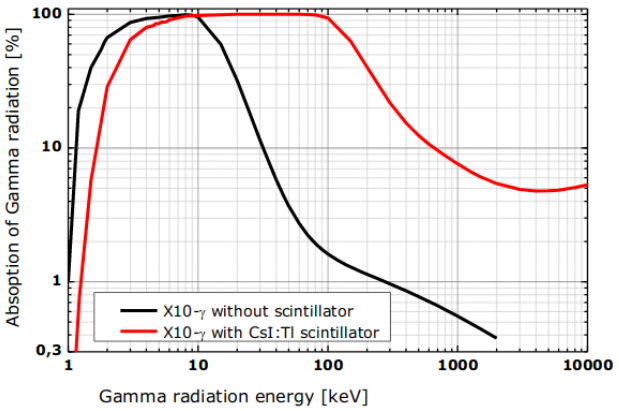


Figure 2: Absorption probability of gamma rays in a solid state detector as a function of their energy with and without scintillator attached. Source:[13]

2.2.1 Interaction Processes:

Since the gammas do not attenuate appreciably in air, we need to consider only the interactions that take place within the detector. The following competing processes define the interaction of photons with matter, or in our case, gamma rays with the scintillation crystal:

Photoelectric effect: The incident gamma ray transfers its entire energy to an electron in the scintillator which then travels along the scintillator material creating a number of scintillation photons proportional to its initial energy.

Compton scattering: This occurs when the incident gamma ray interacts with a free electron and loses some energy in the recoil process. The energy loss is characterized by a loss in frequency of the incident gamma ray since its velocity must be constant. The resultant scintillation photons produced by the recoil is also lower than that of the photoelectric process since the scattered gamma ray still contains a part of its initial energy.

Pair production: For gamma rays with energies(E) greater than 1.022MeV which corresponds to twice the rest mass energy of an electron, an electron-positron pair may be created in accordance with Einstein's mass-energy relation. The energy excess is transferred to the pair. The positron annihilates with the production of a pair of 511keV gamma rays. The net energy deposited can be one of the following

- Initial Energy of the gamma ray(E). This implies that both 511 keV photons deposited their energy in the scintillator, and the excess energy also created a proportional number of scintillation photons.
- One of the photons escaped the material, resulting in the total deposited energy being only $E - 511keV$. This position in the spectrum is called a single escape peak.
- Both the annihilation photons escaped the material, resulting in $E - 2 * 511keV$ being deposited. This creates what is known as a double escape peak.

2.3 Modeling Outcomes

The students at UG level can easily determine the Q_α and Q_β values of the nuclei using Binding energy values obtained from Semi-Empirical Mass Formula based on the Liquid Drop Model. This provides the theoretical basis as to why the sequence, of αs and βs in Thorium series, is the way it is. $Q_\beta \leq 0$ and $Q_\alpha >> 0$ implies, the nucleus is β stable and the only mode of decay is through α. Whenever, $Q_\beta > 0$, it is the more preferred mode of decay.

At the PG level, the students can be guided to observe that most of the β decays are proceeding through parity changing forbidden transitions to the excited states of the daughter nuclei. Further, it is seen that for the transitions to proceed to the ground state of the daughter, we have $\Delta I \geq 2$. An analysis of the log ft values of β-decays in the A $>$ 221 Actinide region has revealed that the first forbidden transitions have comparable values to that of the allowed transitions[12]. The various gammas from the level sequence diagrams in Fig.1 are compiled in Table 1. Even though all these energies are expected to be seen, the limitations and interactions of the detection unit places certain restrictions, which will be discussed later.

3 Experimental Design

There are different ways to design an experiment so as to validate the correctness of the results. One such design strategy that is incorporated here, is to obtain the data through two different techniques and perform a comparative analysis. Franklin [17] discusses this as the first epistomological strategy which is effective in arriving at valid results:

> A hypothesis receives more confirmation from 'different' experiments, where different measurement instruments are used, than from the repetition of the 'same' experiment,where the same apparatus is used each time. This reference to 'different' experiments is due to the fact that the theories that underlie the various apparatus are different; Since the results of an experiment are strongly related to the theories underlying the different apparatus employed, one may regard the results of performing an experiment using different measurement instruments as being obtained from 'different' experiments.

The gammas emitted by the $ThNO_3$ powder are studied using two different detection units. One is the commercially available NaI(Tl) scintillator coupled to a PMT and the other is our own indigenously developed detection unit consisting of CsI(Tl) scintillator mated to a PIN photodiode using an index matching glue. Here, the methods for depositing the energy of the scintillation photons, generated from gammas incident on the NaI(Tl) or CsI(Tl) detectors, are different and hence the measurement instru-

Parent	Process	Daughter	I_p^π	$I_{d*}^\pi \rightarrow I_d^\pi$	I_α / I_β	HF/ logft	E_γ (keV)
$^{232}_{90}Th$	α	$^{228}_{88}Ra$	0^+	$2^+ \rightarrow 0^+$	21.7	0.9	63.8
$^{228}_{88}Ra$	β	$^{228}_{89}Ac$	0^+	$1^- \rightarrow 1^+$	20	6.2	13.5
$^{228}_{89}Ac$	β	$^{228}_{90}Th$	3^+	$2^+ \rightarrow 0^+$	29.9	7.4	968.9
				$2^+ \rightarrow 2^+$			911.2
				$3^- \rightarrow 2^+$	11.6	8.4	338.2
$^{228}_{90}Th$	α	$^{224}_{88}Ra$	0^+	$1^- \rightarrow 0^+$	0.4	10.7	215.9
				$2^+ \rightarrow 0^+$	26.0	0.9	84.4
$^{224}_{88}Ra$	α	$^{220}_{86}Rn$	0^+	$2^+ \rightarrow 0^+$	5.1	1.1	240.9
$^{212}_{82}Pb$	β	$^{212}_{83}Bi$	0^+	$0^- \rightarrow 1^-$	83.1	5.2	238.6
$^{212}_{83}Bi$	α	$^{208}_{81}Tl$	1^-	$3^+ \rightarrow 4^+$	1.1	70	452.9
				$5^+ \rightarrow 4^+$	1.7	280	288.2
				$4^+ \rightarrow 5^+$	69.9	130	39.8
$^{212}_{84}Po$	α	$^{208}_{82}Pb$	18^+	$5^- \rightarrow 3^-$	2.1	$1.7 * 10^9$	570
				$3^- \rightarrow 0^+$	1.0	$7.5 * 10^{10}$	2610
$^{208}_{81}Tl$	β	$^{208}_{82}Pb$	5+	$5^- \rightarrow 5^-$	24.2	5.4	510.8
				$5^- \rightarrow 3^-$			583.2
				$3^- \rightarrow 0^+$	49.1	5.6	2614.5

Table 1: The I^π values for the parent and those of the excited states of the daughter and that of the final state to which the transition takes place are indicated.

ments differ in their principle of operation in producing the final output voltage corresponding to the same gamma energies.

3.0.1 Principle of operation:

In case of the photomultiplier tube(PMT), the scintillation photons eject a proportional number of electrons from the photocathode of the PMT, which then accelerate towards an adjacent dynode which is maintained at a higher potential. The electrons thus ejected are higher in number due to the energy gained whilst accelerating through the potential, and are now directed towards a second dynode which is maintained at an even higher potential. This sequence of events through multiple dynodes results in an amplification of the initial incident photopulse, and a proportional voltage spike is created at the output of the PMT.

On the other hand, the semiconductor photodiode's PN junction is connected in reverse bias mode so as to create a depletion region that acts as an ionization medium to convert their energy into electron-hole pairs which move towards the electrodes due to the applied electric field. The charge thus generated is proportional to the energy of

the incident gamma ray.
The length of the scintillator decides the photopeak efficiency. In other words, a longer scintillator tube results in more gamma rays depositing their entire energy in the crystal, and contribute to the photopeak rather than the Compton edge. The size of our CsI(Tl) detector is $10mm \times 10mm$ with a thickness of $8mm$ as compared to the NaI(Tl) detector available in the lab that has approximately 5 times larger dimensions.

3.1 Modeling the Measurement Apparatus:

The PMT from **Ortec** includes a preamplifier, and its output was input to a shaping amplifier stage by **CAEN(N968)**[15]. The shaper output of 3uS rise time was processed using a **4K USB Multi Channel Analyzer** developed by us[16], and the spectrum was acquired using our **CNSpec** Software. The PMT needs a bias voltage, which is typically in the range of 550 to 1100 V. A bias voltage of 730Volts was maintained for the PMT used here, so as to obtain the 1332 keV peak from ^{60}Co to be somewhat nearer to the half way mark in the 4K channels. For calibration, spectrum from ^{60}Co was taken, and the 1332keV peak was located at channel 1824.86. Then a second spectrum was acquired from ^{137}Cs, and the 662keV peak was located at channel 931.79. These two datapoints were used to create a straight line calibration polynomial. Since these two chosen peaks are far apart, errors in slope calculation arising due to small variations in the gaussian fitting based centroid estimation are minimized.
'GammaSpec-1K' gamma spectrometer setup, that can measure gamma energies upto 3 MeV developed by us (is shown in Figure 3). The various stages of signal processing electronics are shown as an inset. The compact detector unit, consisting of the CsI scintillator mated to PIN photodiode and the associated electronics, is the just the size of a smart phone (with about 4 times the thickness). It is powered by the USB of a laptop. Data from the 1K MCA is acquired via same USB cable and the spectrum builds dynamically in our software 'CNSpec' written in the Python language with FOSS tools[4]. Gaussian curve fitting, log plots, sum of events and other analysis and report requirements are incorporated into the software.
The source is to be placed in front of the 8mm hole behind which the detector is housed. The detector's peak sensitivity is around 540nm[13] and hence it is covered by an Aluminium foil of 20 microns thick, so as to block any stray visible spectrum light which was not generated by the scintillator.
The instrument is factory calibrated using ^{60}Co which has two sharp peaks at 1173 keV and 1332 keV. For this unit, the centroid channel for the 1.33MeV peak was calculated using gaussian fitting to be 503, and the corresponding calibration polynomial is $y = \frac{1332}{503} * x$. The FWHM of this peak is found to be 80 KeV. It is observed that at the lower end of the spectrum, spurious

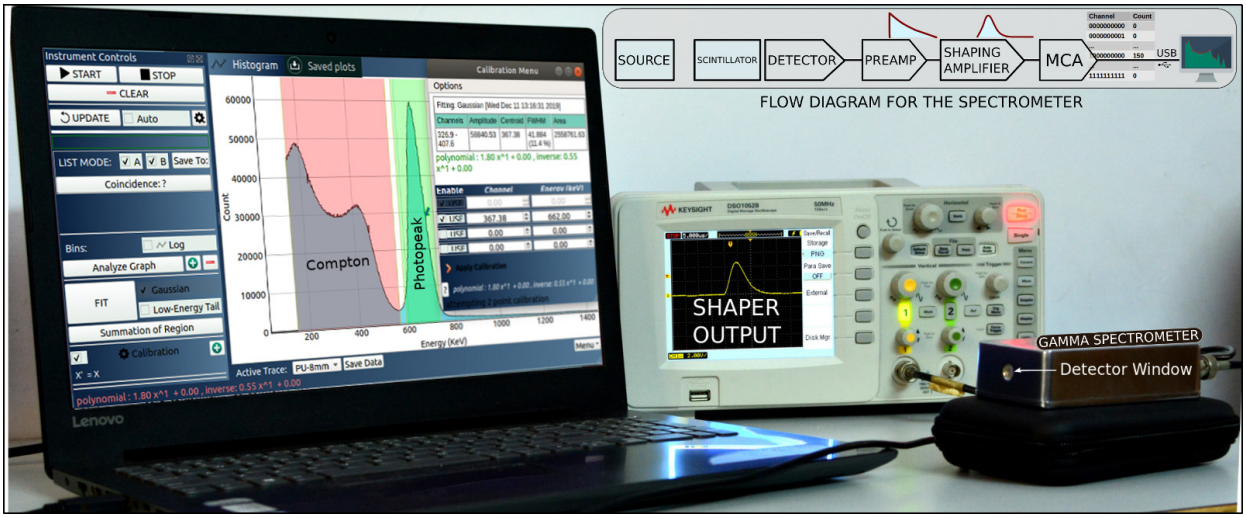


Figure 3: Photograph of the Gamma Spectrometer connected to a laptop showing the spectrum obtained. The shaper output is seen on the oscilloscope next to it, and the electronic circuit block diagram is shown as an inset.

events are getting recorded due to the noise fluctuations, in the absence of a signal. In order to reject this, the first 58 channels, that correspond to gamma energies upto 154keV, are not considered while acquiring spectra.

3.2 Data Acquisition

In the first iteration, the experiment is performed by simply placing approximately 20gm of Thorium Nitrate powder in front of a 2" NaI scintillator attached to a PMT detector manufactured by Ortec. The data was acquired for a period of 2 hours, and the obtained spectrum is shown in Figure 4. Again, 20gm of Thorium Nitrate powder is placed in front of the GammaSpec-1K detector window. The data is acquired for a period of 3.5 hours with a 1K MCA, and the spectrum is shown in Figure 5. Since this spectrometer uses a 1K MCA, it will record higher counts per bin as compared to a 4K MCA with the same input range. This is because the 4K MCA has four times as many channels into which the total incident events can be distributed.

4 Analysis and Discussion

4.1 Determination of Energies in the Spectrum:

Due to the large number of gamma rays with distinct energies emitted by ^{232}Th and its daughters, multiple photopeaks are formed in the spectrum, with most of them being superimposed on the Compton scattered events of gammas with higher energies than them. The width of each photopeak is smeared across a few channels in the spectrum due to the limitations of the de-

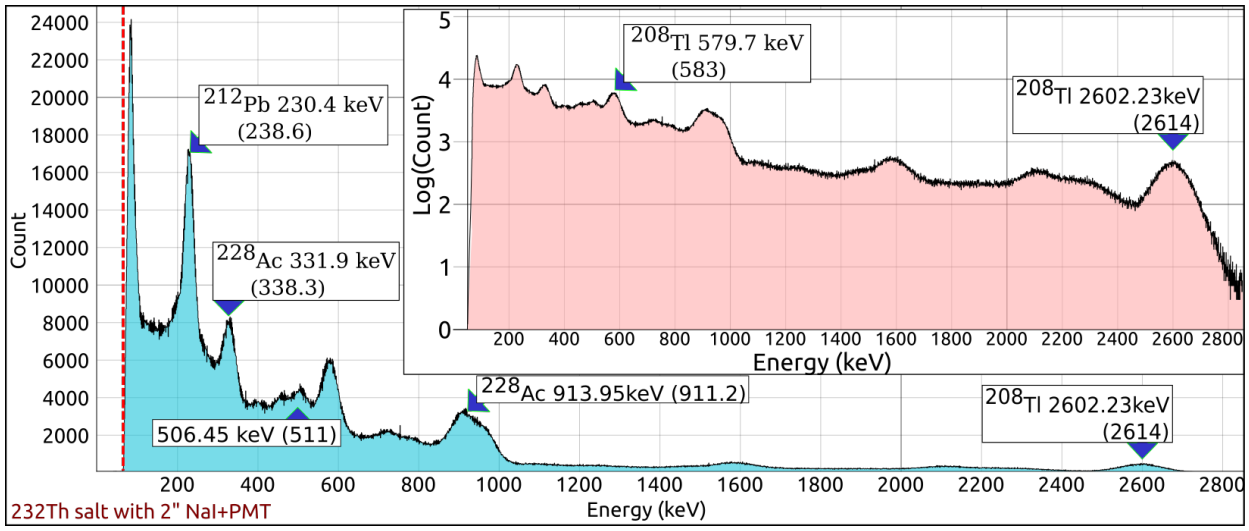


Figure 4: The Gamma spectrum of $ThNO_3$ powder obtained with a commercial PMT from Ortec with 2″ NaI scintillator and a 4K MCA.

Expected (in keV)	238.6	338.3	511	583	911	968.9	2614.5
2″ NaI(tl) + PMT	230.4	331.9	506.5	579.7	913.95	Unclear	2602.2
10mm CsI(Tl)+ Photodiode	235.6	339.0	508.5	580	913.6	966.6*	2607.2

Table 2: Comparison of peak centroids of gamma energies (in keV) obtained from 20gm $ThNO_3$ sample from the two different detectors used. *Peak appears as a shoulder peak, and has been identified manually.

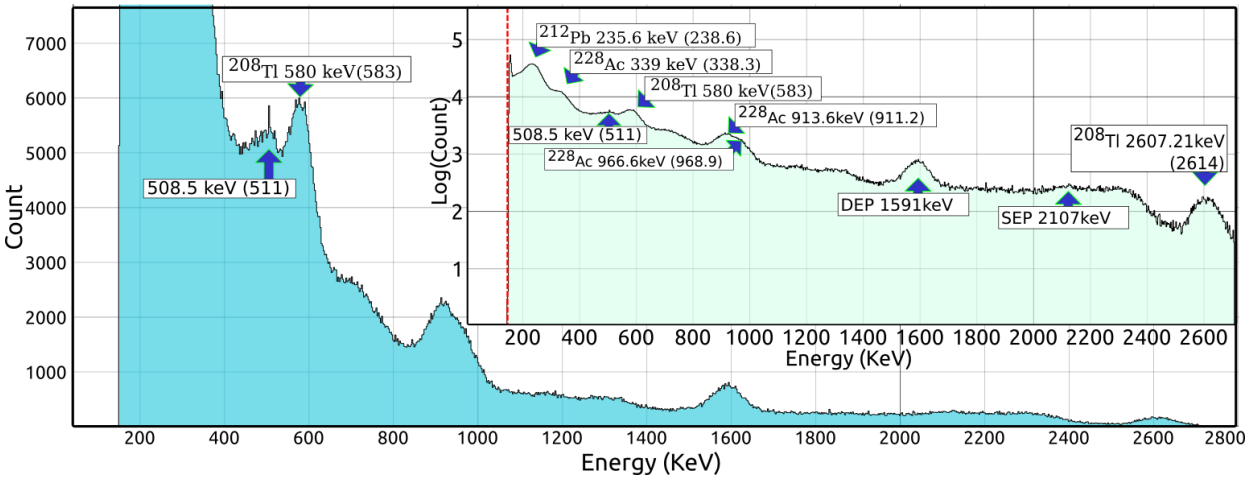


Figure 5: The Gamma spectrum of $ThNO_3$ powder obtained with GammaSpec1K spectrometer with CsI scintillator having volume $10mm * 10mm * 8mm$.

tector, and the centroid channel represents the actual peak position. Multiplying the centroid channel with the calibration polynomial gives us the energy associated with the peak. In order to extract the centroids which resemble a pseudo-gaussian shape, a mathematical least square minimization can be carried out on the channels containing the peak to identify the centroid. However, since this approach fails for small peaks riding on large Compton plateaus such as the 511 keV peak in Figure5, occasionally peak centers must be identified manually by estimating the channel spread of the photopeak. The energies obtained in both the spectra are tabulated in Table 2. We note that most of the energies expected from Table 1 are missing in the actual obtained spectrum. This is firstly due to the non-enriched source we are using, the energies due to transitions from I_α and I_β less than 5% do not emit enough gammas to be observed within the acquisition time. Further restrictions come due to the limitations of the detector units and the interaction processes within the detector medium which are discussed below:

4.1.1 Limitations of the detector unit

- All the energies less than 154 KeV shall be suppressed, corresponding to the 58 channel noise threshold we have set. That is, the gammas with energies that are not highlighted in the last column in Table 1 are to be ignored.

- The energies within the range of the FWHM, which is about 80-100KeV, cannot be resolved and shall appear at their weighted mean value. The resultant gammas to be expected, should form photopeaks in the spectrum at 238.6keV, 338.2, 481.9, 576.6, 940, and 2614.5 keVs.

4.1.2 Interations with the detector medium:

- Compton scattering produces a continuous range of energies below each of the expected energies.

- The gammas with energy greater than 1022 keV could result in pair production. In case of our $ThNO_3$ sample, the daughter product ^{208}Tl is responsible for the emission of gamma rays with 2614 keV energy which is capable of pair-production. The deposition of this energy resulting in pair production, and high chance of both annihilation photons escaping due to small detector volume, could result in a peak at 2612-1022=1590 keV, which is referred to as Double Escape Peak (DEP). Single escape event could result in 2614-511=2103 keV peak.

4.2 Efficiency Calculations for Comparision:

In order to make a comparison between the two detectors, we estimate the normalized efficiency. Even though the times of acquisition are not the same, the number of

counts(N_E) for the low energy peak at 238 keV, medium energy peaks at 913 and 966 keV, and the largest energy peak at 2614.5 keV within the intervals that were chosen (see Table 3) about their photopeak width are determined. The total events N for the whole spectrum is also determined. The ratio $\frac{N_E}{N}$ indicates a normalized efficiency η which can be used to compare the different detectors as long as the source is the same. However, since N_E includes Compton events from high energy peaks if any, the ratio may be overestimated and is only a reliable indicator for the highest energy peak, which in our case is 2614.5 keV. It is observed that the PMT based detector has an order of magnitude better normalized efficiency as compared to our spectrometer for this peak. The reduced overall efficiency of the smaller volume scintillator can be compensated by using longer acquisition time intervals. Since gamma interaction lengths are a function of energy, the smaller volume scintillator also shows a lower normalized efficiency ratio for the 2614keV gamma rays compared with the 238.6keV gamma rays.

5 Applications and Conclusions

5.1 Spectra of Monazite sample obtained using CsI Scintillator mated to PIN photodiode:

A 1gm Monazite sample procured from the sands of Kerala, and glued using Araldite epoxy on a 1" planchette was placed in front of the GammaSpec1K spectrometer, and the data acquired over a period of 14 hours is shown in Figure 6. All the expected peaks are identified and marked in the spectrum. It matches the one for $ThNO_3$ salt obtained in the lab, completely validating the utility of our spectrometer for environmental radiation monitoring.

5.2 Conclusions

We find that even though the volume of the scintillator used in our detector is around 130 times less than a standard 2" NaI scintillator, it is still usable for isotope identification. An added advantage is that high voltages are not required since our scintillator is mated with a photodiode instead of a PMT. The large volume scintillator however, has better photopeak efficiency, and lower Compton scattering, resulting in clearer peak formation. The capability to identify all the relevant peaks in Monazite sample makes our low cost, compact, portable USB powered Gamma Spectrometer an ideal choice for online radiation mapping of large geographical locations.

6 Acknowledgment

The authors thank Inter University Accelerator Centre(IUAC), New Delhi, for taking educational initiatives through their outreach programme, and providing a healthy environment for research and teaching in physics education. Prof. Sastri would like to thank IUAC for awarding Visiting Associateship to facilitate this work. We would

Detector	Total Counts N	Energy(keV)	Energy Interval(keV)	N_E	$\eta = N_E/N$
2″ NaI(Tl)		230.4	194.71-265.22	1079108	0.159
With	6768879	913.95+968.9	840-1030	5443016	0.08
PMT		2602.2	2516.83 - 2689.34	73771	0.01
10*10*8mm		235.6	201.26 - 270.11	856272	0.293
CsI(Tl) with	2914071	913.6+966.6	840 - 1030	122310	0.04
Photodiode		2607.2	2521.05-2693.18	2521.05	0.0024

Table 3: Determination of normalized efficiency for various energy peaks

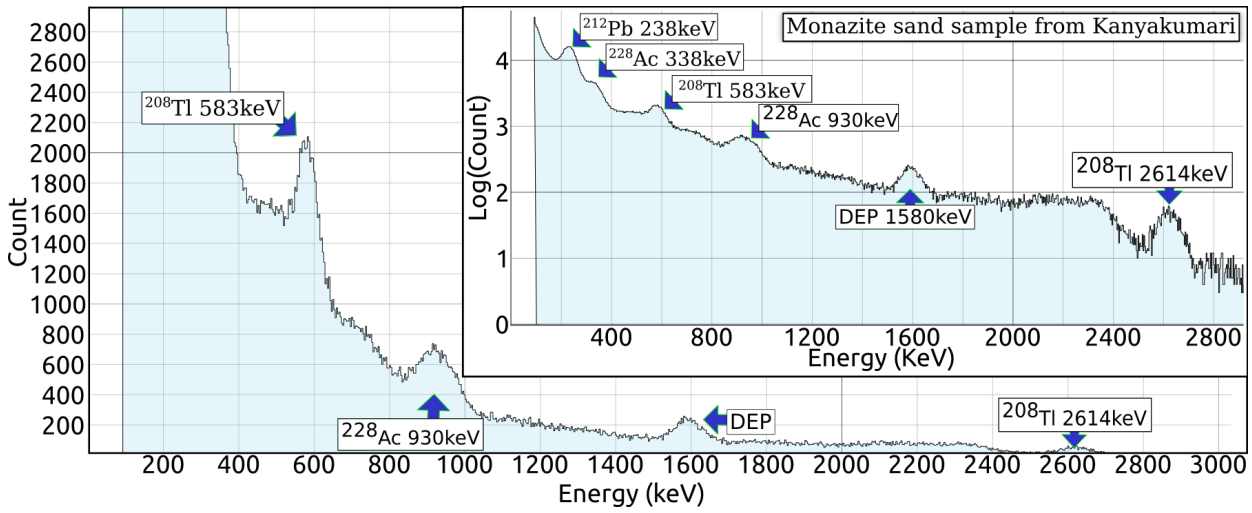


Figure 6: The Gamma spectrum of Monazite sample from the sands of Kerala. The peaks match those obtained from the Thorium Nitrate powder

also like to thank Dr. K Sreekumar from Indian Rare Earths Ltd for introducing us to monazite sand and its expected characteristics.

References

[1] Swapna et.al 'Alpha Spectrum of 212-Bi Source Prepared using Electrolysis of Non-Enriched ThNO3 Salt', Physics Education-IAPT Volume 35: No. 1. Jan-Mar 2019.

[2] Jithin B.P., O.S.K.S Sastri. DAE Symp. on Nucl. Phys. 63 (2018) 1072 (pp. 1072–1073).

[3] Jithin et.al 'Measurement Model of an Alpha Spectrometer for Advanced Undergraduate Laboratories', Physics Education-IAPT Volume 35: No. 1. Jan-Mar 2019.

[4] Source code for CNspec https://github.com/csparkresearch/cnspec/

[5] https://csparkresearch.in/alphaspectrometer/

[6] Zwickl, Benjamin M., Noah Finkelstein, and Heather J. Lewandowski. "The process of transforming an advanced lab course: Goals, curriculum, and assessments." American Journal of Physics 81, no. 1 (2013): 63-70.https://aapt.scitation.org/doi/10.1119/1.4768890

[7] Evaluated Nuclear Structure Data File Search and Retrieval, NNDC, Brookhaven National Laboratory. https://www.nndc.bnl.gov/ensdf/

[8] "Proposed syllabus and Scheme of Examination for B.Sc.(Honors) Physics" https://www.ugc.ac.in/pdfnews/7756304_B.SC.HONOURS-PHYSICS.pdf

[9] Martin, Brian R. Nuclear and particle physics: an introduction. John Wiley & Sons, 2006.

[10] Jithin B.P., O.S.K.S Sastri. DAE Symp. on Nucl. Phys. 64 (2019)

[11] https://csparkresearch.in/gammaspec/

[13] Gamma Detector white paper, First Sensor, DE. https://www.first-sensor.com/cms/upload/datasheets/gamma-ray-detection.pdf

[14] Tuli, Jagdish K. "NUCLEAR WALLET CARDS." (2011).

[15] CAEN N968 Spectroscopy Amplifierhttps://seltokphotonics.com/upload/iblock/dc3/dc32e13030d9a5c2fc1eb002e766554a.pdf

[16] CSpark Research Multi-Channel Analyzers https://csparkresearch.in/mca1k

[17] Franklin, A., "The epistemology of experiment" in Gooding, D. et al. (eds.) (1989) The uses of experiment: Studies in the natural sciences, Cambridge, UK: Cambridge University Press, 437-460.

[12] Sood, P. C. and Jain, Raj Kumar and Sastri, O. S. K. S., Unusual features of β transition rates in heavy deformed nuclei, Vol 69 Issue 5, May 2004, PhysRevC.69.057303, https://link.aps.org/doi/10.1103/PhysRevC.69.057303

Energy density and susceptibility of warm and dense quark matter

Kausik Pal

Department of Physics, Serampore College,
Serampore, Hooghly, 712201, West Bengal, India

kausik.sinp@gmail.com

Submitted on 20-04-2020

Abstract

The expressions for the ground state energy (GSE) density and the spin susceptibility (χ) of warm and dense interacting quark matter have been derived including the corrections due to correlations. For this, we have calculated the GSE of spin-polarized quark matter at non-zero temperature containing the term up to order ($g^4 \ln g^2$). It has been shown that, χ^{-1} changes sign and turns negative at different temperatures for different densities, which indicates a physical instability or impermanence of ordered phase of quark matter.

1 Introduction

The study of properties of interacting hadronic or quark matter has been an area of contemporary high energy physics research for quite some time now [1-6]. Such studies are usually made at extreme conditions of temperature and/or density where perturbative calculations might become meaningful. This is due to the fact that at such extreme conditions, the hadrons are expected to dissolve into their more fundamental constituents viz, the quark and gluon, forming a new state of matter called Quark Gluon Plasma (QGP) and the theory that describes the matter is known as Quantum Chromodynamics (QCD). One extreme limit is the region of low chemical potential and high temperature. This corresponds to conditions very similar to the deconfined phase that might have existed in the early Universe before hadronization. Experimentally the heavy ion collisions provide opportunities to verify some of the theoretical results against measurements. Intense research activities have been performed at Relativistic Heavy Ion Collider (RHIC) experiments at Brookhaven National Laboratory (BNL), the Large Hadron Collider (LHC) experiments at CERN. In contrast, the experiments at GSI will focus on the compressed baryonic matter i.e., matter with high baryonic chemical potential compared to temperature. The low temperature and high chemical potential region corresponds to the phases which may be found at the core of neutron star. Such studies are important to understand the properties of astrophysical compact objects. For example, experiments at Einstein Laboratory, ROentgenSATellite (ROSAT), CHANDRA have been performing measurements to understand the properties of neutron stars, quark stars, pulsars, quasars etc. For

theoretical verification, we need various estimation which depend on temperature, density, pressure, energy etc.

There is a possibility that at the core of the neutron stars the density may go up to 5 ~ 6 times the normal nuclear matter density where the matter is not expected to be in the hadronic phase. In fact, under such scenario one expects that it would be more appropriate to describe the core of such dense stars as degenerate quark matter, which is our main interest in the present manuscript. Due to asymptotic freedom of QCD, it is predicted that hadronic matter can undergo a series of phase transitions like confinement-deconfinement and (or) chiral phase transitions. In the high density limit QCD predicts the existence of color superconducting state. These aside the possibility of spin polarized quark liquid i.e., the existence of ferromagnetic phase in dense quark system has also been suggested by various authors [1,3]. Presence of extraordinarily high magnetic field (~10^{15} Gauss) in the core of neutron stars can be attributed to the existence of spin polarized quark matter [1,3-6]. To examine the possibility of ferromagnetic phase transition in dense quark matter in Ref.[1] a variational calculation was performed where it was shown that quark liquid interacting through one-gluon exchange (OGE) shows spontaneous magnetic instability at low densities.

The underlying mechanism of such a phase transition for slow moving massive quark is similar to what one observes in case of interacting electron gas in a neutralizing positive charge background where the electrons interact only by the exchange interaction and the contribution of the direct term cancels with the background contribution. In case of interacting electron gas, the kinetic energy is minimum in unpolarized state, the exchange energy, on the other hand, favours spin alignment. These are two competing phenomena which also depends on density. It has been seen that the kinetic energy dominates at higher density and as the density is lowered the exchange energy becomes larger at some point turning the electron gas suddenly into a completely polarized state. This is the mechanism of ferromagnetism in electron gas interacting via. Coulomb potential.

The exchange energy for quark matter interacting via one gluon exchange is also attractive and becomes dominant at some density giving rise to ferromagnetism [1,3,4,5]. However, there are similarities and differences between quark matter and electron gas as discussed in Ref.[1]. For slow moving massive quarks the dynamics is very similar to what happens in electron gas, while in the relativistic case a completely different mechanism works when spin dependent lower component of the Dirac spinor becomes important. It should also be noted that the exchange energy is negative for massive strange quark at low densities while it is always positive for massless u and d quark as observed in Ref.[1,3].

The properties of dense and warm quark matter are particularly relevant for the study of various astrophysical phenomena. In the present context, I calculate the energy density at non-zero temperature of spin asymmetric quark matter [3,5] and consequently determine the spin susceptibility [4,6] with corrections due to correlations, i.e., containing terms up to $\mathcal{O}(g^4 \ln g^2)$.

In the present theoretical model, for the computation of ground state energy (GSE) density and the corresponding spin susceptibility, it has been assumed that the QGP matter is composed of weakly interacting massless up and down quarks, and the gluons which are treated as almost free. It is well known that QGP is asymptotically free. Thus one may expect that quarks and gluons interact weakly and the properties of the interacting quark matter might be computable perturbatively.

So it is acceptable to assume the coupling constant $\alpha_c = g^2/4\pi < 1$. This is well known that for weak coupling constant g, perturbation theory can only be worked out to a finite order, so that the series expansion in terms of α_c to converge. It is to be mentioned, although the perturbative expansion converge very slowly, the approach makes predictions consistent.

To calculate the energy densities and other relevant quantities, we use the Lorentz invariant color symmetric forward scattering amplitude of two quarks around the Fermi surface interacting by the one-gluon-exchange [1,3,6]. The direct term does not contribute as the trace of single color matrices vanishes. The first correction due to interaction to the GSE, is given by the exchange energy density. This arises from two quarks interchanging positions in the Fermi sea by exchanging a virtual gluon. However, to compute the ground state energy density, we have to evaluate the diagrams beyond the exchange loop i.e., require to include the contribution of bubble diagrams. Without such corrections, however, the calculations are known to remain incomplete as the higher order terms are plagued with infrared divergences arising out of the exchange of massless gluons, indicating the failure of naïve perturbation series. This can be accomplished by summing a class of diagrams known as bubble diagrams and receives logarithmic corrections. Here, however, the calculation is little tricky, as in this situation one cannot deal with spin averaged values for the polarization tensor. In the present work, we calculate the density temperature dependent ground state energy and the spin susceptibility with correlation corrections within the one-gluon exchange model for warm and dense plasma and also shows how these quantities can be expressed in terms of spin polarization parameter ξ and the temperature T.

2 ENERGY DENSITY AND SPIN SUSCEPTIBILITY

Let us consider the quasiparticles whose spins are all eigenstates of the spin along z direction. Since quarks are spin polarized, the quasiparticles interaction ($f_{pp'}^{ss'}$) have two parts, when spin is parallel ($s = s'$) which is known as spin non-flip ($f_{pp'}^{nf}$) interaction and the spin antiparallel ($s = -s'$) interaction correspond to the spin flip ($f_{pp'}^{f}$) scattering amplitudes, such that [1,3,6]

$$f_{pp'}^{ss'} = f_{pp'}^{nf} + \frac{1}{2} f_{pp'}^{f} \quad (1)$$

where p and p' are the momentum of the quasiparticles. Here, the factor 1/2 is due to the equal scattering possibilities involving spin-up spin-down and spin-down spin-up quarks. Since spin and momentum have randomly oriented, we take the average over the angles θ_1 and θ_2 corresponding to spins s and s'. The angular averaged interaction parameter is given by [3,6]

$$f_{pp'}^{ss'} = N_f N_c^2 \frac{g^2}{9pp'} \int \frac{d\Omega_1}{4\pi} \int \frac{d\Omega_2}{4\pi} [1 + (\hat{p}.\hat{s})(\widehat{p'}.\widehat{s'})] \quad (2)$$

where g is the coupling constant. Here, N_f=2 and N_c=3 are the flavor and color degeneracy factors. In our calculations, it has been assumed that the contributions of u and d quarks are equal. Since we are dealing with ultra-relativistic massless quarks, the quantities which we want to calculate are obtained as a function of temperature (T) and Fermi momentum (p_f). As the Fermi momentum is not a suitable experimental quantity, we calculate the quantities in terms of baryon density. Since the baryon number is 1/3, we have baryon density $n_b = \frac{1}{3} n_q$, where $n_q = n_q^+ + n_q^-$,is the quark density, where the superscript denotes the spin-up (+) and spin-down (-) quarks respectively.

The leading contributions to the energy density of quarks are given by the kinetic, exchange and correlation energy densities

$$E = E_{kin} + E_{ex} + E_{corr} \quad (3)$$

Including the color and flavor degeneracy factors for quarks, the total kinetic energy density for spin-up and spin-down quarks is given by [6]

$$E_{kin} = \frac{3}{(2\pi)^2}\Big(p_f^4 \left[(1+\xi)^{4/3} + (1-\xi)^{4/3}\right] + 2\pi^2 T^2 p_f^2 \left[(1+\xi)^{2/3} + (1-\xi)^{2/3}\right]\Big) \quad (4)$$

Here $\xi = (n_q^+ - n_q^-)/(n_q^+ + n_q^-)$ is the spin polarization parameter with condition $0 \le \xi \le 1$.. The first leading order correction due to interaction to the energy density is the exchange energy. The exchange energy density E_{ex} was calculated in Ref.[3] within Fermi liquid theory approach and also directly evaluating two loop diagrams in Ref.[6]. The analytical expression for the exchange energy density at finite temperature is found to be [6]

$$E_{ex} = \frac{g^2}{(2\pi)^4}\Big(p_f^4 \left[(1+\xi)^{4/3} + (1-\xi)^{4/3} + 2\big(1-\xi^2\big)^{2/3}\right] + \frac{4}{3}\pi^2 T^2 p_f^2 \left[(1+\xi)^{2/3} + (1-\xi)^{2/3}\right]\Big) \quad (5)$$

The next higher order correction to the energy density beyond the exchange term is the correlation energy. The detailed calculation of correlation energy have been given in Ref.[6], which we quote here. The leading $g^4 \ln g^2$ order contribution is given by

$$E_{corr} = \frac{g^4 \ln g^2}{(2\pi)^4}\frac{1}{8}\Big(p_f^4 \left[(1+\xi)^{4/3} + (1-\xi)^{4/3} + 2\big(1-\xi^2\big)^{2/3}\right] + \frac{16}{3}\pi^2 T^2 p_f^2 \left[(1+\xi)^{2/3} + (1-\xi)^{2/3}\right]\Big) \quad (6)$$

It is to be noted that all the energy contribution for various ξ are always positive for massless quarks.

The other quantity which have direct physical relevance is the calculation of spin susceptibility. Once the energy density is known, we can readily calculate the spin susceptibility. The spin susceptibility is determined by the change in energy of the system as its spin are polarized. In the small ξ limit, the energy density can be expanded as [3,6]

$$E(\xi) = E(\xi = 0) + \frac{1}{2}\beta_s \xi^2 + \mathcal{O}(\xi^4) \quad (7)$$

where β_s is the spin stiffness constant which is inversely proportional to the spin susceptibility χ; mathematically $\chi = 2\beta_s^{-1}$. Since energy density involves three leading terms, the spin susceptibility have also three contributions, given by

$$\chi^{-1} = \chi_{kin}^{-1} + \chi_{ex}^{-1} + \chi_{corr}^{-1} \quad (8)$$

With the help of equations (4) – (7), each energy contribution to the susceptibility is

$$\chi_{kin}^{-1} = \frac{p_f^4}{3\pi^2}\left(1 - \frac{\pi^2 T^2}{p_f^2}\right) \quad (9)$$

$$\chi_{ex}^{-1} = -\frac{g^2 p_f^4}{18\pi^4}\left(1 + \frac{\pi^2 T^2}{3p_f^2}\right) \quad (10)$$

$$\chi_{corr}^{-1} = -\frac{(g^4 \ln g^2) p_f^4}{576\pi^6}\left(1 + \frac{4\pi^2 T^2}{3p_f^2}\right) \quad (11)$$

It is to be noted, contribution from kinetic term may be positive or negative depending on the values of T and p_f, while the exchange and correlation contribution is always negative. These three competing terms, as we shall see, turns the susceptibility from positive to negative. Using all the equations (8) – (11), the total spin susceptibility is given by

$$\chi = \chi_P \left[1 - \frac{g^2}{6\pi^2}\left(1 + \frac{4\pi^2 T^2}{3p_f^2} + \frac{\pi^4 T^4}{3p_f^4}\right) - \frac{g^4 \ln g^2}{192\pi^4}\left(1 + \frac{7\pi^2 T^2}{3p_f^2} + \frac{4\pi^4 T^4}{3p_f^4}\right)\right]^{-1} \quad (12)$$

where χ_P is the Pauli susceptibility. The value of χ can be estimated if coupling constant (g), temperature (T) and Fermi momentum (p_f) are exactly known.

3 RESULTS AND DISCUSSION

In figure 1, we have plotted the density dependence variations of ground state energy per baryon density of quark matter for various order parameter ξ at temperature 30 MeV. It has been observed that the energy density is larger with higher value of ξ and energy increases with baryon density. Therefore spin polarized quark matter is more unstable compared with unpolarized quarks.

In figure 2, we have plotted the temperature dependence of the inverse spin susceptibility per baryon density for two different densities. It has been observed that the susceptibilities blow up at different temperatures for two different densities. This can be identified as a physical instability or inconstancy of the quark matter towards the ordered phase. This is equivalent to what happens to the ground state energy as a function of ξ.

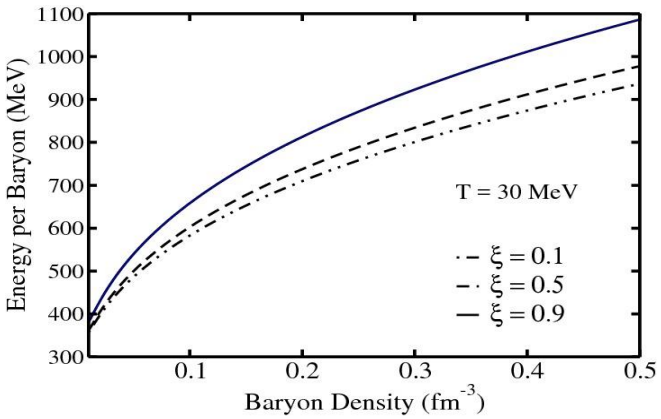


Figure. 1 – Density dependence of ground state energy density of quark matter with three different spin polarization parameter ξ at temperature 30 MeV.

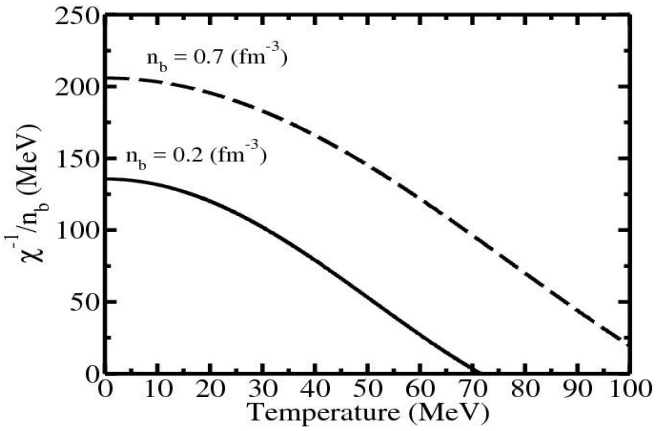


Figure. 2 – Variations of inverse spin susceptibility (χ^{-1}) with temperature of quark matter for two different baryon densities.

Therefore quark matter is more stable with unpolarized quarks. In other words, the equation of state of spin asymmetric quark matter becomes stiffer by increasing the baryon density or the temperature.

Acknowledgments

I am indebted to late A.K. Dutt-Mazumder who introduced me to this topic and his fruitful discussions motivated me to initiate the present work.

References

1. Tatsumi T. “Ferromagnetism of quark liquid”. Phys. Lett. B. 2000; 489: pp. 280.

2. Baym G, Chin S.A. “Landau theory of relativistic Fermi liquids”. Nucl. Phys. A. 1976; 262: pp. 527-538.

3. Pal K, Biswas S, Dutt-Mazumder A.K. "Spin-dependent Fermi liquid parameters and properties of polarized quark matter". Phys. Rev. C. 2009; 79: pp. 015205.

4. Pal K, Dutt-Mazumder A.K. "Spin susceptibility of degenerate quark matter". Phys. Rev. C. 2009; 80: pp. 054911.

5. Pal K, Biswas S, Dutt-Mazumder A.K. "Ground state energy of spin polarized quark matter with correlation". Phys. Rev. C. 2009; 80: pp. 024903.

6. Pal K. "Correlation corrections to the thermodynamic properties of spin-asymmetric QGP matter". Eur. Phys. J. Plus. 2015;130: pp. 23.

P.W. Anderson (1923-2020) - He influenced the course of Physics and Physicists

G. Baskaran

The Institute of Mathematical Sciences and IIT Madras,

Chennai 600113

Department of Physics, Indian Institute of Technology,

Chennai 600 036

Perimeter Institute for Theoretical Physics, Waterloo,
Ontario, Canada

P.W. Anderson was a giant among theoretical physicists in the second half of last century. He passed away at the age of 96 at Princeton, NJ, USA on 29th March 2020. He remained scientifically active till the end. In addition to transforming a field he also educated the physics community about nature of science and paved new paths. A 2006 survey rated him as the most creative living physicists.

Anderson's approach to theoretical physics was somewhat special. He communed with nature by contemplating on body of the available experimental results and phenomena on down to earth materials, often from the point of view of quantum organization. He nurtured a whole field of condensed matter physics over 70 years and influenced several generation of scientists. His work had implications beyond condensed matter - origin of mass of elementary particles (electron, quark etc.) via Anderson-Higgs mechanism, understanding a certain anomalous behavior of pulsars, nano electronics, issues in quantum computation, neural networks, pre-bioltic evolution, complex optimization problems, stock markets etc.

He received the Nobel Prize in 1977 and shared it with van Vleck and Nevil Mott. Anderson was also responsible for about half a dozen Nobel Prizes, directly or indirectly. In general Anderson helped and nurtured good ideas, wherever it came from. He also tried to be just to his profession by being forthright and critical (sometimes in undiplomatic fashion). Beyond science, he longed for peace on earth and a healthy planet; opposed Vietnam war, argued for nuclear disarmament, argued against star wars etc. He leaves behind a world of admirers.

Various factors inspire young ones into science. Often it is teachers. Families also nurture free spirit and a spirit of enquiry. In his Nobel Prize biographical note Anderson says, "At Illinois my parents belonged to a group of warm, settled friends, whose life centered on the outdoors and in particular on the "Saturday Hikers", and my happiest hours as a child and adolescent were spent hiking, canoeing, vacationing, picnicking, and singing around the campfire with this group. They

were unusually politically conscious for that place and time, and we lived with a strong sense of frustration and foreboding at the events in Europe and Asia … . I remember at school was a first-rate mathematics teacher at the University High School, Miles Hartley, ...".

Anderson went to Harvard, finished undergraduate studies in Physics with a good record, and completed a Ph.D. According to him "Graduate school (1945-49) consisted of excellent courses; a delightful group of friends, … , centered around bridge, puzzles, and singing; .." . Joyce Gothwaite became his life partner. Anderson had 73 years long life journey with Joyce. Their daughter Susan was born in 1948. Joyce and Anderson were a wonderful couple and mutually supportive. Behind every successful man there is a woman – Joyce was always behind Anderson.

Anderson grew in the atmosphere of the Bell Telephone Laboratories for 35 years and moved to Princeton for another 35 years. Bell Labs, an industrial research lab at USA, was unique in the world of science and technology in the last century. Graham Bell decided to spent a percent or so of his profit from his telephone company, on open ended basic science, both experiment and theory. Such explorations at Bell Labs resulted in profound advancements in basic science, resulting in several Nobel Prizes. Two Nobel Prizes in extremely contrasting domains are, a) discovery of 30 background radiation from the Big Bang (beginning of the universe), and b} discovery of solid state transistor, which has profoundly modified science, technology and our society in turn.

I am proud to say that Perimeter Institute of Theoretical Physics at Waterloo, Canada, where I have a joint appointment, was born as a result of love and appreciation of Mike Lazaridis for theoretical physics. He says thattt it can help society in the long run, as it has done in the past. Incidentally, Mike established Black Berry mobile phone company. My fond hope is that science philanthropy will grow in India and take Indian science to new heights.

Before we get a glimpse of Anderson's science we will see him as a person. It was 1986. I was talking to Anderson in his office at Princeton. A young undergraduate walked in. Students knew that no prior appointment was necessary to meet this Nobel Laureate in his office. The student elaborated an idea he had. After a while I became impatient. Anderson patiently listened till the end; he did notice my impatience. After the student left, Anderson enlightened me, with his own understanding of a very important point that the student was actually trying to convey, but didn't completely succeed. I woke up. On several occasions Anderson brought out deep meanings in very ordinary sounding seminar talks and pleasantly surprised the speaker and audience.

Anderson was amazingly supportive of his young associates. In 1986, while at Princeton in Anderson's group, I faced some issues in renewal of US visa for third year. News reached Anderson. He told me, `I will call George Schultz (then Secretary of states of the US Government, a Princetonian whom Anderson knew personally) and fix it up'. I felt elated, but politely stopped Anderson. That particular problem got sorted on its own. The third year at Princeton became a memorable period in my scientific career.

Anderson had a joint appointment with Cambridge University for a number of years. In 1962, in his lectures on manybody theory at Cambridge Anderson made a passing remark on the possibility of an anomalous current between weakly coupled superconductors, as a consequence of BCS theory. Brian Josephson, a student at the class took

Andersons suggestion seriously and proved the existence of such an anomalous current using BCS theory and showed it to Anderson. Impressed by the calculation Anderson encouraged young Josephson to publish it. More remarkably, after returning to Bell Labs that year, Anderson and his experimental colleague John Rowell performed experiments that demonstrated the Josephson effect, thereby removing a widely prevailing scepticism in the community on the possibility of such an effect. Tje rest is history and Josephson received his Nobel Prize in 1973.

Two experimental Nobel Laureates Dough Osheroff and Dan Tsui were explicit, in a meeting at Aspen Physics Center (2000) honoring Anderson's 50 years of contribution to condensed matter physics: they said that their Nobel Prize winning discoveries were strongly influenced by Anderson's timely alert and advice about some anomalies (which could have been easily ignored) in their measrements. There are few more such stories.

Anderson was a model builder. He looked at a body of experimental results, related to anomalous behavior. His deep contemplation and appreciation of unusual aspects (which often escaped the community) resulted in simple mathematical models, mostly quantum mechanical. There are more than half a dozen models in the name of Anderson. His analysis of his own model and finding quantitative solution were also unusual. Because of a resonance with nature, his science advanced by leaps and bounds, often resulting in paradigm shifts.

Since Anderson's models abstracted the essence of certain universal quantum and cooperative phenomena in condensed matter, its applicability was universal and went beyond its place of birth. Anderson's insights into superconductors resulted in Anderson-Higgs mechanism of mass generation of elementary particles and the celebrated Higgs particle. His understanding of vortex creep in a superconductor like Niobium resulted in understanding glitches in pulsar periods, as arising from neutron star quakes initiated by vortex creep in neutron superfluid in the crust. Edwards-Anderson spin glass model lead to solution to optimization problem, advances in neural network theory etc.

Anderson's major contributions are broadly, in the field of localization of waves, quantum magnetism and superconductivity. In recent times `manybody localization' phenomenon is a very active field in quantum condensed matter. It is about localization or the non propagating character of multiparticle states in Hilbert space, when it faces disorder and or strong interactions. In 1958 Anderson, inspired by some experiments identified this as an important problem and illustrated its complexity using a toy model in involving one particle Hilbert space. This toy model was partly responsible for his Nobel Prize. Anderson localization was revived in 1979, in a classic paper on scaling theory of localization by Abrahams, Anderson, Licciardello and Ramakrishnan. The evival of study of one particle and manybody localization phenomena continues to generate remarkable experimental and theoretical developments, including throwing light on issues of future quantum computers.

Anderson's decades of work on strongly correlated electron systems and quantum magnetism resulted in the notion of quantum spin liquid (borrowing an early idea of Pauling's resonating valence bond states and applying it to a totally new and experimentally motivated context) in 1973. This

foundational work, a culmination of Andersons unique insights into quantum matters remained relatively unknown (exceptions are Patrick Fazekas, Sriram Shastry, Baskaran and few others). This work, Anderson's own creation became handy in explaining a milestone discovery of high temperature superconductivity in cuprates by Bednorz and Muller, in 1986. Resonating valence bond (RVB) theory of high Tc superconductivity by Anderson changed the course of the field of superconductivity experimentally and theoretically. It also changed the scientific career of many physicists.

I happened to be there at Princeton and became a long-term partner of Anderson on the RVB theory of high Tc superconductivity, from the beginning (to be precise early November 1986). RVB theory continues to influence the field even now. It has answered major and key questions about mechanism of superconductivity and high Tc superconductivity phenomenology. However, many normal state properties at optimal doping and metallurgical complications (to use Anderson's words) in under doped regime remain challenging from many-body theory point of view. Generation of young physicists are actively involved in subfields that arose from RVB theory. Interestingly some of these developments are far removed from known and compelling condensed matter contexts and phenomenology. However, they are exposing unknown territories of the rich world of mathematics to condensed matter physicists.

Anderson's view of physics is nicely summarized in his influential article `More is different' (1972). In this article he emphasized the role of emergence and hierarchical nature of science in general and physical sciences in particular. He emphasizes that what is at work in most of natural world and science is emergence . Hierarchy of new properties, surprises and complexities emerge at each energy, length, time or other relevant scales. This is how science gets created. In this process the web of science and a grand unity also reveals itself – unexpected connections such as connection of Anderson's mechanism of plasmon gap in superconductors to origin of mass of elementary particles emerge. Spin glass model becomes a model for some aspect of brain function via Hopfield's neural network theory and so on.

Recognizing key elements and new notions in the hierarchy and understanding them are challenging in their own right. To make real progress it requires great intellectual efforts, new experimental methods and new mathematics and skills. To put it plainly, Anderson criticized naive attitude such as `what is important and basic is Dirac equation, rest is chemistry'. In Anderson's mind nature is inexhaustible and it will continue to surprise us and keep science and scientists busy if only we have a open mind and follow nature closely.

It was 3rd July 2004, a day before Higgs particle discovery was to be announced officially by CERN. I was visiting Anderson at Princeton (Figure 1). Physics Department was jubilant because of the impending announcement of discovery of `Anderson-Higgs' particle. Anderson told me, `I am happy for Peter (Higgs)... Do I want one more Nobel ? No'. His mind instantly shifted to physics. He started asking, `is Higgs field a collective mode of some fermi sea, has CERN determined spin and quantum numbers of the Higgs particle …'. Physics kept Anderson happy till the very end. Anderson made others happy.

Figure 1. P.W. Anderson and the author at Princeton on 3rd July 2004, the day before CERN officially