PHYSICS EDUCATION

Volume 37, Number 3

In this Issue

Electrodynamics : The Maxwellian Quest

Kaustav Saha[1], Tuhin Sur[2] and Dr. Upendranath Nandi[3]

[1]Student, B.Sc Physics (Hons.), Maulana Azad College
(Affiliated to The University of Calcutta)
[2]Head of the Department, Physics, Scottish Church College
(Affiliated to The University of Calcutta)
[3]Member of the Indian Association of Physics Teachers
unphys@scottishchurch.ac.in

Abstract

The main objective of this article is to present the chronological development of electrodynamics as followed by J. C. Maxwell to the undergraduate students. The four fundamental equations of electrodynamics as presentable to the undergraduate are derived with a focus on the lives of the discoverers. To make the content interesting, some historical aspects of the subject are highlighted. Simple experiments are discussed in support of the evidence of some theoretical calculations. The concept of displacement current which is less understood by the students is presented in detail. We hope that students will be benefited much by reading this article.

1 The Physics of Electromagnetism

Modern life has been oversimplified by the discovery of electromagnetism. Although people have been applying the principles of electromagnetism since ancient times, yet the theory itself was discovered long after. The most common example of electromagnetism which comes with a promise that everyone on Earth is acquainted with, is lightning. Also, the early Greeks had noticed that amber had the property to attract lightweight objects like feathers, small pieces of hay, etc. A student in physics can definitely recall from his knowledge of previous classes that it is due to static electricity. Even ancient Egyptians knew animals like eels or catfish, who gave out electric shocks.

As time went on, compasses changed the way navigation was done in those days.

The Industrial Revolution, as it progressed, moved away from mechanical machines into electrical ones. In today's world, there exist a very few devices in everyday use which do not rely on electromagnetism for their functioning.

Sophistication has really been the key for evolving human life as we see it today. It is generally known that current flowing through the filament of a light bulb heats it, causing the luminosity of the filament. Electric clocks use electric impulses to measure time, which displaces a lever and it is again brought back to its original position by an electromagnet. Combine these two devices, and one might develop a system of traffic lights, the clocks keeping track of when and for how long, which bulb might glow. A modern computer, which is fundamentally built on the principles of electromagnetism can be further used to create an algorithm that automatically controls this, based on the traffic on the road. Speaking of traffic, all automobiles use currents in an electric motor, which generate magnetic fields that rotate the motor shaft and drive engine pistons to compress an explosive mixture of gasoline and air; the spark initiating the combustion is an electric discharge, which in-turn makes up a momentary current flow.[7]

Many distinguished physicists studied the various aspects of this fascinating subject, and reported their observations to the scientific community between the late sixteenth and eighteenth centuries. Sadly, but unsurprisingly, the accounts of various physicists varied. It was not until James Clerk Maxwell, who unified the electromagnetic theory, and published a treatise on it. It was this unification, which led to the discovery of the electromagnetic nature of light. We shall attempt to understand this treatise, and the four equations central to that treatise, known as Maxwells equations.
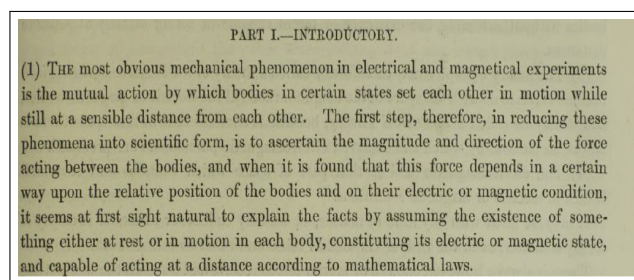
Figure 1: The original treatise by James Clerk Maxwell.

But before diving into his equations, it might be interesting to know a bit more about him.

2   James Clerk Maxwell

Maxwell was born in a comfortable-middle class family, where he received his education from an nonintellectual private tutor, who classified Maxwell as a "sluggish learner" (imagine that!), even though Maxwell had a preternatural memory and a buzzing inquisition. Luckily, his Aunt rescued him and admitted him to school in the Edinburgh Academy. Like so many geniuses in the history of time, his interests ranged further than his school curriculum.[3]

Figure 2: James Clerk Maxwell.

Maxwell published his first research paper at the tender age of just fourteen. It described a generalized series of oval curves that could be traced with pins and thread by analogy with an ellipse. Throughout his career, Maxwell showed keen interest for geometry and mechanical models. Maxwell joined the University of Edinburgh at the age of sixteen. There, he published two more scientific papers. Later in life, Maxwell had a happy union and was appointed as a professor of natural philosophy at King's College, London.

It was here that Maxwell took a keen interest in electromagnetism, and published his two famous theses on electromagnetic field. He even supervised the experimental determination of electrical units for the British Association for the Advancement of Science, and this work in measurement and standardization led to the establishment of the National Physical Laboratory. He also measured the ratio of electromagnetic and electrostatic units of electricity and confirmed that it was in satisfactory agreement with the velocity of light, in agreement with his theory.

It was in the spring and summer of 1867 that Maxwell devoted his energies into writing a dissertation on electricity and magnetism, while on a tour of Italy. It was this research which led him to become one of the greatest physicists in history and his book *Treatise on Electricity and Magnetism* (1873) was famous. Those who were fortunate enough to work alongside Maxwell have said that Maxwell preferred geometrical approach to a problem instead of the conventional mathematical or analytical approach. Indeed, this was a huge advantage for Maxwell, and he was able to deduce several correct equations from faulty arguments.

In this article we will look into one of the key aspects of this book: the four equations proposed by Maxwell which were alleged to be the fundamental equations for the physics of electromagnetism.

3 The four equations that define electromagnetism

We will be talking about each of the four equations in quite some detail later in this article, but to first give a general idea, the statements of the four equations fundamental to the electromagnetic theory are given below:

- Electric field diverges from electric charge.

- There are no isolated magnetic poles, but the Coulomb force acts between the poles of a magnet.

- Electric fields are produced by changing magnetic fields.

- Circulating magnetic fields are produced by changing electric fields and by electric currents.

Maxwell modified the last law, originally proposed by Ampre, to include the time-variation of the field.

We shall proceed to understand each law by understanding the origin and try to develop the mathematics essential to give a mathematical form to each statement discussed above. However, we should be extra careful to distinguish between the law, originally proposed by other physicists, and how Maxwell modified it to create his four equations to describe the entire physics of electromagnetism.

4   The Gauss's Laws

The German mathematician *Carl Friedrich Gauss* is generally regarded as one of the *greatest mathematicians of all times* for his contributions to Number Theory, Probability Theory, the Planetary Astronomy, and, especially the **Potential Theory** (including **Electromagnetism**) on the consequences of which our discussion is based on.

It shall indeed be intriguing to see how Gauss' works in mathematics have snowballed into the corollaries that have been drawn from his hypothesis' in Physics. And before we focus our discussion on his contributions in context of this article, let us have a glimpse of the greatness of the mathematician in him.

4.1   Early Life

Gauss was the only child of poor parents. That he was a calculating prodigy made him rare among mathematicians, and he retained the ability to do elaborate calculations in his head most of his life. Impressed by this ability, his teachers, along with his devoted mother recommended him to the duke of Brunswick in 1791, who granted him financial assistance to continue his education, at first locally and then to pursue mathematics at the University of Gttingen from 1795 to 1798. Gausss pioneering work gradually established him as the times' pre-eminent mathematician, first in the German-speaking world and then further into his field of work.[2]

4.2   The Law of Flux: Origins of Potential Theory

Gauss published works on number theory, the mathematical theory of map construction, and many other subjects. In the 1830s he became interested in terrestrial magnetism and participated in the first worldwide survey of the Earths magnetic field (in-

fact, to measure it, he invented the magne-tometer). With his Gttingen colleague, the physicist Wilhelm Weber, he made the first electric telegraph, but a certain parochialism prevented him from pursuing the invention energetically. *Instead, he drew important mathematical consequences from this work for what is today called* **potential theory**, *an important branch of mathematical physics arising in the* **study of electromagnetism** *and gravitation.*

The term potential theory arises from the fact that, in 19th century physics, the fundamental forces of nature were believed to be derived from potentials which satisfied Laplaces equation[1]. Hence, *potential theory was the study of functions that could serve as potentials.*

Figure 3: Gauss and the Gaussian World in Physics, the reader for his/her interest can look up his contributions to geometrical constructions of polygons and astronomy as alluded to in the figure.

Originally studies were related to the

---

[1] The widely useful Laplace's Differential equation in physics, $\nabla^2 V = 0$ the solutions of which occur in problems of electrical, magnetic, and gravitational potentials.

properties of forces which follow the law of gravitation. In the statement of this law given by Sir Issac Newton in 1687 the only forces considered were the forces of mutual attraction acting upon two material particles of small size or two material points. These forces are directly proportional to the product of the masses of these particles and inversely proportional to the square of the distance between them. Thus, the first and the most important problem from the point of view of celestial mechanics and geodesy was to study the forces of attraction of a material point by a finite smooth material body - a spheroid and, in particular, en ellipsoid (since many celestial bodies have this shape)[2].

Lagrange established that the field of gravitational forces, as it is called now, is potential field. He introduced a function which was called a potential function in 1828 by G. Green and later in 1840 by Gauss – just a potential. At present, the achievements of this initial period are included in courses on ce-

---

[2] It is noteworthy to mention here that of the works which gave Gauss substantial recognition, the second publication was his rediscovery of the asteroid Ceres. Its original discovery in 1800 had caused a sensation but it vanished behind the Sun before enough observations could be taken to calculate its orbit with sufficient accuracy to know where it would reappear. Many astronomers competed for the honour of finding it again, but Gauss won with a novel method for dealing with errors in observations, today called the method of least squares. Thereafter Gauss worked for many years as an astronomer and published a major work on the computation of orbitsthe numerical side of such work was much less onerous for him than for most people.

lestial mechanics.

Meanwhile, we see that *Gauss and his contemporaries* had already discovered that the method of potentials can be applied not only to solve problems in the theory of gravitation but, in general, to solve a wide range of problems in mathematical physics, in particular, **in electrostatics and magnetism**. In this connection, potential became to be considered not only for the physically realistic problems concerning mutual attraction of positive masses, but also for problems with some equivalent of masses but having arbitrary sign in the problem, **i.e., charges**.[1]

5 Gauss's Law in Electrostatics

In physics, *the Gauss' Flux Theorem* is one of the most important theorems of electrostatics that relates the distribution of electric change to the resulting electric field. In integral form; it states that

> The flux of electric field out of any arbitrary closed surface is proportional to the electric charge enclosed by it, irrespective of how it is distributed.

This form is most useful when, by symmetry reasons, a closed surface can be found along which the field is uniform. When no such symmetry exists, Gauss' Law can be used in its differential form, which states that

> Divergence of the electric field is proportional to the local density of charge.

The law was first formulated by Joseph-Louis Lagrange in 1773, followed by *Carl Friedrich Gauss in 1813*, both in the context of the attraction of ellipsoids, as we mentioned earlier. It is, as we see, one of Maxwell's four equations, which form the basis of *classical electrodynamics*.

Let's begin our quantitative discussion with the simplest possible case: a single point charge q placed at the origin, whose field at position *r* is given as

$$E(r) = \frac{1}{4\pi\epsilon_0}\frac{q}{r^3}r. \qquad (1)$$

We might now sketch a few representative vectors as in Fig. 4-(a). Because the field falls off as $\sim 1/r^2$, the vectors get shorter as we move further from the origin. But conventionally, we connect up the arrows as in Fig. 4-(b), to form the field lines. It must be noted that this doesn't omit any information as regards the strength of the field since this data has now shifted from the length of the lines to their density in space, since this model involves at any instant a finite number of lines being drawn for our purpose, which precisely ought to be drawn consistently enough to include an accurate sense of the field[3]. Such diagrams are therefore also very convenient for representing more complicated fields.

In this model so described, the *flux* of E

---

[3]That is, if a charge q gets 8 lines (say), a charge of 2q deserves 16, which of-course, must be drawn preserving the symmetry of the situation, as illustrated in a subsequent diagram (Fig-5).

through a surface S, given as

$$\phi_E = \int_S E \cdot \hat{n}\, dS$$

is the *measure of the amount of field lines passing through S*. For a *given sampling rate*, the *flux* is proportional to the number of field lines drawn, and hence $E \cdot \hat{n}\, dS$ is proportional to the *number-density of the perpendicular component of field-lines through the infinitesimal area dS*.

This suggests that the flux through any *closed* surface can be a measure of the to-

tal charge inside. For, field lines that originate on a positive charge must either terminate[4] on a negative charge inside or pass out through the surface.

In case of a point charge q at the origin, the flux of E through a spherical surface of radius r with the polar and azimuthal angles θ and φ is

$$\oint E \cdot \hat{n}\, dS = \oint \frac{1}{4\pi\epsilon_0}\left(\frac{q}{r^2}\hat{r}\right)\cdot(r^2\sin\theta\, d\theta\, d\varphi\, \hat{r}) = \frac{1}{\epsilon_0}q \tag{2}$$
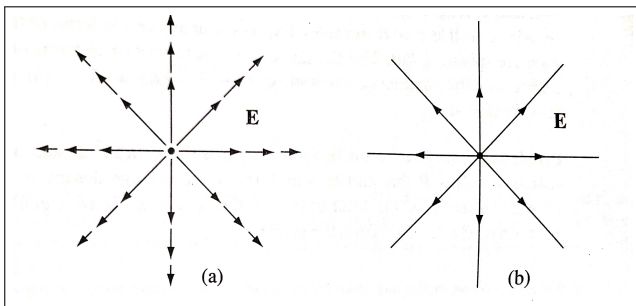

Figure 4: The Schematic representation of Fields.

Figure 5: Charges and respective field lines associated with closed surfaces.

which is independent of dimensions of the Gaussian surface. Now, for a bunch of n charges scattered about, the equation is modified according to the Principle of Superposition as

$$E = \sum_{i=1}^{n} E_i.$$

And thus, the flux through a surface that encloses them all is

This statement is, of-course, axiomatic; but can be satisfactorily understood by the following argument. Consider the charge

---

[4]Lines cannot terminate in mid-air because if they did, the divergence of E would not be zero, and (as we shall soon see) that cannot happen in empty space.
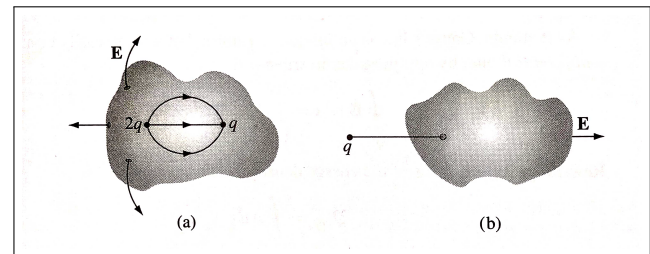
$$\oint E \cdot \hat{n}\, dS = \sum_{i=1}^{n} \oint E_i \cdot \hat{n}\, dS = \sum_{i=1}^{n} \frac{1}{\epsilon_0} q_i = \frac{1}{\epsilon_0} Q_{enc} \qquad (3)$$

where Q_{enc} is the total charge enclosed within[5]

q_i radiating photons at a constant rate in a spherically symmetric manner. Since E is an inverse-square field, the total number of photons n_p passing per unit time through S is proportional to the flux associated with it. Now, if the charge lies on the surface itself, n_p is obtained, due to spherical-symmetry, as the fraction $\frac{1}{4\pi}$ of the inside solid-angle Ω_i subtended by the surface S (open/closed). If S is a closed surface, we have $\Omega_i = 4\pi$.

the surface. This is the quantitative statement for Gauss' Law in Electrostatics. Although we note that it contains no additional information that was not already present in Coulomb's Law of Electrostatic(from which our discussion with this subject begins at many levels), yet this is a very powerful statement, as we shall see in subsequent discussions.

As it stands, Gauss' Law is an integral equation, but can be easily modified into a differential one by applying Gauss' Divergence Theorem :

$$\oint_S E \cdot \hat{n}\, dS = \int_V (\nabla \cdot E)\, dV \qquad (4)$$

Rewriting Q_{enc} in terms of the charge density ρ, we have $Q_{enc} = \int_V \rho\, dV$. So, Gauss' Law becomes $\int_V (\nabla \cdot E)\, dV = \int_V \left(\frac{\rho}{\epsilon_0}\right) dV$, which must hold for *any* volume. Hence, the integrands must be equal:

$$, \boxed{\nabla \cdot E = \frac{1}{\epsilon_0}\rho}$$

(**Gauss' Law in differential form**)

5.1 Applications

Gauss' Law is valid for *any* distribution of charges and for *any* closed surface. If we know the charge distribution, and if the problem has enough symmetry to let us evaluate the integral in Gauss' Law, we can find the field. Or, if we know the field, we can use the differential form (widely stated as the First of Maxwell's Relations in Electrodynamics) to find the charge distribution, such as charges on conducting surfaces.

In practical problems, we often encounter situations in which we want to know the electric field caused by a charge distribution on a conductor. *When excess*[6] *static charge is placed on a solid conductor, it resides entirely on the surface, not in the interior of the material*. Here is a simple, intuitive proof. We know from the nature of conductors that the field E at all points in the inte-

---

[6]By excess, we mean charges other than the ions and free electrons that make up the neutral conductor.

rior is zero[7]. Suppose we construct a Gaussian surface inside it. Because $E = 0$ at all points on it, *Gauss' Law requires that the net charge inside the surface is also zero. If this surface were to collapse to a point, there would be no charge at the point. And since this argument holds at all points within, we conclude that there is no charge at any point in the interior*, therefore arriving at what is known as *Electrostatic Shielding*, which also means that you are pretty safe from lightning strikes inside your car unless you touch the metal body! That's not too bad a chance of survival when struck in a thunderstorm for my readers : so what that they might get roasted; being electrocuted can altleast be ruled out!

6 Gauss's Law in Magnetostatics

Gauss' Law for magnetism, the second of four Maxwell's Equations that underline classical electrodynamics, is essentially a way of formulating the following observation.

> MAGNETIC MONOPOLES DO NOT EXIST AND THUS, RATHER THAN MAGNETIC CHARGES, THE BASIC ENTITY FOR MAGNETISM IS THE MAGNETIC DIPOLE[8].

---

[7]if E were not zero, charges would not be static in the conductor.

[8]If magnetic monopoles are ever found, the law would have to be modified, as elaborated later in our discussion.

6.1 Development

This idea of non-existence of magnetic monopoles dates back to 1269, as promulgated by a French-scholar Peter Peregrinus, whose work influenced the English physicist William Gilbert in his publication *De Magnete (1600)*. In the early 1800s, Faraday reintroduced this law after many contemporaries of his time had repeatedly failed to identify a magnetic monopole. The original idea that attributed the magnetism of lodestones to two different "magnetic fluids" (or what was termed as "effluvia"), a north-pole fluid at one end and a south-pole fluid at the other, which attracted and repelled each other in analogy to positive and negative electric charge, was discarded with advancement in the understanding of electromagnetism which showed that the magnetism of lodestones was properly explained not by magnetic monopole fluids, but rather by a combination of electric currents, the electron magnetic moment, and the magnetic moments of other particles. and this statement of non-existence of monopoles by Faraday subsequently made way into Maxwell's EM-field equations.

6.2 Intuitive Approach

Mathematically, the law is to be stated as

> THE MAGNETIC FIELD B IS A SOLENOIDAL VECTOR, THAT IS, HAS A DIVERGENCE EQUAL TO ZERO.

It must here be noted that the law is not any formal mathematical consequence but rather, only the formulation of a physical observation or truth. But we physicists must, at our level, try to arrive at this statement from the knowledge developed in Magnetostatics so far, and therefore, we will now deduce this statement from the Biot-Savart's Law that provides the magnetic field $B(x,y,z)$ due to a steady, continuous charge-flow (current), which for the general case of a volume current distribution $J(x',y',z')$ reads

$$B(r) = \frac{\mu_0}{4\pi} \int \frac{J(r') \times}{2} d\tau' \qquad (5)$$

where   is the separation vector between the primed and unprimed co-ordinates $r'(x',y',z')$ and $r(x,y,z)$ and $d\tau'$ the volume element $dx'\, dy'\, dz'$, as shown in Fig-6.
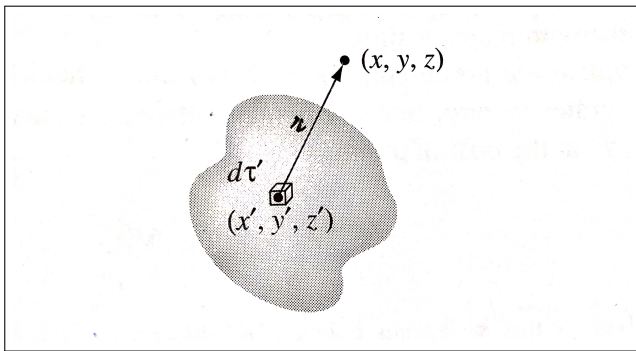

Figure 6: The Distribution, Source and field points.

The integration is over the primed co-ordinates; the divergence and curl of B are w.r.t the unprimed co-ordinates. Applying the divergence to the above equation, we ob-

tain

$$\nabla \cdot B = \frac{\mu_0}{4\pi} \int \nabla \cdot \left(J \times \underset{2}{} \right) d\tau' \qquad (6)$$

But for any two vectors P and Q, it we have the product-rule
$\nabla \cdot (P \times Q) = Q \cdot (\nabla \times P) - P \cdot (\nabla \times Q)$. Invoking this in the above equation, we have

$$\nabla \cdot \left(J \times \underset{2}{} \right) = \underset{2}{} \cdot (\nabla \times J) - J \cdot (\nabla \times) \qquad (7)$$

But $\nabla \times J = 0$ since J does not depend on the unprimed variables, while the curl of the inverse-square vector-field   is also zero. Therefore, we arrive at

$$\boxed{\nabla \cdot B = 0} \qquad (8)$$

which is the differential form of the law.

6.3 Interpretation

Field Lines and Non-existence of Monopole : The magnetic field B, like any vector field, can be depicted by field lines[9]. Gauss' Law of Magnetism is therefore equivalent to the statement that the field-lines neither have a beginning nor an end: they either form closed-loops by winding back unto itself, or they extend to infinity.

However, if magnetic monopoles were to be discovered, then, like Gauss' law of Electrostatics, the law for magnetism would

---

[9]as modelled in our discussion on the Gauss' law in Electrostatics, i.e, here, a set of curves pointing to the direction of the field at all points and whose areal-density is proportional to the magnitude of the field

have to be modified to state that $\nabla \cdot B$ would be proportional to the magnetic charge density in space.

Transversality-Requirement : For the Monochromatic plane-wave, this equation imposes on a solution of the wave-equation for E and B, the constraint of perpendicularity of B to the direction of propagation of the wave. For a plane-wave propagating along Z, the field has a form $B(z,t) = B_0 e^{i(kz-\omega t)}$, which, in order to satisfy the Gauss' Law, must have $(B_0)_Z = 0$. The same constraint is imposed on the Electric-field component by Gauss' Law in free space devoid of any charge density. That is to say, *electromag-*
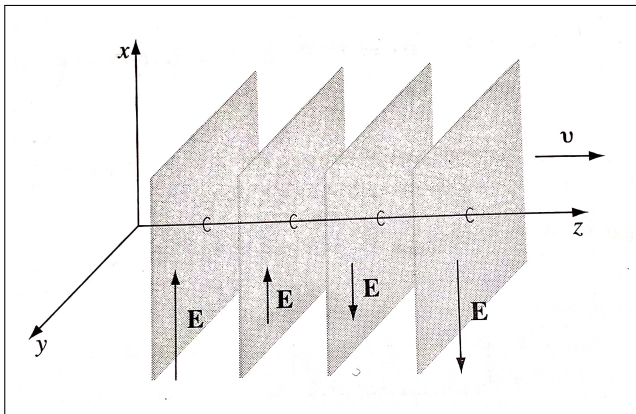
7 Faraday's Law of Electromagnetic Induction

In the year 1831, Michael Faraday came up with the revolutionary concept of electromagnetic radiation. Though subsequently in the year 1832, Joseph Henry too independently proposed this idea, Faraday got the advantage of being first.
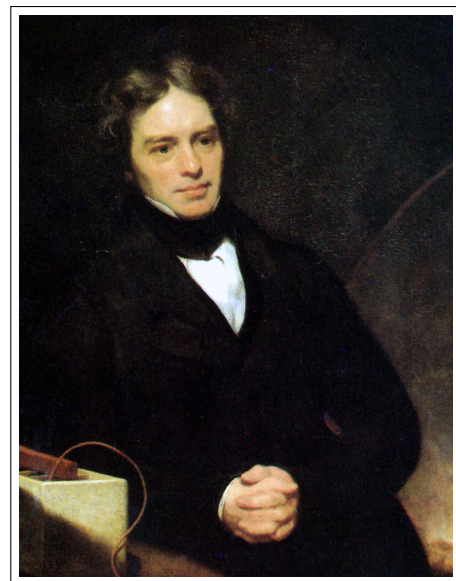

Figure 8: Michael Faraday.

Figure 7: The Plane-wave Propagation of EM oscillations.

netic waves are transverse, the electric and magnetic fields are perpendicular to the direction of propagation as shown in Fig-7.

7.1 Early Life

Faraday was born into a small Christian family and educated himself in a church Sunday School. The economic condition of his family was not very good, his father being a blacksmith, and so Faraday took up bookbinding at a very early age. Unlike other bookbinders even today, Faraday used to actually read some of the books he brought in for rebinding. There, he was

fascinated by an article on electricity in the third edition of the Encyclopdia Britannica. This sparked his interest in this subject and he performed simple experiments on electrochemistry. Faraday later was a personal assistant to Sir Humphrey Davy and he proceeded to become a great chemist. It was Faraday, in fact who discovered benzene.[4]
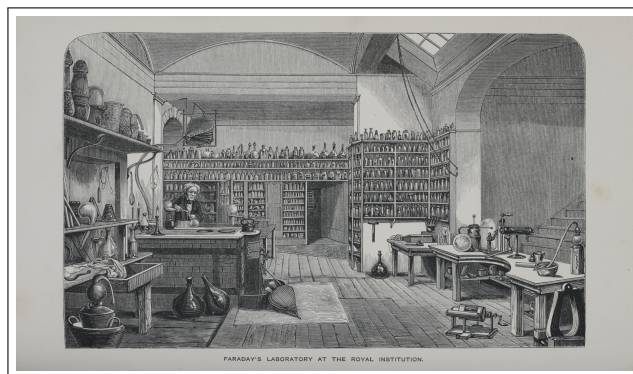

Figure 9: Faraday's Laboratory.

Even after being a chemist, Faraday was still intrigued about electricity. Discoveries made by Hans Christiaan rsted and André-Marie Ampère led Faraday to contemplate on the nature of electricity. He believed that electricity did not flow like water on a pipe, but rather as some kind of vibration or force that was somehow transmitted as the result of tensions created in the conductor. In the spring of 1831, Faraday began to work with Sir Charles Wheatstone on the theory of sound. There, it was demonstrated that a dynamic cause can create a static effect, as was seen with light powder spread on iron plates when these plates were thrown into vibration by a violin bow. Also, the phenomenon of acoustic induction was noticed by Faraday, when he saw that he could induce patterns on one plate by bowing another one nearby. Since Faraday believed current was a vibrational phenomenon, he decided to draw an analogy between sound and current, and prove that *currents can be induced*. So he devised an experiment, as explained below.

7.2 Faraday's Experimental Setup

Faraday, on August 29, 1831 took an an iron ring (called a *torus*) and then he wrapped two insulated conducting wires around the opposite sides of the torus[10]. Then, he connected one of these two wires to a galvanometer, and the other one to a battery via a switch. Indeed, there was no physical contact between these two wires for conduction to occur, since the wires were effectively insulated and situated "not-so-very-close" to each other (being on the opposite ends of the torus, the detailed setup as shown in Fig-10). Thus it was safe to assume that there would be no effect on the current flowing through the coil connected through the galvanometer, when the coil containing the battery was switched on. But Faraday obtained quite peculiar results, though not quite contrary to his expectations.

---

[10]As we will see later, the torus acts as a "highway" for the magnetic lines of force passing through it, i.e. it offers better transport of magnetic field through it.
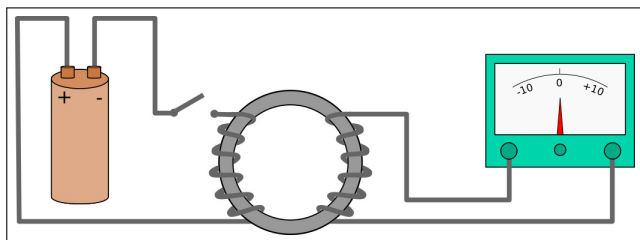
Figure 10: Faraday's Experimental Setup.

7.3 The Results

The galvanometer showed spikes (i.e. there was an *instantaneous* flow of current though the coil) when the switch in the other coil was turned on or off. Furthermore, the spikes were in opposite directions as according to whether the switch was turned on or off. Faraday was well versed in the various recently discovered properties of electromagnets, and he concluded that these were "electrical waves" travelling via the torus from one coil to the other.

7.4 Conclusions

After an extensive study, Faraday confirmed that when the switch was turned on there was a change in the *magnetic flux* from zero, and after a brief instant it attained a constant maximum value. The same phenomenon could be observed when the switch was turned off, only that the change was in the opposite direction (from the maximum attained value to zero). Therefore, Faraday argued that there must be some induced potential which was equal to the rate of change of magnetic flux, which enabled the current to flow. And thus, he came up with his law,

THE ELECTROMOTIVE FORCE AROUND A CLOSED PATH IS EQUAL TO THE NEGATIVE OF THE TIME RATE OF CHANGE OF THE MAGNETIC FLUX ENCLOSED BY THE PATH.

Mathematically, this can be stated as

$$\text{EMF} = -\frac{d\phi}{dt} \tag{9}$$

7.5 The Induced Potential

The induced potential we mentioned earlier, refers to the EMF (electromotive force). Though this term is actually a misnomer, because it is not a force, but a potential, this has been accepted as the standard convention.

The EMF is represented by the symbol ε.

The negative sigh in equation (9) can be obtained from the *Lenz's Law*, deduced in 1834 by the Russian physicist Heinrich Friedrich Emil Lenz which states that

AN INDUCED ELECTRIC CURRENT FLOWS IN A DIRECTION SUCH THAT THE CURRENT OPPOSES THE CHANGE THAT INDUCED IT.

In classical mechanics, any kind of resisting forces (for example: air drag, viscous force, friction) were all negated in general, because they hinder the motion of a body. We can indeed extend this convention here because the electromotive "force" is opposed by the change in flux.

We introduced a new term in our discussion here, *magnetic flux*. It is basically defined as:

THE MAGNETIC FLUX THROUGH A SURFACE IN AN INSTANT IS THE SURFACE

INTEGRAL OF THE NORMAL COMPONENT OF THE MAGNETIC FIELD OVER THAT SURFACE AT THAT INSTANT.

Mathematically,

$$\phi(t) = \int_S B(t) \cdot dS \qquad (10)$$

Here, $\phi(t)$ represents the magnetic flux passing through the surface S at a time t, and $B(t)$ is the magnetic field vector at that same instant.

Unfortunately, no one heeded to Faraday's work, until James Clerk Maxwell who understood the importance of Faraday's work, and in 1861-62 used it to qualitatively bring up his famous electromagnetic theory. Although he did not use the original Faraday's law, but used the one slightly modified by Oliver Heaviside and called it the Maxwell-Faraday equation.

7.6 Deriving the Maxwell-Faraday equation

Rather than presenting it directly, let us attempt to arrive at the Maxwell-Faraday equation from the preliminary concepts and equations shown above.

From electrostatics, we can get the following equation:

$$V = \int E \cdot dl \qquad (11)$$

Also, in our case, V is the total emf in the circuit, i.e.

$$V = \varepsilon_{total} = \oint_{circuit} E \cdot dl \qquad (12)$$

Using vector analysis, we can rewrite equation (12) as:

$$\varepsilon_{total} = \oint_{circuit} E \cdot dl = \int_S (\nabla \times E) \cdot dS \qquad (13)$$

This is the **Stokes' Theorem**.

Now we can say, from equations (9) and (10):

$$\varepsilon_{total} = -\frac{d\phi}{dt} = -\frac{d}{dt}\left(\int_S B(t) \cdot dS\right) \qquad (14)$$

We may interchange the integral and differential operators, since they are mutually independent here. However we switch from the $\frac{d}{dt}$ operator to the $\frac{\partial}{\partial t}$ operator. Therefore, we get

$$\varepsilon_{total} = -\int_S \frac{\partial B(t)}{\partial t} \cdot dS \qquad (15)$$

From equations (13) and (15), we get

$$\boxed{\int_S (\nabla \times E) \cdot dS = -\int_S \frac{\partial B(t)}{\partial t} \cdot dS} \qquad (16)$$

This is the Integral form of the **Maxwell-Faraday** relation. Here, in equation (16), the choice of the region of integration is completely arbitrary, and hence the integrands must themselves be equal. Therefore, we get

$$\boxed{\nabla \times E = -\frac{\partial B(t)}{\partial t}} \qquad (17)$$

This the **Maxwell-Faraday** relation that we shall conveniently be using in its differential form.

7.7 Interpretation

An electric field is capable of producing magnetic fields, (say, when a current flows

through a conductor). In a similar manner, a changing magnetic field in time can generate electric fields. This is one of the many finer examples of the fact that

NATURE LOVES SYMMETRY.

The ability of one field to generate the other is the most fundamental aspect for the existence of an electromagnetic wave. One field looks after the other as they propagate in space, a schematic representation of which is shown in Fig-11.
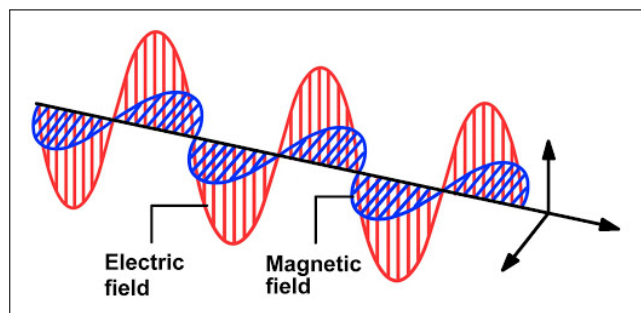

Figure 11: An electromagnetic wave.

7.8   Applications

- Transformers, electrical generators, induction motors and alternators.

- Induction cookers.

- Electric and Hybrid vehicles for magnetic induction.

8   Ampre's Law

In the year 1831, Michael Faraday came up with the revolutionary concept of electromagnetic radiation. Though subsequently in the year 1832, Joseph Henry too independently proposed this idea, Faraday got the advantage of being first.

8.1   Early Life

André-Marie Ampère was a French physicist who taught himself advanced mathematics at the young age of twelve. His early career was mostly predominated on teaching, he was a teacher of physics and chemistry at École Centrale in Bourg-en-Bresse, mathematics at École Polytechnique, astronomy and even philosophy in the University of Paris and finally he was elected to the prestigious chair in experimental physics at the Collège de France. Ampère was also invited to join the Institut Impriale, where he immersed himself in scientific inquiries and later he was elected into the Academy of Sciences.[5]

"Had Ampre died before 1820, his name and work would likely have been forgotten."

8.2   Ampre's Inspiration

In 1820, a friend (and eventual eulogist) Franois Arago demonstrated the work of the Danish physicist Hans Christiaan rsted that a magnetic needle is deflected by a current-carrying wire in close proximity before the members of the French Academy of Sciences. Ampre was so inspired by this that he completely motivated himself into the development of a relation between the

Figure 12: André-Marie Ampère.

physics and mathematics that rsted's experiments unraveled. Ampre made quite a few significant discoveries of his own, such as the mutual interation between two current-carrying wires with their cause and effect. But his most important contribution is the *Ampre's Law*, which was taken up by James Clerk Maxwell in his formulation of the electromagnetic theory.

In the language of mathematics (calculus) the law can be stated as:

THE LINE INTEGRAL OF THE MAGNETIC FIELD AROUND AN ARBITRARILY CHOSEN PATH IS PROPORTIONAL TO THE NET ELECTRIC CURRENT ENCLOSED BY THE PATH.

Hence, mathematically

$$\oint B \cdot dl = \mu_0 I_{enc} \qquad (18)$$

Now, let us derive this.

8.3   Derivation

As we have seen before, the Biot-Savart law is given as:

$$B(r) = \frac{\mu_0}{4\pi} \int \frac{J(r') \times}{2} d\tau' \qquad (19)$$

Let us for the present purpose apply the curl operator to the above expression. We get,

$$\nabla \times B = \frac{\mu_0}{4\pi} \int \nabla \times \left(J \times \frac{}{2} \right) d\tau' \qquad (20)$$

It is time for us to use a vector identity. The integrand in equation (20) is of the form $\nabla \times (M \times N)$. We use the identity

$$\nabla \times (M \times N) = (N \cdot \nabla) M - (M \cdot \nabla) N + M (\nabla \times N) - N (\nabla \times M)$$

to reduce the integrand to

$$\nabla \times \left(J \times \frac{}{2} \right) = J \left(\nabla \cdot \frac{}{2} \right) - (J \cdot \nabla) \frac{}{2} \qquad (21)$$

The terms involving the derivatives of J can be neglected because J is independent of the unprimed space variables x, y and z, as I have shown in the discussion on Gauss' Law for Magnetism.

It can be shown very easily that term within brackets in the first term reduces to

$$\nabla \cdot \left(\frac{}{2} \right) = 4\pi\delta^3() \qquad (22)$$

Here $\delta^3()$ is the Dirac-Delta function in 3D coordinates in the variable $= r - r'$, the analogue of which over the 2D coordinates has the nature as illustrated in Fig-13. We shall be using a property of this function, namely

$$\int_V f(r)\delta^3(r - r')dV = f(r')$$

when the region of integration includes r'. For evaluating the second term, we use a property of partial derivatives (which is of course, what the ∇ consists of)

$$\frac{\partial f(x - x')}{\partial x} = -\frac{\partial f(x - x')}{\partial x'}$$

This is so because R only depends on the *difference* between the coordinates[11]. We can

extend this to the ∇ operator so that the second term in equation (21) can be written as

$$- (J \cdot \nabla)_{\bar{2}} = (J \cdot \nabla')_{\bar{2}} \qquad (23)$$

Let us deal with the x-component first. We get,

$$(J \cdot \nabla') \left(\frac{x - x'}{3}\right) = \nabla' \cdot \left[\frac{(x - x')}{3}J\right] - \left(\frac{x - x'}{3}\right)(\nabla' \cdot J) \qquad (24)$$
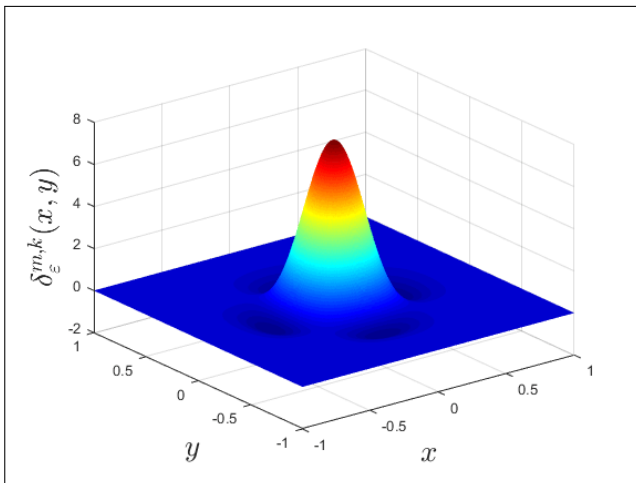
Figure 13: A three dimensional Dirac-Delta function.

using the product rule. For steady currents, the divergence of J is zero. Hence,

$$\left[-(J \cdot \nabla)_{\bar{2}}\right]_x = \nabla' \cdot \left[\frac{(x - x')}{3}J\right] \qquad (25)$$

This term makes a contribution

$$\int_V \nabla' \cdot \left[\frac{(x - x')}{3}J\right] d\tau' = \oint_S \frac{(x - x')}{3}J \cdot dS' \qquad (26)$$
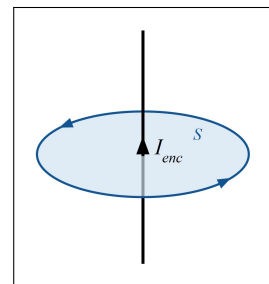
Figure 14: Here I_{enc} is the net current flowing. If there are multiple wires crossing S, then for any arbitrary choice currents have to be considered positive, and negative for the opposite direction.

Therefore, equation (20) can be integrated as

$$\nabla \times B = \frac{\mu_0}{4\pi} \int J(r') 4\pi \delta^3() d\tau' = \mu_0 J(r) \qquad (27)$$

[11]The (') is used to denote the source coordinate. Or else, we mean the field coordinate.

$$\int (\nabla \times B) \cdot dS = \oint B \cdot dl = \mu_0 \int J \cdot dS = \mu_0 I_{enc} \qquad (28)$$

the terms in reference to Fig-14. Thus, in *Integral form*,

$$\boxed{\oint B \cdot dl = \mu_0 I_{enc}} \qquad (29)$$

And, in *Differential form*,

$$\boxed{\nabla \times B = \mu_0 J} \qquad (30)$$

8.4 A Simple Thought Experiment

Maxwell was convinced that the original Ampre's law was missing something. We will try to understand the flaw in it by considering a simple thought experiment, as described below and illustrated through Fig-15:
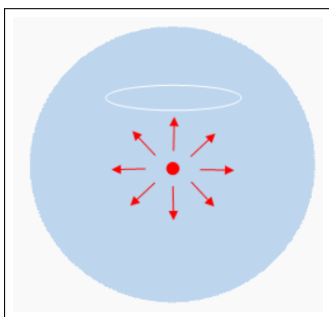

Figure 15: The spherically symmetric blob of charge in the middle of the large vat of solidified jello moves outward due of Coulomb repulsion.

Let us take a hypodermic syringe and inject a spherically symmetric blob of charge in the middle of a large vat of solidified jello (we assume it conducts electricity). From electrostatic Coulomb interaction, it is clear that they will repel each other as they move outward (and that too in a spherically symmetric fashion). Now since the charges are *moving* outward, they must generate some sort of a magnetic field. Since we are considering a spherically symmetric condition, the magnetic field must also be so. But a spherically symmetric magnetic field rises from magnetic monopoles, and we know

MAGNETIC MONOPOLES DO NOT EXIST.

So, there can be no magnetic field generated in the vat. Again, let us consider a circular path above the center of the blob. Since there are moving charges (currents) threading through the surface enclosed by the path, we must have a magnetic field, from Ampre's law. We apparently come to a paradox![10]

Now the question comes that why Ampre and other great minds of his time could not see this anomaly, even though a large number of experiments were performed by Ampere himself? Because no one at the time thought about looking into *changing* fields, all experiments were carried out in the steady state.

8.5   Maxwell's Modifications

Maxwell solved this paradox, by introducing an additional term to the previously unmodified equation. He originally proposed his modified version of the Ampere's law by considering a circuit with a parallel plate capacitor. Let us take an *Amprean loop*[12]
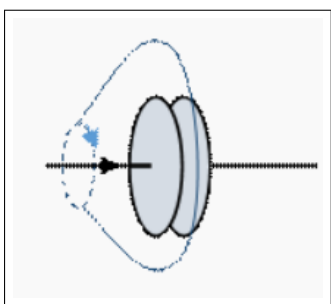

Figure 16: The Amprean loop as considered

at any point close to the capacitor, as shown in Fig-16. We can view the surface of the Amprean loop as a soap film, which can be distorted however we want. If we take it to be flat, then the same current is "punching" in and out, and Ampre's law stands valid. However, when we distort the surface such that we might include one of the plates of the capacitor, then we have the same current "punching" in, but none going out. But then, the divergence of J comes out as nonzero!

To solve this, Maxwell considered the fact that there is a changing electric field inside the capacitor, as the plate charges up. Since he had already seen that a changing magnetic field can generate an electric field,

---

[12]This is basically the closed path enclosing the wire.

he believed (on account of the universal affection towards symmetry) a changing electric field too generates a magnetic field.

8.6   The Ampre-Maxwell Relation

Maxwell thus brought up a new term, which he coined as the *displacement current*[13] which was proportional to the rate of change of the electric field. Mathematically,

$$\nabla \times B = \mu_0 \left(J + J_d \right) = \mu_0 \left(J + \epsilon_0 \frac{\partial E}{\partial t} \right)$$
$$(31)$$

This is the **Maxwell-Ampre** relation in ***differential form***, and this is valid for all situations. For steady state conditions,

$$\frac{\partial E}{\partial t} = 0$$

In the ***Integral form***,

$$\oint B \cdot dl = \mu_0 I_{enc} + \epsilon_0 \frac{d}{dt} \left(\int E \cdot dS \right) \quad (32)$$

We can get this by combining the equations

$$I = \frac{dq}{dt} \qquad (33)$$

$$\int E \cdot dS = \frac{q}{\epsilon_0} \qquad (34)$$

8.7   Interpretation

Maxwell's fourth equation, or the Amperes law in a symmetric formulation, adds an-

---

[13]Maxwell picked this term considering the movement of charges across a dielectric, which might be present in the parallel plate capacitor. Both positive and negative charges are deflected and hence *displaced*. When the electric field is time-varying, the rate of displacement changes too.

other source of magnetism – changing electric fields. Magnetic field can hence be generated in two ways:

- By an electric current, or by the movement of charges;

- By a time varying electric field, which generates a magnetic field such that it circles the electric field.

To understand how the charge and current densities ρ and J may be thought of as the sources of the electromagnetic fields, we consider in wave propagation problems, these densities to be localized in space; for example, they are restricted to flow on an antenna. The generated electric and magnetic fields are radiated away from these sources and can propagate to large distances to the receiving antennas. The qualitative mechanism by which Maxwell's equations give rise to propagating electromagnetic fields is shown in Fig-17 below.
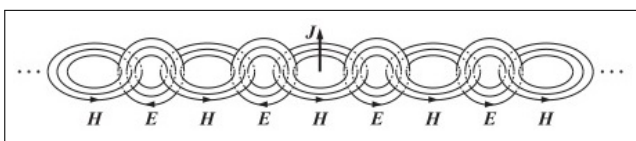

Figure 17: EM wave generation mechanism.

Here, a time-varying current J on a linear antenna generates a circulating and time-varying magnetic field of strength H, which through Faradays law generates a circulating electric field E, which through Amperes law again generates a magnetic field, and so on. The cross-linked electric and magnetic fields propagate away from the current source.

The Ampre-Maxwell relation is also another shot at charge conservation principle. Let us see how it is so.

For any volume, if the current leaving the volume be greater than the current getting into it, some charge must be consumed from within the volume. That is,

$$\int J \cdot dS = -\frac{dq}{dt} \qquad (35)$$

This is the charge conservation principle. Also, from equations (33) and (34), we get

$$\frac{d}{dt}\left(\int E \cdot dS\right) = \frac{1}{\epsilon_0}\frac{dq}{dt} \qquad (36)$$

Combining these two equations, we arrive at equation (32).

The Ampre-Maxwell relation is also very useful in calculating the magnetic field in cases where certain symmetry exists: it offers a qualitative yet easy method to do so. Indeed, doing the same from the Biot-Savart law would be much more tedious.

9   The Displacement Current

We have come across the term "Displacement current" in our discussion on the Maxwell-Ampere relation. Now, we shall be dealing with its implications and consequences in details.

$$\int_S (\nabla \times B) \cdot dS = \mu_0 \int_S J \cdot dS + \mu_0 \, \epsilon_0 \int_S \frac{\partial E}{\partial t} \cdot dS \tag{37}$$

Using Stokes' theorem, the above equa-   tion takes the form

$$\oint B \cdot dl = \mu_0 \, I_C + \mu_0 \, \epsilon_0 \int_S \frac{\partial E}{\partial t} \cdot dS \tag{38}$$

Here $I_C = \int_S J \cdot dS$ is the conduction current. This equation (38) gives the relationship between magnetic field along a loop and the rate of change of magnetic flux through an open surface whose boundary is the loop.

The second term (except μ_0) of equation (38) can be written in the following form:

$$\epsilon_0 \int_S \frac{\partial E}{\partial t} \cdot dS \;=\; \int_S \epsilon_0 \, \frac{\partial E}{\partial t} \cdot dS \;=\; \int_S J_D \cdot dS = I_D \; (say) \tag{39}$$

where $J_D = \epsilon_0 \frac{\partial E}{\partial t}$ is the displacement F. The expression is very similar to that given for conduction current. So this current density also represents some sort of current I_D, often called "displacement current". This correction was introduced by Maxwell to Ampere's law.

Now, the current carried by conductors due to the flow of charge is called "**Conduction Current**". The new term added is the current that flows due to the changing electric field and is called "**Displacement Current**" or "**Maxwell's Displacement Current**". It is to be noted that the displacement and conduction currents have the same physical effects. Here are some points to mention:

1. In cases where the electric field does not change with time, like steady electric fields in a conducting wire, the displacement current may be zero.

2. Both currents are present in different regions of the space.

3. Since a perfectly conducting or insulating medium does not exist, in most cases both the currents can be present in the same region.

4. In cases where there is no conduction current but a time-varying electric field, only displacement current is present. In such a scenario, we have a magnetic field even when there is no conduction current source nearby.

9.2 Equality of conduction and displacement currents

The conduction current I_C is given by

$$I_C = \int_S J \cdot dS = \frac{dq}{dt} \qquad (40)$$

Let dq be the charge stored in the capacitor of capacitance C in time dt. The macroscopic electric field inside the capacitor is

$$E = \frac{V}{d} \qquad (41)$$

where d is the separation between the plates of the capacitor. The capacitance of the parallel plate capacitor of cross-section S and separation d is given by

$$C = \frac{\epsilon_0 \, S}{d}. \qquad (42)$$

Now, the displacement current I_D is given by

$$I_D = \epsilon_0 \int_S \frac{\partial E}{\partial t} \cdot dS = \epsilon_0 \frac{\partial E}{\partial t} S$$

$$I_D = \epsilon_0 \frac{\partial \frac{V}{d}}{\partial t} S = \frac{\epsilon_0 \, S}{d} \frac{\partial V}{\partial t} = C \frac{\partial V}{\partial t} = \frac{\partial q}{\partial t} = I_C. \qquad (43)$$

This equation shows that the displacement current I_D is equal to the conduction current I_C.

9.3 Interesting note about displacement current

Maxwell added the displacement current term to Ampere's law in order to guarantee the charge conservation. We recast below equation (37) as

$$\boxed{\nabla \times B = \mu_0 \, J + \mu_0 \, \frac{\partial D}{\partial t}.} \qquad (44)$$

Here we introduce the quantity: *Electric Displacement vector*[14] given by

$$D = \epsilon_0 E.$$

We find that the divergence of the right-hand side of equation (44) is zero as a consequence of charge conservation. The extra term is called the displacement current.

In summary, we have shown that although the flux of the real current through a loop is not well-defined, if we form the sum of the real current and the displacement current then the flux of this new quantity through a loop is well-defined. Of course, the displacement current is not a current at all. It is, in fact, associated with the generation of magnetic fields by time-varying electric fields. Maxwell came up with this rather curious name because many of his ideas regarding electric and magnetic fields were completely wrong. For instance, Maxwell believed in the ether, and he thought that electric and magnetic fields were some sort of stresses in this medium. He also thought

---

[14]a field introduced to care of free and bound charges in the region, whose time derivative accounts for the Displacement Current.

that the displacement current was associated with displacements of the ether (hence, the name). The reason that these misconceptions did not invalidate his equations is quite simple. Maxwell based his equations on the results of experiments, and he added in his extra term so as to make these equations mathematically self-consistent. Both of these steps are valid irrespective of the existence or non-existence of the ether.

"But, hang on a minute," someone might say, "One cannot go around adding terms to laws of physics just because he/she feels like it"! The field equations are derived directly from the results of famous nineteenth century experiments. If there is a new term involving the time derivative of the electric field which needs to be added into these equations, how come there is no corresponding nineteenth century experiment which demonstrates this? We have Faraday's law which shows that changing magnetic fields generate electric fields.

Why is there no "Joe Bloggs" law[15] that says that changing electric fields generate magnetic fields? This is a perfectly reasonable question. The answer is that the new term describes an effect which is far too small to have been observed in nineteenth century experiments. Let us demonstrate this.

First, we shall show that it is comparatively easy to detect the induction of an elec-

tric field by a changing magnetic field in a desktop laboratory experiment. The Earth's magnetic field is about 1 Gauss (that is, 10^{-4} Tesla). Magnetic fields generated by electromagnets (which will fit on a laboratory desktop) are typically about one hundred times bigger that this. Let us, therefore, consider a hypothetical experiment in which a 100 Gauss magnetic field is switched on suddenly. Suppose that the field ramps up in one tenth of a second. We ask the question - What electromotive force is generated in a 10 centimeter square loop of wire located in this field?

Faraday's law is written

$$\epsilon = -\frac{\partial}{\partial t} \oint B \cdot dS \sim \frac{BA}{t} \qquad (45)$$

where $B = 0.01$ Tesla is the field-strength, $A = 0.01$ m^2 is the area of the loop, and $t = 0.1$ seconds is the ramp time. It follows that $\epsilon \sim 1$ millivolt. Well, one millivolt is easily detectable. In fact, most hand-held laboratory voltmeters are calibrated in millivolts. It is clear that we would have no difficulty whatsoever detecting the magnetic induction of electric fields in a nineteenth century style laboratory experiment.

Let us now consider the electric induction of magnetic fields. Suppose that our electric field is generated by a parallel plate capacitor of spacing one centimeter which is charged up to 100 volts. This gives a field of 10^4 volts per meter. Suppose, further, that the capacitor is discharged in one tenth of a second. The law of electric induction is obtained by integrating equation (44), and ne-

---

[15]"Joe Bloggs" is a placeholder name to represent the average man on the street. It is used to represent the random layman.

glecting the first term on the right-hand side. Thus,

$$\oint_{2\pi r} B \cdot dl = \epsilon_0 \mu_0 \frac{\partial}{\partial t} \int_{\pi r^2} E \cdot dS \qquad (46)$$

Let us consider a loop of 10 centimeters square, as shown in Fig-18. What is the magnetic field generated around this loop (we could try to measure this with a Hall probe). Very approximately, we find that
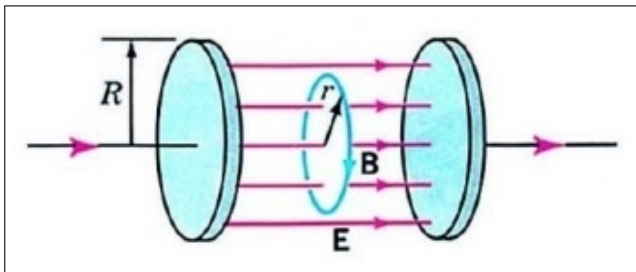

Figure 18: The parallel plate capacitor of plate-area $A = \pi R^2$, consider a loop of radius r = 1.78 cm spanning an area $\sim$10 cm sq.

$$rB \sim \epsilon_0 \mu_0 \frac{Er^2}{t} \qquad (47)$$

where $r = 0.02$ meters is the dimensions of the loop, B is the magnetic field-strength, $E = 10^4$ volts per meter is the electric field, and $t = 0.1$ seconds is the decay time of the field. We find that $B \sim 10^{-9}$ Gauss. Modern technology is unable to detect such a small magnetic field, so we cannot really blame Faraday for not noticing electric induction in 1830.

"So," one might say, "why did we bother mentioning this displacement current thing in the first place if it is undetectable?"

YES, THIS IS A PERFECTLY FAIR QUESTION.

The answer is that the displacement current is detectable in some experiments. Suppose that we take an FM radio signal, amplify it so that its peak voltage is one hundred volts, and then apply it to the parallel plate capacitor in the previous hypothetical experiment. What size of magnetic field would this generate? Well, a typical FM signal oscillates at 10^9 Hz, so time t in the previous example changes from 0.1 seconds to 10^{-9} seconds. Thus, the induced magnetic field is about 10^{-1} Gauss. This is certainly detectable by modern technology. So, it would seem that if the electric field is oscillating fast then electric induction of magnetic fields is an observable effect. In fact, there is a virtually infallible rule for deciding whether or not the displacement current can be neglected in equation (44). If electromagnetic radiation is important then the displacement current must be included. On the other hand, if electromagnetic radiation is unimportant then the displacement current can be safely neglected. Clearly, Maxwell's inclusion of the displacement current in equation (44) was a vital step in his later realization that his equations allowed propagating wave-like solutions. These solutions are, of course, electromagnetic waves.

9.4 Conservation of Charge

Maxwell added the displacement current term to Ampere's law in order to guaran-

$$\nabla \times B = \mu_0 \, J + \mu_0 \, \frac{\partial D}{\partial t}$$

Taking divergence of equation (44), we get

$$\nabla \cdot (\nabla \times B) = \mu_0 \, \nabla \cdot J + \mu_0 \, \nabla \cdot \frac{\partial D}{\partial t}$$

$$= \mu_0 \, \nabla \cdot J + \mu_0 \, \frac{\partial (\nabla . D)}{\partial t} = \mu_0 \, \nabla \cdot J + \mu_0 \, \frac{\partial \rho}{\partial t} \quad (48)$$

Here, we have used the relation

$$\nabla \cdot E = \frac{\rho}{\epsilon_0} \qquad \longrightarrow \qquad \nabla \cdot (\epsilon_0 \, E) = \rho \qquad \longrightarrow \qquad \nabla \cdot D = \rho$$

Value of left hand side of equation (48) is identically zero. Equating the right hand side of this equation to zero, we get

$$\mu_0 \, \nabla \cdot J + \mu_0 \, \frac{\partial \rho}{\partial t} = 0 \quad \Longrightarrow \quad \nabla \cdot J + \frac{\partial \rho}{\partial t} = 0 \tag{49}$$

This is the differential form of charge conservation law. This is the equation of continuity for charge conservation.

9.5 Charges are safe on the surface of a conductor

Integrating both sides of the charge conservation equation over a closed volume V surrounded by the surface S, as shown in Fig-19 and using the divergence theorem, we obtain the integrated form of equation (49):

$$\int_V \nabla \cdot J \, dV + \int_V \frac{\partial \rho}{\partial t} \, dV = 0 \implies \oint_S J \cdot dS = -\frac{d}{dt} \int_V \rho \, dV \tag{50}$$

The left-hand side of equation (50) represents the total amount of charge flowing outwards through the surface S per unit time. The right-hand side of equation (50) represents the amount by which the charge is decreasing inside the volume V per unit time. In other words, charge does not disappear into (or created out of) nothingness—it decreases in a region of space only because it flows into other regions. This is the law of conservation of charge.
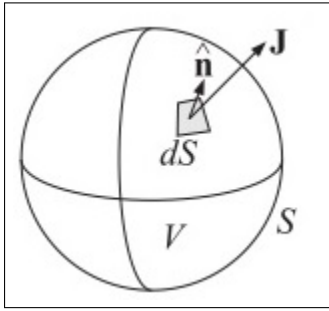
Another consequence of equation (50)

Figure 19: Flux outwards through the surface.

is that in good conductors, there cannot be any accumulated volume charge. Any such charge will quickly move to the surface of the conductor and distribute itself in such a way to make the surface an equipotential one.

9.5.1 An Interesting Note!

Let us start from equation for conservation of charge which is recast as

$$\nabla \cdot J + \frac{\partial \rho}{\partial t} = 0 \qquad (51)$$

Now, inside the conductor, we have $D = \epsilon E$ and $J = \sigma E$ where σ is the conductivity of the conductor[16]. Using these re-

---

[16]The current density J can be simply expressed in terms of the magnitude of conventional current I as it's value per unit surface area. Over an area A of a conductor, we have

$$J = \frac{I}{A} = \frac{V}{RA} = \frac{V}{RA}$$

from Ohm's Law. Writing $V = El$ across the length of flow l, and expressing R in terms of resistivity ρ, or conductivity $\sigma = 1/\rho$, we can easily show the vectorial relation

$$J = \sigma E$$

lations and equation (51), we obtain

$$\nabla \cdot J = \sigma \, \nabla \cdot E = \frac{\sigma}{\epsilon} \, \nabla \cdot D = \frac{\sigma}{\epsilon} \, \rho$$

and

$$\boxed{\frac{\partial \rho}{\partial t} + \frac{\sigma}{\epsilon} \, \rho = 0} \qquad (52)$$

This equation (52) has the solution of the form

$$\rho(r, t) = \rho_0(r) \, e^{-\frac{\sigma t}{\epsilon}} \qquad (53)$$

where $\rho_0(r)$ is the initial volume charge distribution. The solution shows that the volume charge disappears from inside and therefore it must accumulate on the surface of the conductor, and how interestingly, we arrive at the result of what the Gauss Law of Electrostatics had led us to in its implications!

The exponential factor of equation (53) defines a 'time constant' T_{rel}. This is the relaxation time of charge inside the conductor and is given by

$$T_{rel} = \frac{\epsilon}{\sigma}. \qquad (54)$$

This relaxation time T_{rel} has very small value for good conductors. For example, in copper,

$$T_{rel} = \frac{\epsilon}{\sigma} = \frac{8.85 \times 10^{-12}}{5.7 \times 10^7} = 1.6 \times 10^{-19} \text{ sec.}$$

By contrast, T_{rel} is of the order of days in a good dielectric. For good conductors, the above argument is not quite correct because it is based on the steady version of Ohm's

law, $J = \sigma E$, which must be modified to take into account the transient dynamics of the conduction charge.

It turns out that the relaxation time T_{rel} is of the order of the collision time, which is typically 10^{-14} sec.

10   Concluding Discussions

The reader now knows how Maxwell brought together all the work that had been done by brilliant physicists such as Oersted, Coulomb, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. We have, through our discussion, paraphrased in this text the Maxwells equations, rather than just presenting numerically, to stress upon how intuitive learning in Physics is of much necessity to the core and crux of this field. Here, the equations have illustrated how apparently simple mathematical statements can elegantly unite and express a multitude of concepts why mathematics is the language of science.

We have seen that the Maxwells equations span the major laws of electricity and magnetism. However, what might not be so apparent is the symmetry that Maxwell introduced in his mathematical framework. Especially important is his addition of the hypothesis that changing electric fields create magnetic fields. This is exactly analogous (and symmetric) to Faradays law of induction and had been suspected for some time, but fits beautifully into Maxwells equations.

Symmetry is apparent in nature in a wide range of situations. As famously professed by the German theoretical physicist H. Weyl —

THE MATHEMATICAL LAWS GOVERNING NATURE ARE THE SOURCE OF SYMMETRY IN NATURE.

In contemporary research, symmetry plays a major part in the search for subatomic particles using massive multinational particle accelerators such as the new Large Hadron Collider at CERN, about which the reader can pursue his exploration for further interests. Herein, to comment rather vaguely on the driving propositions behind such experimental pursuit: Maxwells complete and symmetric theory showed that electric and magnetic forces are not separate, but different manifestations of the same thing – the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four fundamental forces in nature – Gravitational, Electrical (or the EM together), and the Strong and Weak nuclear forces.

10.1   The Prediction and Pursuit

The consequences of Maxwell's Equations for Electromagnetic Waves remained a topic for rigorous pursuit after Maxwell. Since changing electric fields create relatively weak magnetic fields, they could not be easily detected at the time of Maxwells hypothesis. However, Maxwell realized that oscil-

lating charges, like those in AC circuits, produce changing electric fields. He predicted that these changing fields would propagate from the source like waves generated on a lake by a stone dropped into the water.

The waves predicted by Maxwell would consist of, as discussed by us in The Interpretation of the Maxwell-Faraday Relation, oscillating electric and magnetic fields (the EM wave). Electromagnetic waves would be capable of exerting forces on charges great distances from their source, and they might thus be detectable. Maxwell calculated that electromagnetic waves would propagate at a speed given by the equation —

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

— the verification [9] of which by Maxwell's Equations as applied to Equations of Wave Theory is left as an exercise to the reader.

Plugging in the values for μ_0 and ϵ_0 into the equation for c , we find that

$$c = \frac{1}{\sqrt{\left(8.85 \times 10^{-12} \frac{C^2}{N \cdot m^2}\right)\left(4\pi \times 10^{-7} \frac{T \cdot m}{A}\right)}} = 3 \times 10^8 \text{ m/s}$$

which is the speed of light. Maxwell therein concluded that light is an electromagnetic wave having *such wavelengths that it can be detected by the eye*.

However, the Physicists' intuition must claim that other wavelengths should exist – it remained to be seen if they did. If so, Maxwells theory and remarkable predictions would be verified, the greatest victory of physics since Newton. Experimental verification indeed followed within a few years, but not before Maxwells death.

10.1.1 Finally : The EM-Wave Detection

The German physicist H.Hertz (18571894) was the first to generate and detect EM waves of arbitrary wavelength in the laboratory[11]. He performed a series of experiments that not only confirmed the existence of electromagnetic waves, but also verified that they travel at the speed of light. Hertz used an AC LCR (inductor-capacitor-resistor) circuit, that shows dissipated power resonance at a frequency $f_0 = \frac{1}{2\pi\sqrt{LC}}$, and connected it across a loop of wire in manner as shown in Fig-20. High voltages induced across the gap in the loop produced visible sparks that served as evidence of the current in the circuit and that helped generate the electromagnetic waves. Across the laboratory, another loop was attached to another LCR set-up, which could be tuned (just like the dial on a radio) to the same resonant frequency as the first and thus, was made to receive the waves. This loop also had a similar gap across which sparks

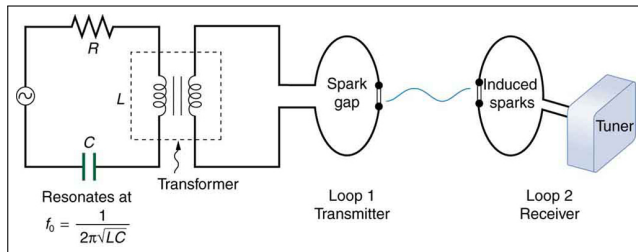were generated, giving solid evidence that EM waves had been received.

Figure 20: Hertz's Experimental Setup.

Hertz also studied the reflection, refraction, and superposition patterns of the electromagnetic waves he generated and thus verified their wave character. He determined the wavelength λ from the interference patterns, and knowing their frequency f from the AC circuit, he was able to calculate their speed of propagation v using the equation $v = f\lambda$. Hertz was thereby able to prove that EM waves travel at the speed of light, laying down the truth in Maxwell's hypothesis once and for all.

While drawing down this discussion, its worth remembering Boltzmann's remark

I COULD SAY THAT MAXWELL'S SUCCESSORS CHANGED NOTHING IN HIS EQUATIONS BUT THE LETTERS. BUT... ONE MUST NEVERTHELESS MARVEL NOT AT THE FACT THAT SOMETHING COULD BE ADDED TO THESE EQUATIONS, BUT MORE AT THE FACT THAT SO LITTLE HAS BEEN ADDED TO THEM.

It would be good to know at this concluding point that at Maxwell's 100th birth-anniversary, Niels Bohr had indeed pointed out "... it was even proposed that new terms be added to the celebrated equation of Maxwell. However;
MAXWELL'S THEORY TURNED OUT TO BE TOO CONSISTENT AND TOO ELEGANT TO ADMIT SUCH MODIFICATION !!"

References

[1] Potential theory: the origin and applications, Reading: Irina Markina, Department of Mathematics, University of Bergen, Johannes Brunsgate 12, Bergen 5008, Norway.

[2] Gray, J. John. "Carl Friedrich Gauss." Encyclopedia Britannica, https://www.britannica.com /biography/Carl-Friedrich-Gauss.

[3] Domb, Cyril. "James Clerk Maxwell". Encyclopedia Britannica, https://www.britannica.com/biography /James-Clerk-Maxwell.

[4] Williams, L. Pearce. "Michael Faraday". Encyclopedia Britannica, https://www.britannica.com/biography /Michael-Faraday.

[5] Shank, J.B.. "Andr-Marie Ampre". Encyclopedia Britannica, https://www.britannica.com/biography /Andre-Marie-Ampere.

[6] Griffiths, David J, Introduction to electrodynamics; 4th ed., Pearson,

Boston, MA, 2013, Re-published by Cambridge University Press in 2017, https://cds.cern.ch/record/1492149.

[7] Britannica, The Editors of Encyclopaedia. "Maxwell's equations". Encyclopedia Britannica, https://www.britannica.com/science/Maxwells-equations.

[8] https://www.maxwells-equations.com/

This site aims to teach Maxwell's Equations in as simple and intuitive a manner as possible.

[9] Foundations of Electricity and Magnetism, Basudev Ghosh, Books & Allied Ltd. 1 January 2008.

[10] Feynman, Richard Phillips, Robert B. Leighton, and Matthew Linzee Sands. 1964. Feynman lectures on physics. Vol. 2, Vol. 2. Reading: Addison-Wesley.

[11] https://courses.lumenlearning.com/physics/chapter/24-1-maxwells-equations-electromagnetic-waves-predicted-and-observed/

An intuitive and simple summary of Maxwell's equations, Lumen Physics.

Towards understanding the properties of Quark Gluon Plasma using a MIT bag model

Kausik Pal[1]

[1]Department of Physics, Serampore College, Serampore-712201, India.
kausik.sinp@gmail.com

Abstract

We present the thermodynamical equilibrium properties with emphasis on the equation of state (EOS) of spin asymmetric quark gluon plasma (QGP) using a MIT bag model within one gluon exchange (OGE) interaction. We have evaluated the QGP partition function, total energy, pressure, entropy, Helmholtz free energy, specific heat and incompressibility dealing with the basic idea of statistical mechanics. The analytical expressions which we have presented are valid for light quarks only. Here we have taken massless limit of quarks which are weakly interacting and massless gluons are almost free. Moreover, we have studied the quark hadron phase transition and reveal the QCD phase diagram in the $T - \mu$ plane.

1 Introduction

Exploring and understanding the properties, especially the equation of state (EOS), of extremely high temperature and high density matter formed in the process of high energy heavy ion collision where quarks and gluons define the relevant degrees of freedom is the fundamental issues in contemporary nuclear and high energy physics. This new phase of matter, the Quark Gluon Plasma (QGP), is governed by the principles of Quantum Chromodynamics (QCD) and is the result of a phase transition from the normal nuclear matter, has a great significance to understanding the origin of the universe, intersteller evolution and the nuclear structure, reactions etc [1].

QCD is an asymptotically free theory i.e., the interactions between the quarks and gluons become weaker as the mutual distance decreases or as the exchanged momentum increases. Similarly, the QCD coupling constant also becomes weak at high temperature and/or density allowing the formation of QCD matter by which we mean phases of matter whose properties are governed not by the hadrons (mesons and baryons) but by more fundamental constituents of matter *viz.* the quarks and gluons. Such a phase where color is deconfined

and interaction becomes screened i.e., any length scale of interest is greater than the screening length of the interaction, is known as QGP [1]. The most experimentally accessible way to characterize the QCD phase diagram is in the plane of temperature (T) and the baryon chemical potential (μ), as shown in fig.(1) where the two extreme regions can be identified easily close to the T and μ axis. The early universe evolved close to the temperature axis in the phase diagram of strongly interacting matter. Matter in the core of compact stellar objects, like neutron stars, is close to the quark chemical potential axis, at values μ around $400 - 500$ MeV [1].
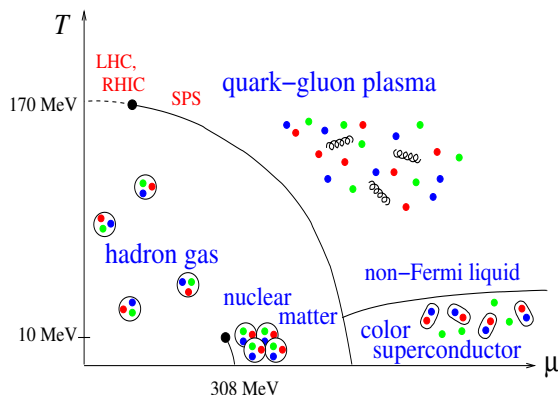

Figure 1: Schematic QCD phase diagram with boundaries that defines various states of QCD matter.

Lattice QCD calculations have established the existence of QGP phase at temperature larger than ~ 150 MeV for zero net-baryon density [2]. Depending on the number of quark flavors and the masses of the quarks, the transition between ordinary hadronic matter to the QGP could be a ther-modynamic phase transition of first order, of second order or simply a crossover transition [3, 4]. In addition to the confined-deconfined transition, another phase transition is postulated, known as chiral phase transition which might take place at high temperature ($T \sim \Lambda_{QCD} \sim 200$ MeV) where the broken chiral symmetry can be restored[2]. However, the exact locations of the phase boundary and predictions on the critical point or the order of transition differ among the various lattice groups, depend on model assumptions [3, 4]. It is to be mentioned that the deconfinement phase transition is expected to be first order if the quark mass is infinitely heavy [3], while the chiral phase transition could be of second order with two massless flavors [4]. We in the present model particularly interested in the region where QCD matter corresponds to low temperature and high chemical potential which permits us to use Fermi Liquid Theory (FLT)[5] to compute the thermodynamic quantities perturbatively [6, 7].

In order to create the QGP in the laboratory via collisions of heavy nuclei at ultra-relativistic energies, to study experimentally the QCD phase structure under controlled conditions, a scan over beam energies is employed. By creating systems with a broad range of initial conditions, we access the properties of the matter by studying the evolution of these systems as they expand, cool, and possibly undergo phase transitions. The different reaction trajectories cross the phase boundary at different T and μ values and

that allows us to access interesting features in the phase diagram. Interpolation between the extreme conditions of this phase diagram provides the early universe conditions towards the T axis and compact steller objects on the μ axis. Experiments at the RHIC and the LHC have provided compelling evidence of the formation of a deconfined state of quarks and gluons (QGP) matter close to $\mu \simeq 0$. These experimental search of the QGP needs plausible theoretical estimates of various signals which depend on energy, pressure, entropy, transition temperature and the equation of state (EOS), etc. In this manuscript tries to give an overview of the thermodynamical quantities concerned with the equilibrium properties of the QGP using a MIT bag model [6, 7]. We assume quarks and gluons interact weakly, as they are asymptotically free. Using quasiparticle model, the thermodynamical properties of the QGP might be computable perturbatively from the knowledge of QCD partition function. It should be mentioned, that the terms in the perturbative series, which are analytically computable, are rapidly decreasing in magnitude at high temperatures where the coupling constant $\alpha_c = \frac{g^2}{4\pi} < 1$, is weak. Therefore, only the first few terms of the perturbative series are actually necessary to obtain physically reasonable values for the QCD partition function. Phase transition between quarks and hadrons are associated with the evolution of thermodynamic quantities like QCD partition function ($\mathcal{Z}$), energy density (E), pressure ($\mathcal{P}$),

entropy density ($\mathcal{S}$), Helmholtz free energy ($\mathcal{E}$) as well as a set of response functions, like specific heat (C_v), incompressibility (K), etc [6, 7]. In the present work, we undertake to present all those characteristics of weakly interacting spin asymmetric QGP matter at nonzero quark chemical potential in terms of temperature and spin polarization parameter $\xi = (n_q^+ - n_q^-)//(n_q^+ + n_q^-)$, where n_q^+ and n_q^- correspond to the densities of spin-up and spin-down quarks, respectively. The spin polarization parameter ξ has the condition $0 \leq \xi \leq 1$. When $\xi = 1$, system becomes completely spin polarized, whereas, $\xi = 0$ indicates the system becomes unpolarized. Moreover, we estimate the deconfinement temperature T_c for quark-hadron phase transition and examine the behavior of the equation of state.

2 MIT bag model and thermodynamics of QGP

A phenomenological description of quarks that are confined in color singlet hadrons and never seen in isolation is provided by the bag model. In the MIT bag model quarks are treated as massless particles inside a bag of finite dimension and infinitely massive outside the bag. A rigid boundary condition is implemented confining the quarks, also making the restriction that only those bound quarks should exist inside the bag cavity and that different bags should not overlap. Confinement in the model is the result of the balance of the bag pressure $\mathcal{B}$,

which is directed inward, and the stress arising from the kinetic energy of the quarks. Here, the bag pressure $\mathcal{B}$ is introduced to take into account the difference between the perturbative and the non-perturbative vacuum. In short, the bag constant is the difference between the energy densities of non-interacting and interacting quarks. If the quarks are confined in the bag, the gluons should also be confined in the bag. The total color charge of the matter inside the bag must be colorless.

Within our model, we consider the Lorentz invariant scattering amplitude of two massless quarks (u and d quarks) around the Fermi surface by the one gluon exchange interaction. The direct or Hartree term vanishes because the trace of single color matrices becomes zero. Thus the leading contribution to the energy density next

to the kinetic energy is the exchange (Fock) term. We are dealing with quasiparticles whose spins are all eigenstates of the spin along a given direction, viz. z. The quasiparticle interaction $(f_{pp'}^{ss'})$ can be decomposed into two parts, spin may be either parallel $(s = s')$ or antiparallel $(s = -s')$ corresponding to spin nonflip $(f_{pp'}^{\mathrm{nf}})$ or spin flip $(f_{pp'}^{\mathrm{f}})$ scattering amplitudes, such that[7, 8]

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

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, i.e. $f_{pp'}^{+-} = f_{pp'}^{-+}$, where superscripts denote the spin indices. Since spin and momentum have no preferred direction, we take the angular averaged interaction parameter which is given by [7, 8]

$$f_{pp'}^{ss'} = \frac{g^2}{9pp'} \int \frac{\mathrm{d}\Omega_1}{4\pi} \int \frac{\mathrm{d}\Omega_2}{4\pi} \left[1 + (\hat{p} \cdot \hat{s})(\hat{p}' \cdot \hat{s}') \right], \qquad (2)$$

where g is the coupling constant.

To understand the heavy ion processes, we restrict ourselves to the symmetric nuclear matter i.e., the initial state with equal neutron and proton densities i.e., $n_{\mathrm{n}} = n_{\mathrm{p}}$ so that $n_{\mathrm{u}} = n_{\mathrm{d}}$, which means the contributions

of u and d quarks are equal. For simplicity, we consider massless quarks, for which all the relevant quantities will be function of temperature (T) and Fermi momentum (p_f). Since the baryon number of a quark is $\frac{1}{3}$, we have baryon density $n_b = \frac{1}{3}n_q$ and the quark density n_q is given by

$$n_q = \sum_{s=\pm} n_q^s = N_c N_f \sum_{s=\pm} \int_0^\infty \frac{d^3p}{(2\pi)^3} n_p^s(T) = \frac{1}{\pi^2} \sum_{s=\pm} \left[p_f^{s\,3} + \pi^2 T^2 p_f^s \right]. \tag{3}$$

Here, $N_c = 3$ and $N_f = 2$ are the color and flavor degeneracy factors, $n_p^s(T)$ is the Fermi distribution function and p_f^s is the spin dependent Fermi momentum which can be represented as a function of n_b and T.

In order to understand the phase structure of strongly interacting matter, one has to know its equation of state. The equation of state is determined by the partition function, the resulting expression for Fermionic partition function for each quarks is written as

$$\frac{\ln \mathcal{Z}_q}{V} = N_c \sum_{s=\pm} \int_0^\infty \frac{d^3p}{(2\pi)^3} \left[\ln \left(1 + e^{-\beta(p-\mu)} \right) + \ln \left(1 + e^{-\beta(p+\mu)} \right) \right] \tag{4}$$

In the chiral limit i.e., $m_q \to 0$, the ana-lytical result of the above integral becomes

$$\frac{\ln \mathcal{Z}_q}{V} = \frac{1}{2\pi^2} \left[\frac{7\pi^4 T^3}{30} + \sum_{s=\pm} \left(\frac{p_f^{s\,4}}{4T} + \frac{\pi^2 p_f^{s\,2} T}{2} \right) \right] \tag{5}$$

where we have used $\beta = \frac{1}{T}$ in the Boltz-mann factor unit. Similarly, for massless gluons the partition function is given by

$$\frac{\ln \mathcal{Z}_g}{V} = 16 \int_0^\infty \frac{d^3k}{(2\pi)^3} \left[\ln \left(1 - e^{-\beta k} \right) \right] = \frac{8}{45} \pi^2 T^3 \tag{6}$$

where 16 is the degeneracy of gluons. Using above eq.(5), alternatively we obtain the same result of the quark number density as given by eq.(3), by evaluating

$$n_q = N_f \sum_{s=\pm} T \left[\frac{\partial}{\partial p_f^s} \left(\frac{\ln \mathcal{Z}_q}{V} \right) \right]_T \quad (7)$$

The QGP equation of state of free massless quarks and gluons can be derived from the above equations. But, the non-trivial feature of the bag model EOS is that the vacuum pressure generates positive and has a contribution to the bag energy density. Hence,

the partition function for the QGP state consisting of perturbatively interacting quarks and gluons including non-perturbative vacuum contribution in terms of bag constant $\mathcal{B}$. Thus the total partition function of QGP state can be written as

$$\ln \mathcal{Z}_{QGP} = \ln \mathcal{Z}_q + \ln \mathcal{Z}_g + \ln \mathcal{Z}_{vac} \quad (8)$$

where the vacuum partition function $\frac{\ln \mathcal{Z}_{vac}}{V} = -\frac{\mathcal{B}}{T}$. With these considerations, resulting expressions for the energy density and pressure of QGP phase are given by

$$E_{QGP} = E_q + E_g + \mathcal{B}, \quad \mathcal{P}_{QGP} = \mathcal{P}_q + \mathcal{P}_g - \mathcal{B}. \quad (9)$$

Now, the leading contributions to the energy density of quarks is given by

$$E_q = E_{kin} + E_{ex}. \quad (10)$$

The total kinetic energy density for massless spin-up and spin-down quarks, including the color and flavor degeneracy factors for quarks, is

$$
\begin{aligned}
E_{kin} &= N_c N_f \sum_{s=\pm} \int \frac{d^3 p}{(2\pi)^3} p^s n_p^s(T) \\
&= \frac{3}{(2\pi)^2} \left\{ p_f^4 [(1+\xi)^{4/3} + (1-\xi)^{4/3}] + 2\pi^2 T^2 p_f^2 [(1+\xi)^{2/3} + (1-\xi)^{2/3}] + \frac{14}{15} \pi^4 T^4 \right\}
\end{aligned}
$$
$$(11)$$

For spin asymmetric quarks, the exchange energy density consists of two terms

$E_{ex} = E_{ex}^{\text{nf}} + E_{ex}^{\text{f}}$, and can be determined by   evaluating the following integrals

$$E_{ex}^{\text{nf}} = \frac{N_f N_c^2}{2} \sum_{s=\pm} \int \int \frac{d^3 p}{(2\pi)^3} \frac{d^3 p'}{(2\pi)^3} f_{pp'}^{\text{nf}} \, n_p^s(T) \, n_{p'}^s(T), \tag{12}$$

$$E_{ex}^{\text{f}} = N_f N_c^2 \int \int \frac{d^3 p}{(2\pi)^3} \frac{d^3 p'}{(2\pi)^3} f_{pp'}^{\text{f}} \, n_p^s(T) \, n_{p'}^s(T). \tag{13}$$

The analytical expression for the total   exchange energy density is found to be

$$E_{ex} = \frac{g^2}{(2\pi)^4} \Big\{ p_f^4 [(1+\xi)^{4/3} + (1-\xi)^{4/3} + 2(1-\xi^2)^{2/3}]$$
$$+ \frac{4}{3}\pi^2 T^2 p_f^2 [(1+\xi)^{2/3} + (1-\xi)^{2/3}] + \frac{4}{9}\pi^4 T^4 \Big\}. \tag{14}$$

The resulting gluon contribution to the energy density is

$$E_g = 16 \int \frac{d^3 k}{(2\pi)^3} \frac{k}{e^{k/T} - 1} = \frac{8}{15}\pi^2 T^4, \tag{15}$$

Our system is ultra-relativistic, the QGP equation of state is determined by the relation between the pressure and the energy density:

$$\mathcal{P}_{QGP} = \frac{1}{3} E_{QGP}. \tag{16}$$

Once the value of $\mathcal{P}_{QGP}$ is known, the entropy density of the system is determined by equating

$$\mathcal{S}_{QGP} = \frac{\partial}{\partial T}(\mathcal{P}_{QGP}). \tag{17}$$

The Helmholtz free energy of the system is given by

$$\mathcal{E}_{QGP} = E_{QGP} - T\mathcal{S}_{QGP}. \tag{18}$$

For the numerical estimation of all these quantities, we take $\alpha_c = \frac{g^2}{4\pi} = 0.2$, as the coupling constant of QCD and the bag pressure $\mathcal{B} = 208$ MeV fm^{-3} for zero hadronic pressure [7].

Fig.(2) shows the variation of Helmholtz free energy with baryon density (upper panel) and with temperature (lower panel) for different polarization parameter ξ. We see, free energy is larger with higher value of ξ and decreases (increases) when temperature (baryon density) is increased.
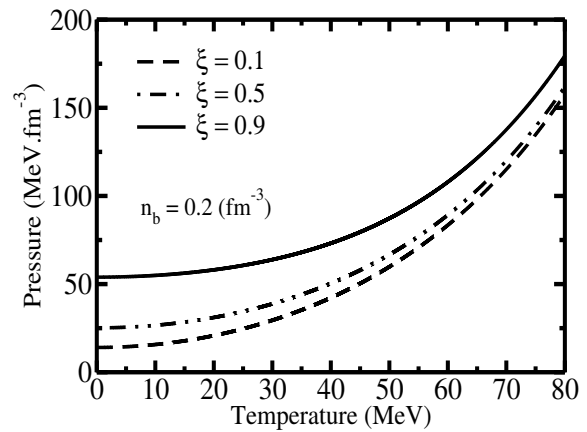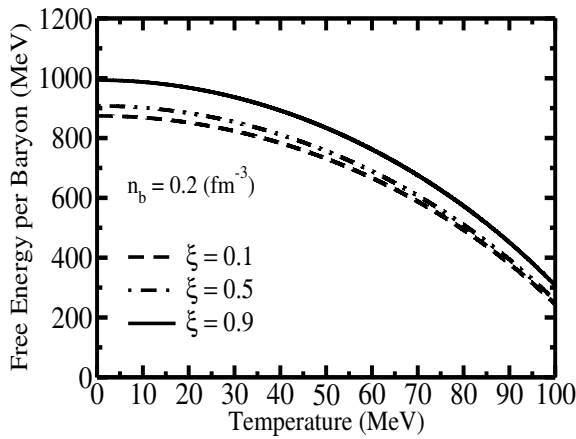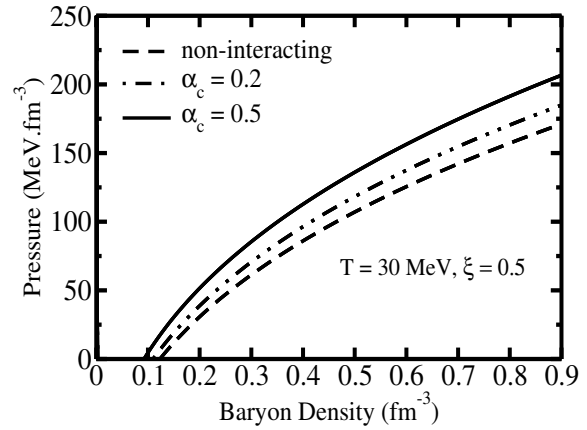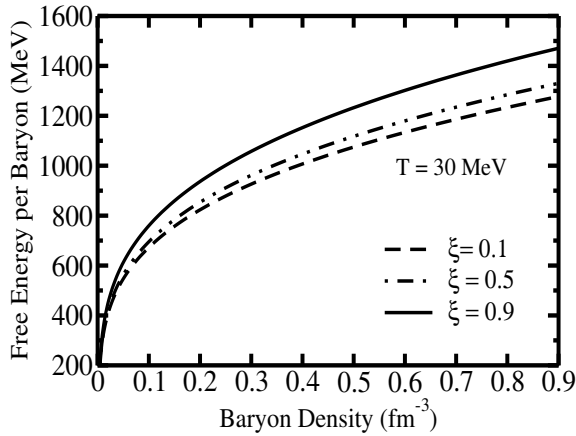
Figure 2: Helmholtz free energy per baryon.

Figure 3: Variation of pressure.

This indicates QGP is more stable when quarks are unpolarized.

Equation of state are useful to describe the properties of the matter, it relates the state variables which delineates the state of matter such as pressure, temperature etc. The pressure is illustrated as a function of density with different coupling constant in the upper panel of fig.(3). It has been observed that the pressure increases with baryon density and also raises as interaction strength between quarks is large.

Thus the interaction makes easier the quarks transition to the deconfined phase at lower density and an increase of the interaction strength pushes the pressure of quarks to be similar to the bag pressure at smaller baryon density. Therefore, the interaction of the QCD coupling constant reduces the value of the density to reach a transition. In the lower panel of fig. 3, the variation of pressure has been shown as a function of temperature with different polarization parameter ξ. We see, greater pressure produced when tem-

perature increases and also the pressure of QGP is larger for polarized quarks than for unpolarized ones. This observation indicates that the interaction as well as the order parameter (ξ) makes the EOS stiffer, as an EOS is generally said to be stiffer than another EOS if greater pressure is produced for a given density or temperature.
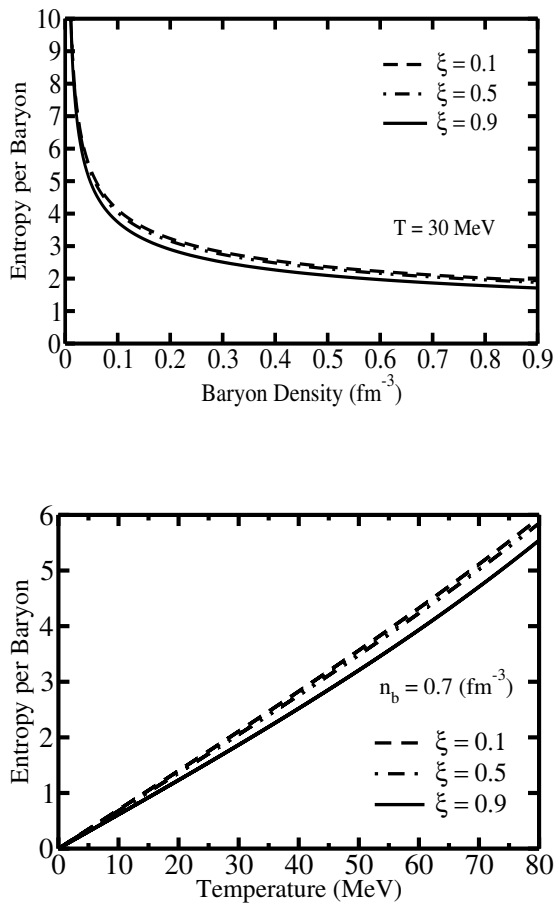

Figure 4: Entropy per baryon.

Phase transitions are determined by studying the entropy of the matter, connected to experimental observables of quark-gluon plasma formation in heavy

ion collisions. The variations of entropy per baryon as a function of baryon density with different values of order parameter ξ is shown in the upper panel of fig. 4. We observe that the entropy decreases with increasing baryon density and at low density, entropy is approximately the same for all the values of the order parameter at a fixed temperature. From the lower panel of fig. 4, we notice that the entropy per baryon in the QGP, is an increasing function of temperature and decreases with increasing ξ. The numerical estimates suggest that the entropy per baryon is continuous along the phase boundary which indicates the possibility that the transition may be a smooth crossover. A crossover transition means the thermodynamic properties change rapidly within a narrow range of values T and μ, but the pressure and all its derivatives remain continuous. Since hadrons being a color singlet cluster of confined quarks, an inference to be drawn from their nature is that a progressive transition to quark matter may occur where hadrons and quarks coexist and interact with one another.

Fig. 5, illustrates the QGP phase diagrams for different ξ at a bag pressure $\mathcal{B} = 208$ MeV fm^{-3}. We find that the critical temperature is independent of the polarization parameter but critical density is different for unpolarized and polarized QGP. Further, we notice, greater values of bag pressure correspond to the larger values of the critical temperature for the deconfined phase transition. When the bag pressure is about 442

MeV fm^{-3}, the estimated results for critical temperature is $T_c = 170$ MeV, which is consistent with the lattice results [2]. The location and existence of such critical point is extremely model dependent [3, 4]. However, as in our model, it is a crossover transition, the exact location of the transition is somewhat ambiguous.

3 Specific heat and incompressibility

Specific heat or heat capacity is a thermodynamic quantity characterizing the phase transition of the system. The specific heat at constant volume (C_v) is defined as the quantity of energy needed to raise the temperature of a system by one unit of temperature, it reads as

$$
\begin{aligned}
C_v &= \left(\frac{\partial E_{QGP}}{\partial T}\right)_v \\
&= \left(1 + \frac{g^2}{18\pi^2}\right)\left[(1+\xi)^{2/3} + (1-\xi)^{2/3}\right]3p_f^2 T + \left(\frac{74}{15} + \frac{g^2}{9\pi^2}\right)\pi^2 T^3.
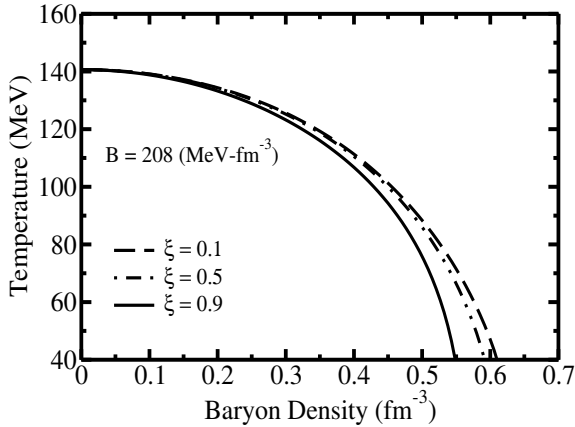\end{aligned} \tag{19}
$$

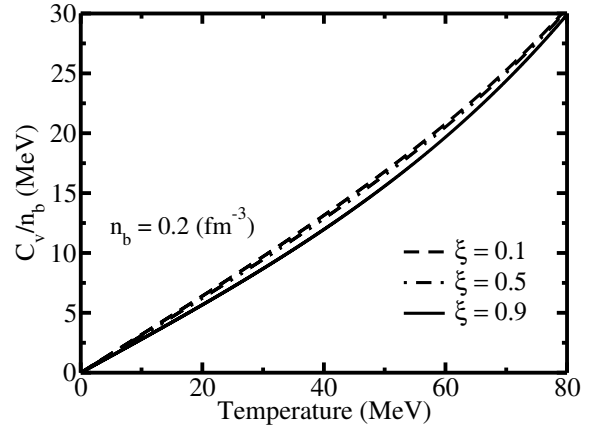Figure 5: The quark hadron phase diagram with different ξ.

Figure 6: Specific heat of QGP for various ξ.

The variation of specific heat as a function of temperature for various ξ is shown in fig. 6. We find higher values of specific heat for unpolarized matter compared to the polarized matter. This is because the specific heat is a measure of energy fluctuation with temperature and the fluctuation is smaller when QGP is polarized. We also observe that the specific heat indicates com-

plete continuity in its behavior along the phase boundary.

Another fundamental quantity to understand the EOS is the incompressibility or compression modulus. A measure of the curvature of the energy density with respect to the number density about the minimum is given by the incompressibility. Mathematically, it is given by

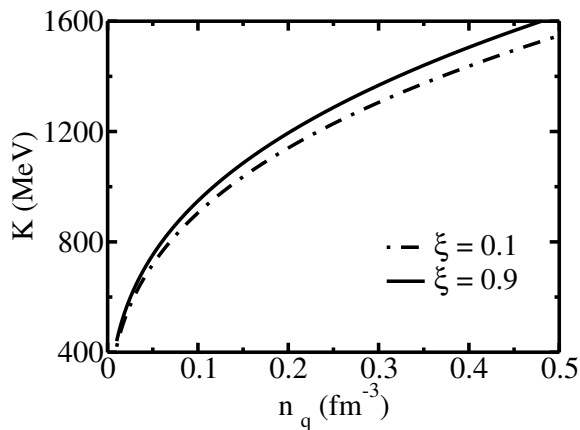$$K = 9n_q \frac{\partial^2 E}{\partial n_q^2} \qquad (20)$$

Figure 7: Variation of incompressibility with density.

We have shown the density dependencies of the incompressibility in fig. (7) with various ξ. This shows, for a higher value of the order parameter ξ, the incompressibility becomes larger for the same value of number density. Thus the numerical values for incompressibility reveals that the EOS for polarized QGP is stiffer than unpolarized one.

4   Summary and conclusion

In the present work a MIT bag model has been used to compute the thermal equilibrium properties and to explore the phase diagram of QGP composed of spin asymmetric massless quarks and gluons, in various regions of temperature and baryon density. The existence of such new phases of quark matter can be ascertained only by the study of quark-hadron phase transition and searching the signatures left by such phase transition. Several observables are associated with the estimation of the free energy, entropy, pressure, specific heat, incompressibility which primarily depends on baryon density and temperature. Observed results of free energy which enhances with increase of spin asymmetry parameter suggests that the QGP with unpolarized quark is energetically favorable. We have found that the entropy is a decreasing function of density and an increasing function of temperature. This is because at high density most of the particles in the system are hadrons and number of degrees of freedom are small while at high temperature most of the particles in the system are partons with large number of degrees of freedom. Within our phenomenological model, we have illustrated the variation of pressure and incompressibility as a function of temperature and baryon density which indicate the increase of temperature or density makes the stiffer equation of state. We have estimated the critical temperature for quark-hadron phase transition by constructing a phase boundary, which is very

close to the lattice QCD calculations. We have clearly observed a continuous variation of entropy density or specific heat along the phase boundary, indicating a crossover from the hadronic phase to quark-gluon phase. The crossover of the entropy is an important signal of QGP production. The inclusion of multi-gluon-exchange processes for multi-flavor systems may strengthen our conclusion.

References

[1] Hwa R.C. *"Quark Gluon Plasma, Vol. 1, 2 & 3"*. World Scientific (1990, 1995, 2004).

[2] Karsch F. "Lattice results on QCD thermodynamics". Nucl.Phys.A. 2002; 698: pp.199.

[3] Fodor Z, Katz S.D. "Critical point of QCD at finite T and μ, lattice results for physical quark masses". JHEP 2004; 04: pp.050.

[4] Stephanov M, Rajagopal K, Shuryak E. "Event-by-event fluctuations in heavy ion collisions and the QCD critical point". Phys.Rev.D. 1999; 60: pp.114028.

[5] Baym G, Pethick C. *"Landau Fermi-Liquid Theory: Concepts and Applications"*. John Wiley & Sons, Inc, (1991).

[6] Modarres M, Mohamadnejad A. "The thermodynamic properties of weakly interacting quark gluon plasma via the one gluon exchange interaction". Phys.Part.Nucl.Lett. 2013; 10: pp.99.

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

[8] Pal K. "Energy density and susceptibility of warm and dense quark matter". Physics Education 2020 (April-June); 36/2/8: pp.1-6.

Demonstrating the essence of Liouville's theorem using examples

Reshma. P[1], Prasanth. P[2] and K. M. Udayanandan[3]

[1]Department of Physics, S.N.Polytechnic College, Kanhangad, Kerala, India.
[2]Department of Physics, Govt. Engineering College, Thrissur, Kerala, India.
[3]Former H O D, NAS College, Kanhangad, Kerala, India.
udayanandan@gmail.com

Abstract

In almost every textbook on dynamics, classical or statistical mechanics, the reference to phase space is connected with Liouville's theorem(LT). Many books just give the proof of it with out giving its essence or significance in detail. In this article we try to demonstrate the essence of LT using some examples.

1 Historical basis

Liouville's theorem named after Joseph Liouville(1809-82) played a key role in classical, statistical and Hamiltonian mechanics in the 19th century[1]. In 1842 the application of this theorem to dynamics was made by Carl Gustav Jacob Jacobi who recognized that the differential equation class that Liouville has studied were same as Hamilton equations [2, 3, 4] and the conserved property was a product of differentials. In his 1871 papers [5, 6] Boltzmann specified the use of Jacobi's principle of the Last Multiplier to show that a volume of phase space is conserved despite the larger number of degrees of freedom and large number of collisions that take place. This was the first derivation of what today is known as Liouville's theorem [7].

2 Phase space

The concept of phase space was developed in the late 19th century by Ludwig Boltzmann [6], Henri Poincare [8], and Josiah Willard Gibbs [9]. In classical statistical mechanics [10, 11], the phase space is the space of all possible states of a physical system; by state we mean the positions q of all the objects in the system and momenta p . One needs both the position and momentum of system in order to determine the future behavior of that system. Suppose that there are

N particles in a system, there will be 3N position coordinates q_1, q_2....q_{3N} , and 3N momentum coordinates, 3N momentum coordinates p_1, p_2....p_{3N} . Each 6N- tuple denotes a possible state of the mechanical system and the space of all possible states is the phase space. Newtonian mechanics implies that states at one time evolve to a unique point in phase space after a fixed interval of time. The sequence of the micro states of a Newtonian system over the time is referred to as its trajectory in phase space. If we want to study and predict the properties of a dynamic mechanical system, such a mapping from Cartesian space to 6N dimensional phase space is essential.

3 Definition of LT

Liouville's theorem states that the density of representative points (DORP) $\rho(q, p, t)$ is constant along the phase trajectories of the phase points [12, 13]. It can be mathematically expressed as

$$\frac{d\rho(q, p, t)}{dt} = 0 \qquad (1)$$

$$\rho = constant$$

with respect to change in time or we can say the density of representative points is always conserved. This is a very important property of the representative points.

4 Consequence of LT

Phase space volume

$$V_{p,q} = \int \frac{\delta N}{\rho} \qquad (2)$$

where N is the total number of representative points. So

$$\frac{dV_{pq}}{dt} = -\int \frac{1}{\rho^2} \frac{d\rho}{dt} \delta N + \frac{1}{\rho} \frac{d\delta N}{dt} \qquad (3)$$

Both terms on RHS are zero. First term vanishes by the definition of the theorem given by Eq(1). Regarding the second term, δN is the number of phase points in an elemental volume and each phase point represents a definite system. System can neither be created or destroyed. Hence the number of representative points in the elemental volume remains fixed. So the second term in Eq(3) vanishes. Here RHS becomes completely zero and so LHS is also zero. Hence the phase space volume does not change during the evolution. The fact that each point in phase space belongs to a unique trajectory in phase space suggests that time evolution of a set of points in phase space moves like a blob of fluid. This blob of fluid can be incompressible like water, which conserves its volume as it moves, or they can be compressible, like air, which responds to differences in pressure by changing its volume. Liouville's theorem states that phase space volumes are incompressible. In the next section we will give a rigorous proof for the conservation of LT, using the properties of phase space.
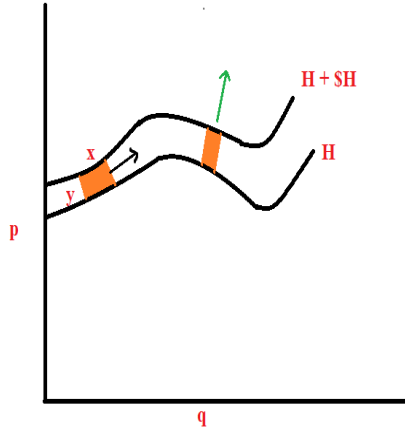
Figure 1:

5 Conservation of volume for phase space flow-A rigorous proof

We had already showed that the volume of a phase space flow is conserved for Hamiltonian systems. We can show this explicitly for a 1D conservative Hamiltonian system using the definition of velocity in phase space. For this let us consider a small rectangular region bounded by two paths in phase space, with energies given by H and $H + \delta H$ respectively as in Figure 1.

The magnitude of the flow vector field is given by

$$|\vec{v}| = \left|\left(\frac{\partial q}{\partial t}, \frac{\partial p}{\partial t}\right)\right| = \left|\left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q}\right)\right|$$

We know that magnitude of $x\hat{i} + y\hat{j}$ is $(x^2 + y^2)^{\frac{1}{2}}$. So

$$|\vec{v}| = \left[\left(\frac{\partial H}{\partial p}\right)^2 + \left(\frac{\partial H}{\partial q}\right)^2\right]^{\frac{1}{2}}$$

Now

$$\nabla H = \frac{\partial H}{\partial p}\hat{i} + \frac{\partial H}{\partial q}\hat{j}$$

So

$$|\nabla H| = \left[\left(\frac{\partial H}{\partial p}\right)^2 + \left(\frac{\partial H}{\partial q}\right)^2\right]^{\frac{1}{2}}$$

or

$$|\vec{v}| = |\nabla H|$$

Then the length x of the area A will be

$$x \approx |\vec{v}|\delta t = |\nabla H|\delta t$$

Now we have to find y in order to find the area A. For that by definition of gradient we have

$$|\nabla H| \approx \frac{\delta H}{y}$$

$\frac{\delta H}{y}$ means magnitude of variation of H with respect to y. Then the area A is

$$A = xy = |\nabla H|\delta ty = \frac{\delta H}{y}\delta ty = \delta H\delta t$$

So what this means is that for any point along the two trajectories, the area will be given by $\delta H\delta t$. But δH and δt are our choice which are constants. Thus the area of the element is conserved as it moves through phase space. In the next section we will consider some examples where we will consider only 2N dimensional space, so the area will be conserved.

6   Demonstration using examples

The interpretation of LT is that if you follow a point as it moves through phase space, then the DORP in the near vicinity of this point remains constant and the consequence is that by Eq(3) volume or area will be conserved. We will consider 4 examples to support the above arguments.

1. **Free particle**
   For a free particle the energy is $\frac{p^2}{2m}$. Hence momentum is independent of q. So the phase space trajectory is given by a straight line.

2. **A particle thrown upwards**
   For a particle thrown upwards the energy is $\frac{p^2}{2m} - mgq$. Here the curves in phase space are rightward-opening parabolas.

3. **A freely falling body**
   For a freely falling body, the energy is $\frac{p^2}{2m} + mgq$. Here the curves in phase space are leftward-opening parabolas.

4. **A harmonic oscillator**
   For a harmonic oscillator the energy is $\frac{p^2}{2m} + \frac{1}{2}mw^2q^2$ and the phase space trajectory is an ellipse.

   We will show that area will be conserved for all the examples during the flow of phase points, which is the consequence of LT.

6.1   Free particle

Consider a free particle with Hamiltonian

$$H = \frac{p^2}{2m} \qquad (4)$$

where p is the momentum and m is the mass of the particle. The Hamilton's equations give

$$\dot{p} = -\frac{\partial H}{\partial q} = 0 \qquad (5)$$

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m} \qquad (6)$$

Integration of these equations yields

$$p = \text{a constant} \qquad (7)$$

$$q = \frac{p}{m}t + \text{constant} \qquad (8)$$

At $t = 0$, let $p = p_0$ and $q = q_0$, then

$$p = p_0$$

;

$$q = q_0 + \frac{p}{m}t$$

Consider the rectangle *abcd* at time $t = 0$. Taking mass as $m = 1$,

$$p_a = p_b = 1; p_c = p_d = 2$$

$$q_a = q_d = 0; q_b = q_c = 3$$

we get Figure 1. After a time t

$$p'_a = p_a$$

$$p'_b = p_b$$

$$p'_c = p_c$$

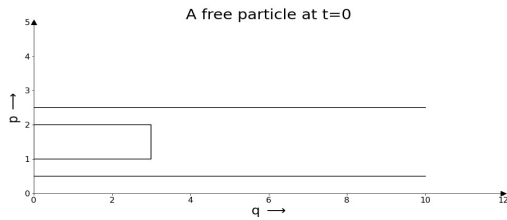$$p'_d = p_d$$

But

$$q'_a = q_a + \frac{p_a}{m}t$$

Figure 2:

Figure 3:

$$q'_b = q_b + \frac{p_b}{m}t$$

$$q'_c = q_c + \frac{p_c}{m}t$$

$$q'_d = q_d + \frac{p_d}{m}t$$

Let $t = 3$ and we get

$$p'_a = 1$$

$$p'_b = 1$$

$$p'_c = 2$$

$$p'_d = 2$$

$$q'_a = 0 + \frac{1}{1}3 = 3$$

$$q'_b = 3 + \frac{1}{1}3 = 6$$
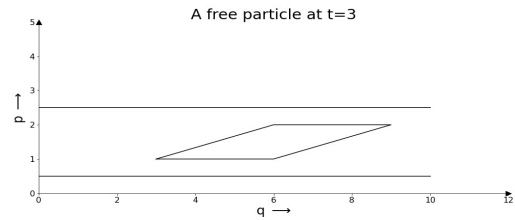
$$q'_c = 3 + \frac{2}{1}3 = 9$$

$$q'_d = 0 + \frac{2}{1}3 = 6$$

After the evolution of time it is found that the rectangle changes to parallelogram. This parallelogram is plotted in Figure 3. In Figure 3 we can see the transformation of the phase space area. Now let us carry out calculations to show that irrespective of the transformation the area in both the cases remains the same. Let us consider the rectangle with length $q_b - q_a$ and breadth $p_d - p_a$. Therefore,

$$\text{Area of the rectangle} = (q_b - q_a)(p_d - p_a)$$

$$= 3 \times 1 = 3$$

For the parallelogram, the base length is $q'_b - q'_a$ and the height is h and hence

$$\text{Area of the parallelogram} = (q'_b - q'_a)(h)$$
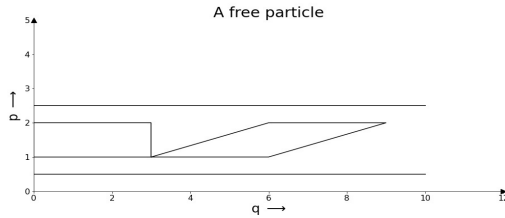
$$= 3 \times 1 = 3$$

Figure 4:

Thus we arrive at the inference that the area remains the same, which is the essence of LT.

6.2 A body thrown upwards

Consider a particle thrown upwards with Hamiltonian

$$H = \frac{p^2}{2m} - mgq \qquad (9)$$

The Hamilton's equations give

$$\dot{p} = -\frac{\partial H}{\partial q} = mg \qquad (10)$$

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m} \qquad (11)$$

Integrating Eq (8) we get

$$p = mgt + constant$$

Substituting Eq (8) in Eq (9)

$$\dot{q} = gt + constant$$

At $t = 0, v = v_0$, then

$$\frac{dq}{dt} = gt + v_0$$

$$dq = gtdt + v_0dt$$

Integrating we get

$$q = \frac{1}{2}gt^2 + v_0t + constant$$

At $t = 0$, $p = p_0$ and $p = p_0$ then

$$p = p_0 + mgt \qquad (12)$$

$$q = q_0 + \frac{p_0}{2m}t + \frac{1}{2}gt^2 \qquad (13)$$

Substituting for $t = \frac{p-p_0}{mg}$ and simplifying we get

$$q = q_0 + \left(\frac{p^2 - p_0^2}{2m^2g} \right)$$

which represents a parabola with slope $\frac{dp}{dq} = \frac{m^2g}{p}$

Consider the rectangle $abcd$ at time $t = 0$ and the corresponding parallelogram $a'b'c'd'$ after a time t.

$$p_a = p_b \; ; \; p_c = p_d$$

$$q_a = q_d \; ; \; q_b = q_c$$

After a time t

$$p'_a = p_a + mgt \; ; \; p'_b = p_b + mgt$$

$$p'_c = p_c + mgt \; ; \; p'_d = p_d + mgt$$

and

$$q'_a = q_a + \frac{p_a}{2m}t + \frac{1}{2}gt^2 \; ; \; q'_b = q_b + \frac{p_b}{2m}t + \frac{1}{2}gt^2$$

$$q'_c = q_c + \frac{p_c}{2m}t + \frac{1}{2}gt^2 \; ; \; q'_d = q_d + \frac{p_d}{2m}t + \frac{1}{2}gt^2$$

After the evolution of time it is found that the rectangle changes to parallelogram at a later time. Let $m = g = 1$,we get
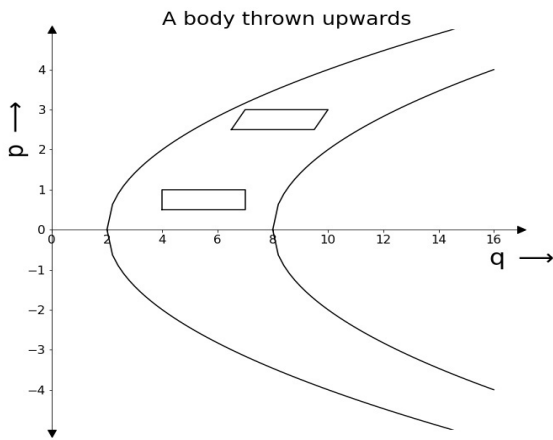
A body thrown upwards

Figure 5:

t=0	t=2
$p_a = 0.5$	$p'_a = 2.5$
$p_b = 0.5$	$p'_b = 2.5$
$p_c = 1$	$p'_c = 3$
$p_d = 1$	$p'_d = 3$
$q_a = 4$	$q'_a = 6.5$
$q_b = 7$	$q'_b = 9.5$
$q_c = 7$	$q'_c = 10$
$q_d = 4$	$q'_d = 7$

In Figure 5 we can see the transformation of the phase space area. Here

Area of rectangle $= (7 - 4)(1 - 0.5) = 1.5$

Area of parallelogram $= (9.5 - 6.5)(0.5) = 1.5$

We can see that the shape changes but the area remains the same.

6.3   Freely falling body

Consider a particle thrown upwards with Hamiltonian

$$H = \frac{p^2}{2m} + mgq \qquad (14)$$

The Hamilton's equations give

$$\dot{p} = -\frac{\partial H}{\partial q} = -mg \qquad (15)$$

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m} \qquad (16)$$

Integrating we get

$$p = -mgt + constant$$

Using the same procedure as a body thrown upwards we arrive at

$$q = -\frac{1}{2}gt^2 + v_0 t + constant$$

At $t = 0$, $p = p_0$ and $q = q_0$ then

$$p = p_0 - mgt \qquad (17)$$

$$q = q_0 + \frac{p_0}{2m}t - \frac{1}{2}gt^2 \qquad (18)$$

Substituting for $t = \frac{p_0 - p}{mg}$ and simplifying we get

$$q = q_0 - \left(\frac{p^2 - p_0^2}{2m^2 g} \right)$$

which represents a parabola with slope $\frac{dp}{dq} = -\frac{m^2 g}{p}$

Consider the rectangle $abcd$ at time $t = 0$ and the corresponding parallelogram $a'b'c'd'$ after a time t.
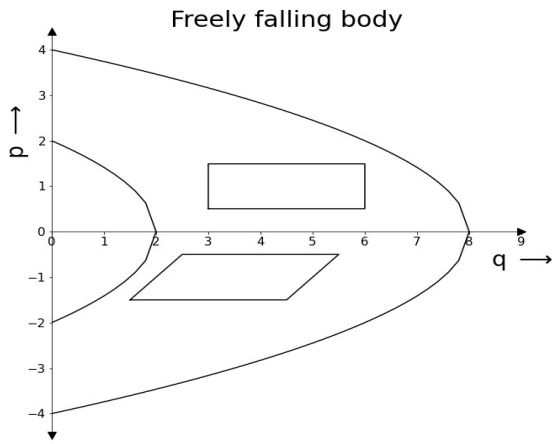
$$p_a = p_b \ ; \ p_c = p_d$$

Figure 6:

$$q_a = q_d \ ; \ q_b = q_c$$

After a time t

$$p'_a = p_a - mgt \ ; \ p'_b = p_b - mgt$$

$$p'_c = p_c - mgt \ ; \ p'_d = p_d - mgt$$

and

$$q'_a = q_a + \frac{p_a}{2m}t - \frac{1}{2}gt^2 \ ; \ q'_b = q_b + \frac{p_b}{2m}t - \frac{1}{2}gt^2$$

$$q'_c = q_c + \frac{p_c}{2m}t - \frac{1}{2}gt^2 \ ; \ q'_d = q_d + \frac{p_d}{2m}t - \frac{1}{2}gt^2$$

After the evolution of time it is found that the rectangle changes to parallelogram at a later time. Let $m = g = 1$

t=0	t=2
$p_a = 0.5$	$p'_a = -1.5$
$p_b = 0.5$	$p'_b = -1.5$
$p_c = 1.5$	$p'_c = -0.5$
$p_d = 1.5$	$p'_d = -0.5$
$q_a = 3$	$q'_a = 1.5$
$q_b = 6$	$q'_b = 4.5$
$q_c = 6$	$q'_c = 5.5$
$q_d = 3$	$q'_d = 2.5$

Figure 6 we can see the transformation of the phase space area. Here

$$\text{Area of rectangle} = (6 - 3)(1.5 - 0.5) = 3$$

$$\text{Area of parallelogram} = (4.5 - 1.5)(1) = 3$$

6.4 Harmonic oscillator

Consider a particle in a harmonic oscillator with Hamiltonian

$$H = \frac{p^2}{2m} + \frac{1}{2}mw^2q^2 \qquad (19)$$

Hamilton's equations give

$$\dot{p} = -\frac{\partial H}{\partial q} = -mw^2q \qquad (20)$$

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m} \qquad (21)$$

We have

$$m\frac{d}{dt}\left(\frac{\partial H}{\partial p}\right) + \left(\frac{\partial H}{\partial q}\right) = 0$$

$$m\ddot{q} + mw^2q = 0$$

$$\ddot{q} = -\omega^2 q$$

The general solution of this equation can be written as

$$q = A\cos\omega t + B\sin\omega t$$

and

$$v = \dot{q} = -\omega A\sin\omega t + \omega B\cos\omega t$$

At t=0, $q = q_0$ and $v = v_0 = \omega b$ therefore, $b = \frac{v_0}{\omega}$ Hence the solution can be written as

$$p = p_0\cos\omega t - m\omega q_0\sin\omega t \qquad (22)$$

$$q = q_0 \cos \omega t + \frac{p_0}{m\omega} \sin \omega t \qquad (23)$$

We have total energy of simple harmonic oscillator

$$E = \frac{p^2}{2m} + \frac{1}{2} m\omega^2 q^2$$

$$\frac{p^2}{2mE} + \frac{q^2}{\frac{2E}{m\omega^2}} = 1$$

$$\frac{p^2}{a^2} + \frac{q^2}{b^2} = 1$$

which represents an ellipse where $a^2 = 2mE$ and $b^2 = \frac{2E}{m\omega^2}$

Put $m = 1$, $a = 3$, $b = 6$ then we get $E = 4.5$ and $\omega = 0.5$. Consider the rectangle $abcd$ at time $t = 0$ and the corresponding parallelogram $a'b'c'd'$ after a time t.

$$p_a = p_b \; ; \; p_c = p_d$$

$$q_a = q_d \; ; \; q_b = q_c$$

After the evolution of time it is found that the rectangle changes to parallelogram at a later time. Let $m = g = 1$,

t=0	$t = \frac{\pi}{4\omega}$
$p_a = 1$	$p'_a = -1.4$
$p_b = 1$	$p'_b = -2.8$
$p_c = 2$	$p'_c = -2.1$
$p_d = 2$	$p'_d = -0.7$
$q_a = 6$	$q'_a = 5.6$
$q_b = 10$	$q'_b = 8.5$
$q_c = 10$	$q'_c = 9.9$
$q_d = 6$	$q'_d = 7.0$

Figure 7 we can see the transformation of the phase space area. Here

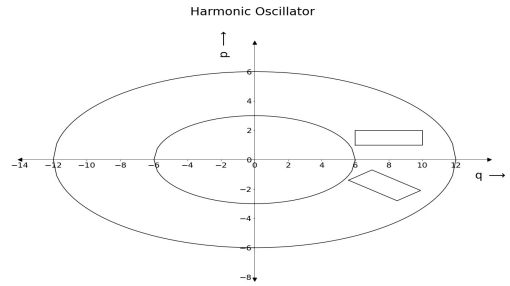Area of rectangle = Area of parallelogram = 4 units.

Figure 7:

7 Free particles flow in phase space with a different initial condition

Here in this section we will consider the initial shape of the flow to be a triangle instead of a rectangle. This is to give a pictorial demonstration that any initial shape will evolve to a correspondingly new shape. We take the free particle Hamiltonian once again to show this. Taking the equations from Section 6.1

$$p = \text{a constant} \qquad (24)$$

$$q = \frac{p}{m}t + \text{constant} \qquad (25)$$

At $t = 0$, let $p = p_0$ and $q = q_0$, then

$$p = p_0$$

$$q = q_0 + \frac{p}{m}t$$

After a time $t = \frac{\pi}{4\omega}$

$$p'_a = p_a \cos\frac{\pi}{4} - m\omega q_a \sin\frac{\pi}{4} \; ; \; p'_b = p_b \cos\frac{\pi}{4} - m\omega q_b \sin\frac{\pi}{4}$$

$$p'_c = p_c \cos\frac{\pi}{4} - m\omega q_c \sin\frac{\pi}{4} \; ; \; p'_d = p_d \cos\frac{\pi}{4} - m\omega q_d \sin\frac{\pi}{4}$$

and

$$q'_a = \frac{p_a}{m\omega}\sin\frac{\pi}{4} + q_a \cos\frac{\pi}{4} \; ; \; q'_b = \frac{p_b}{m\omega}\sin\frac{\pi}{4} + q_b \cos\frac{\pi}{4}$$

$$q'_c = \frac{p_c}{m\omega}\sin\frac{\pi}{4} + q_c \cos\frac{\pi}{4} \; ; \; q'_d = \frac{p_d}{m\omega}\sin\frac{\pi}{4} + q_d \cos\frac{\pi}{4}$$

Consider the triangle abc at time $t = 0$. Taking mass $m = 1$,

$$p_a = p_b = 0; p_c = 2$$

$$q_a = 0; q_b = 2 \; ; q_c = 1$$

After a time t

$$p'_a = p_a$$

$$p'_b = p_b$$

$$p'_c = p_c$$

But

$$q'_a = q_a + \frac{p_a}{m}t$$

$$q'_b = q_b + \frac{p_b}{m}t$$

$$q'_c = q_c + \frac{p_c}{m}t$$

t=0	t=1
$p_a = 0$	$p'_a = 0$
$p_b = 0$	$p'_b = 0$
$p_c = 2$	$p'_c = 2$
$q_a = 0$	$q'_a = 0$
$q_b = 2$	$q'_b = 2$
$q_c = 1$	$q'_c = 3$

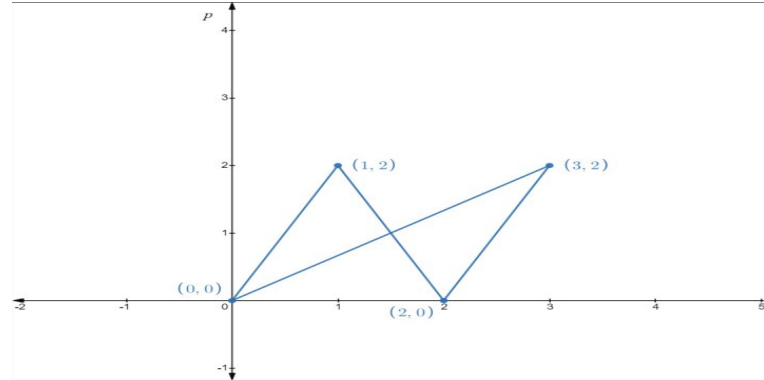Figure 8:

Here

Area of a triangle $= \sqrt{s(s-a)(s-b)(s-c)}$ (26)

where $s = \frac{a+b+c}{2}$. In the first triangle using Eq (26) we get

$$\text{Area } = 1.9\text{units}$$

and for the second triangle

$$\text{Area } = 1.9\text{units}$$

We can see that the shape changes, but the area remains the same.

8   Relevance of Liouville's theorem in statistical mechanics

Liouville's theorem has proved its utility to its greatest extent in the field of many particle systems. It is considered to be a fundamental theory in classical statistical mechanics. The insight it provides in the description of motion of a system in terms of time evolution of phase space distribution function made it to form the cornerstone of statistical mechanics. We list out some of the applications of LT for its better understanding.

1. Photons in Phase Space: Liouvilles theorem describes a number of properties associated with light propagation, including brightness and specific intensity. We see that the specific intensity of photon remains conserved as the radiated energy is constant over any area of interest.[14]

2. First Law of Thermodynamics: Liouvilles theorem is used to construct the proof of first law of thermodynamics[15]

3. Poincares recurrence theorem: Since the volume of phase space element remains invariant due to Liouvilles theorem the system will always return to its initial state, clearly in agreement with the assumptions of Poincares recurrence theorem.[16]

4. Connection to Quantum mechanics: Quantum Liouville theorem derived from the classical principles helps to study the time evolution of operators, expectation values, and density operators in quantum mechanics.[17, 18, 19]

9   Conclusions

In this article we make an attempt to demonstrate the essence of LT, that volume (area for 2D systems) will be conserved during phase space flow, with the help of some simple examples. There are always simple ways to understand any difficult theory in physics. We hope our examples will help the students to get a clear picture about the meaning of the Liouville's theorem, which can help them to find more uses for the theorem.

Acknowledgements

References

[1] J. Liouville (1838). J. Math. Pures Appl. 3, 342

[2] C. G. J. Jacobi. Vorlesungen ber Dynamik, (Georg Reimer, Berlin 1866)

[3] W. R. Hamilton(1834). Philos. Trans. R. Soc. London 124, 247

[4] W. R. Hamilton(1835). Philos. Trans. R. Soc. London 125, 95

[5] L. Boltzmann(1868). Wiener Berichte 58, 517

[6] L. Boltzmann(1871). Wien. Ber. 63, 679

[7] D. D. Nolte. The Tangled Tale of Phase Space, Chapter 6 in Galileo Unbound: A Path Across Life, the Universe and Everything (Oxford University Press, 2018)

[8] H. Poincare(1890). Acta mathematica 13,1

[9] Gibbs, Josiah Willard. Principles in Statistical Mechanics, (New York: Charles Scribner's Sons, 1902)

[10] W. Greiner, L. Neise, H. Stcker. Thermodynamics and Statistical Mechanics, (springer, 2001) ISBN 9780387942995

[11] H.Goldstein, J. L. Safko, C.P. Poole. Classical Mechanics, (Addison-Wesley 2002) ISBN 978-0-201-65702-9

[12] R. K. Pathria and P. D. Beale. Statistical Mechanics, (Butterworth, 2009)

[13] Huang K . Statistical Mechanics, Second Edition, (John Wiley and Sons, 1987)

[14] Implications of Liouville's theorem on the apparent brightness temperatures of solar radio bursts, Melrose, D. B. and Dulk, G. A., Solar Physics vol. 116, no. 1, 1988, p. 141-156

[15] "Liouville's Theorem as a Subtle Statement of the First Law of Thermodynamics."Nyambuya, G.G. , Progress in Physics, vol. 15, no. 3, Oct. 2019, pp. 178

[16] Marius Oltean, Luca Bonetti, Alessandro D. A. M. Spallicci, Carlos F. Sopuerta., Entropy theorems in classical mechanics, general relativity, and the gravitational two-body problem, Phys. Rev. D, 2016, 94 (6), pp.064049.

[17] A new look at the quantum Liouville theorem P. T. Leung and G. J. N, Journal for Foundations and Applications of Physics, vol. 7, No. 1 (2020)

[18] Does Liouvilles theorem imply quantum mechanics, C. Syros, International Journal of Modern Physics B, Vol. 13, No. 02, pp. 161-189 (1999)

[19] Quantum mechanical version of classical Liouville theorem, Xie Chuan-Meia, Fan Hong-Yib Chin. Phys. B, 2013, Vol. 22(3): 030501

A simplistic approach to Ergodic hypothesis

Prasanth. P[1], Reshma. P[2] and K. M. Udayanandan[3]

[1]Department of Physics, Govt. Engineering College, Thrissur, Kerala, India.
[2]Department of Physics, S.N.Polytechnic College, Kanhangad, Kerala, India.
[3]Former H O D, NAS College, Kanhangad, Kerala, India.

udayanandan@gmail.com

Submitted on 02-07-2021

Abstract

Ergodic hypothesis is the basic theory of statistical mechanics, which is very difficult to comprehend by many students. Simplified literature regarding it, in the undergraduate level is almost unavailable. In this article we are trying to make a simplistic approach to the Ergodic hypothesis without compromising the fundamental concepts. A short description about Landau's non ergodic approach is also given at the end of the article.

1 Introduction

During the 1870s and 1880s [1, 2, 3], various forms of the Ergodic hypothesis were used by Boltzmann in his works on the foundations of statistical mechanics. Boltzmann's hypothesis consisted in assuming that the average of a function in time, obtained by following the points of its trajectory would be equal to the phase average. Later, the ensemble method became a basic tool of statistical mechanics through the work of Gibbs. In 1902 J.W.Gibbs [4] devised and introduced the concept of ensemble average based on the realization that a system at equilibrium must have a very large number of microscopic states. Instead of tracking the time evolution of the system over the microscopic states as required by Newtonian mechanics, we consider mentally a large number of systems which have the same macroscopic properties called an ensemble. As an alternative to avoid time averaging Gibbs work helped in making the Ergodic hypothesis assume a central role in statistical mechanics. Let us consider any observable property, say X. The ensemble average is a simple average over all the members of the ensemble.

$$< X >= \lim_{N->\infty} \frac{1}{N} \sum_{i=1}^{N} n_i X_i$$

Taking $\frac{n_i}{N} = p_i$

$$< X >= \lim_{N->\infty} \sum_{i=1}^{N} p_i X_i$$

where p_i is the probability of a member of the ensemble to be in the ith microscopic state. Thus time average was replaced by ensemble average by Gibbs. In one of our earlier article[5] we had shown that using canonical ensemble method, it is very easy to obtain the properties very easily. Here in this article we will try to obtain van der Waals equation of state using both time average and ensemble average methods.

2 Ergodic Hypothesis[EH] for Real gases

Consider N identical particles in a volume V. In the case of an ideal gas particles are considered to be free and do not interact, whereas in real gases particles interact through a potential u(r) where r is the inter particle distance. The virial equation [6] of an ideal gas can be written as

$$P(T, \rho) = \rho kT \qquad (1)$$

where P is the pressure, T the absolute temperature and $\rho = \frac{N}{V}$ the number density. For a real gas

$$P(T, \rho) = kT \left[\rho + \sum_{i=2}^{N} B_i(T) \rho^2 \right]$$

where $B_i(T)$ are the expansion coefficients called virial coefficients. Here for our demonstration of the application of EH we will take only the first correction or second virial coefficient $B_2(T)$ and hence

$$P(T, \rho) = kT[\rho + B_2 \rho^2] \qquad (2)$$

$$\beta P = [\rho + B_2 \rho^2] \qquad (3)$$

where $\beta = \frac{1}{kT}$. Let us try to find from virial theorem, the expression for B_2 in terms of time average and ensemble average.

2.1 Virial theorem

Consider the virial

$$G = \sum_i p_i \cdot r_i$$

where G is considered to be a quantity, which is the product of the momentum and the position of the particle [7, 8]. Taking the derivative,

$$\frac{dG}{dt} = \sum_{i=1}^{N} \dot{p}_i \cdot r_i + \sum_{i=1}^{N} p_i \cdot \dot{r}_i = \sum_{i=1}^{N} F_i \cdot r_i + \sum_{i=1}^{N} \frac{p_i^2}{m}$$

Taking the time average we get RHS is equal to,

$$\sum_{i=1}^{N} \frac{1}{\tau} \int_0^\tau (F_i \cdot r_i) dt + \sum_{i=1}^{N} \frac{1}{\tau} \int_0^\tau \left(\frac{p_i^2}{m} \right) dt$$

We define

$$\left\langle \frac{dG}{dt} \right\rangle = \frac{1}{\tau} \int_0^\tau \frac{dG}{dt} dt \qquad (4)$$

Hence

$$\left\langle \frac{dG}{dt} \right\rangle = \sum_{i=1}^{N} \langle F_i \cdot r_i \rangle + \sum_{i=1}^{N} \frac{\langle p_i^2 \rangle}{m}$$

For a bounded system, $\langle \frac{dG}{dt} \rangle$ is zero, and hence we get

$$\sum_{i=1}^{N} \langle F_i \cdot r_i \rangle = -2 \langle K \rangle \qquad (5)$$

Here $K = \frac{p_i^2}{2m}$ is the kinetic energy a function of momenta, and F_i the total force acting on the i-th particle which is the sum of internal and external forces (the last ones coming from the walls confining the gas).

2.2   Ideal gas

Let us find the equation of state for free particles using virial theorem in this section and in next section we will try to obtain equation of state for real system. The method given below is available in books but for continuity we will make a repetition here. We have $F_i = F_i^{int} + F_i^{ext}$. For an ideal gas there are no internal forces. So

$$\sum_{i=1}^{N} \langle F_i^{ext} \cdot r_i \rangle = -2\langle K \rangle \qquad (6)$$

Since the external forces on the walls of the container is due to the pressure, applying Gauss theorem we get

$$\sum_{i=1}^{N} \langle F_i^{ext} \cdot r_i \rangle = -P \int_A \vec{r} \cdot \vec{dA} = -3PV$$

where dA is the area element on the surface of the container. From Tolman definition of equipartition theorem [9], for N particles in 3D the average kinetic energy

$$\langle K \rangle = \frac{3}{2} NkT$$

Substituting in (6) we get

$$P = \frac{NkT}{V} = \rho kT$$

$$\beta P = \rho$$

We have no difficulty in obtaining the equation of state.

2.3   Real gas

For real gases since the interaction plays an important role, the internal force comes into

play. The particles of the gas interact pairwise, i.e. it is assumed that u is the sum of energy terms, each depending only on the distance apart r_{ij} of two particles i and j. In the system of N particles there are $\frac{N(N-1)}{2}$ different pairs which can be formed, so that the assumption is that u is the sum of $\frac{N(N-1)}{2}$ terms. Then for large N

$$\sum_{i,j} F_{ij}^{int} \cdot r_{ij} = \frac{N^2}{2} \sum_i F_i^{int} \cdot r_i = -\frac{N^2}{2} \sum_i r_i \frac{du}{dr_i}$$

2.3.1   Time Average

In terms of time average we define

$$-\frac{N^2}{2} \left\langle r \frac{du}{dr} \right\rangle = -\frac{N^2}{2} \frac{1}{\tau} \int_0^{\tau} \left(r \frac{du}{dr} \right) dt$$

Therefore substituting in (5) we write,

$$-3PV - \frac{N^2}{2} \frac{1}{\tau} \int_0^{\tau} r \left(\frac{du}{dr} \right) dt = -3NkT$$

The equation of state becomes

$$P = \frac{NkT}{V} - \frac{N^2}{6V} \frac{1}{\tau} \int_0^{\tau} r \left(\frac{du}{dr} \right) dt$$

$$\beta P = \rho - \left(\frac{\beta}{6} \right) \rho^2 \frac{V}{\tau} \int_0^{\tau} r \left(\frac{du}{dr} \right) dt$$

Comparing we write,

$$B_2(T) = -\left(\frac{\beta}{6} \right) \frac{V}{\tau} \int_0^{\tau} r \left(\frac{du}{dr} \right) dt$$

How to solve this without the knowledge of time, it is impossible. This is how founders of statistical mechanics were forced to adopt the ensemble theory.

2.3.2 Ensemble Average

In terms of pair distribution function $g(r)$ we can write

$$-\frac{N^2}{2}\left\langle r\frac{du}{dr}\right\rangle =$$

$$-\frac{N^2}{2}\int\int r\left(\frac{du}{dr}\right)g(r_2-r_1)\frac{dr_1 dr_2}{V^2}$$

$$=-\frac{N^2}{2V}\int_0^\infty r\frac{du}{dr}g(r)\,4\pi r^2 dr$$

$g(r)$ is the probability of finding a pair of particles separated by a distance r. Substituting in (5) and carrying out suitable rearrangement we get the equation of state ,

$$P=\frac{NkT}{V}-\frac{1}{6}\frac{N^2}{V^2}\int_0^\infty g(r)\left(r\frac{du}{dr}\right)4\pi r^2 dr$$

$$\beta P=\rho-\frac{\beta}{6}\rho^2\int_0^\infty g(r)\left(r\frac{du}{dr}\right)4\pi r^2 dr$$

We have as a first approximation $g(r)=e^{\left(\frac{-u}{kT}\right)}$, therefore

$$B_2=-\frac{\beta}{6}\int_0^\infty e^{-\beta u}\left(r\frac{du}{dr}\right)4\pi r^2 dr$$

Integrating by parts we get

$$B_2=-2\pi\int_0^\infty(e^{-\beta u(r)}-1)r^2 dr$$

Let us consider the Sutherland potential

$$u(r)=\begin{cases}\infty;\ r<r_0\\ -u_0\left(\frac{r_0}{r}\right)^6;\ r\geq r_0\end{cases}\qquad(7)$$

We write

$$\int_0^\infty(e^{-\beta u(r)}-1)r^2 dr=$$

$$\left[-\int_0^{r_0}r^2 dr+\int_{r_0}^\infty\left(e^{\left(\frac{u_0}{kT}\left(\frac{r_0}{r}\right)^6\right)}-1\right)r^2 dr\right]$$

If we assume that $\frac{u_0}{kT}\ll 1$

$$\left(e^{\left(\frac{u_0}{kT}\left(\frac{r_0}{r}\right)^6\right)}-1\right)\approx\frac{u_0}{kT}\left(\frac{r_0}{r}\right)^6$$

and hence

$$\int_0^\infty(e^{-\beta u(r)}-1)r^2 dr=-\frac{r_0^3}{3}\left[1-\frac{u_0}{kT}\right]$$

Therefore,

$$B_2=\frac{2\pi r_0^3}{3}\left[1-\frac{u_0}{kT}\right]$$

Substituting in (3) we arrive at the equation of state as

$$P=\frac{NkT}{V}\left[1+\frac{N}{V}\frac{2\pi r_0^3}{3}\left[1-\frac{u_0}{kT}\right]\right]$$

$$\left(P+\frac{2\pi r_0^3 u_0}{3v^2}\right)\left(v-\frac{2\pi r_0^3}{3}\right)=kT$$

$$\left(P+\frac{a}{v^2}\right)(v-b)=kT$$

which is the van der Waals equation of state for a real gas where $v=\frac{V}{N}$, $a=\frac{2\pi r_0^3}{3}u_0$ and $b=\frac{2\pi r_0^3}{3}=4\left[\frac{4\pi}{3}\left(\frac{r_0}{2}\right)^3\right]$.

3 Why the time and the ensemble averages are same?

Ergodic hypothesis says time average and ensemble average are same. Immediate question will be what will be underlying reason for it. We with some examples will show that the probability in both cases will be same which helps us to equate both. This is not a proof of Ergodic hypothesis, but

only the logic behind the hypothesis. We can demonstrate this using some simple examples.

1. **Free particle**

   Consider a free particle bouncing between two walls of length L, the probability that the particle is in between x and $x + dx$ is

   $$P(x)\,dx = \frac{2dt}{T} \qquad (8)$$

   and velocity,

   $$v = \frac{dx}{dt}$$

   If T is the total time then, $v = \frac{L}{T}$. Comparing and rearranging we get,

   $$P(x)\,dx = \frac{2dx}{L}$$

   Now in phase space for a free particle $E = \frac{p^2}{2m}$ and momentum $p = \sqrt{2mE}$. Since the particle is bouncing between walls of container of length L, total area $= \sqrt{2mE}L$. Consider that our oscillatory energy changes from E to $E + dE$ then change in area, $dA = \frac{\sqrt{m}}{\sqrt{2E}}L\,dE$. probability,

   $$P(x)\,dx = \frac{2dp\,dx}{dA} \qquad (9)$$

   we have

   $$dp\,dx = \frac{m}{p}dE\,dx \qquad (10)$$

   Substituting and on simplification we get probability,

   $$P(x)\,dx = \frac{2dx}{L}$$

2. **Harmonic Oscillator**

   Consider a single particle making a one-dimensional oscillation for a long time. Let

   $$x = A\sin\omega t$$

   $$\frac{dx}{dt} = A\omega\cos\omega t$$

   Considering (8) and substituting velocity, $v = \frac{dx}{dt}$. and period of oscillation $T = \frac{2\pi}{\omega}$, the probability that the particle is in between x and $x + dx$ is

   $$P(x)\,dx = \frac{\omega}{\pi}\frac{dx}{A\omega\cos\omega t}$$

   Using the relation $\sin^2\omega t + \cos^2\omega t = 1$, we get

   $$P(x)\,dx = \frac{dx}{\pi\sqrt{A^2 - x^2}}$$

   For a harmonic oscillator, $E = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2$ and $p = \sqrt{2mE - m^2\omega^2 x^2}$. Since a particle having harmonic oscillator energy will trace an ellipse in the phase space, we have total area $= \pi\sqrt{2mE}\sqrt{\frac{2E}{m\omega^2}} = \frac{2\pi E}{\omega}$. If our oscillatory energy changes from E to $E + dE$ then change in area is $dA = \frac{2\pi}{\omega}dE$. Considering (9) and (10) and by suitable substitution, probability

   $$P(x)\,dx = \frac{\omega m}{\pi}\frac{dx}{\sqrt{2mE - m^2\omega^2 x^2}}$$

   But $E = \frac{1}{2}m\omega^2 A^2$ then,

   $$P(x)\,dx = \frac{dx}{\pi\sqrt{A^2 - x^2}}$$

3. **Freely falling body**

Consider a particle with mass m moving in a gravitational field and let the motion be one dimensional. Let

$$x = h - \frac{1}{2}gt^2$$

$$dx = gt\,dt$$

For a freely falling body $h = \frac{1}{2}gT^2$ and substituting in (8), the probability that the particle is in between x and $x + dx$ is,

$$P(x)\,dx = \frac{dx}{\sqrt{h(h-x)}}$$

If we substitute $h = A$, it becomes

$$P(x)\,dx = \frac{dx}{\sqrt{A(A-x)}}$$

For the particle in a gravitational field, $E = \frac{p^2}{2m} + mgx$ and $p = \sqrt{2mE - 2m^2gx}$. Since a particle will trace a parabola in the phase space, we have total area $= \frac{2}{3}\sqrt{2mE}\frac{E}{mg} = \frac{2\sqrt{2}}{3\sqrt{mg}}E^{\frac{3}{2}}$. Consider energy changes from E to $E + dE$ then change in area is $dA = \frac{\sqrt{2E}}{g\sqrt{m}}dE$. Considering (9) and (10) and by suitable substitution, probability

$$P(x)\,dx = \frac{mg}{\sqrt{E(E-mgx)}}dx$$

But $E = mgA$,

$$P(x)\,dx = \frac{dx}{\sqrt{A(A-x)}}$$

Here we see that in the above examples the probability in both the cases is the same which shows time average and ensemble average are the same.

4   An example from daily life

Suppose we wish to know the mean wage(macroscopic quantity, W) of a laborer l. Let us monitor the tasks during a period of 30 days and then calculate the average.

$$W = \mathbf{T}(\mathbf{d}) = \frac{1}{30}\sum_{d=1}^{30} T(d)$$

Now instead of 30 days if we want to know the average in just one day. Here let us consider L laborers with similar tasks as that of l. If the same task is done by L laborers in just one day the the average will be

$$\langle T(b)\rangle = \frac{1}{L}\sum_{b=1}^{L} T(b)$$

The family of laborers can be called as an ensemble. Mathematically EH says,

$$W = \mathbf{T}(\mathbf{d}) = \langle T(b)\rangle$$

Our daily life experience has no confusion in understanding that both will be same.

5   Landau's approach on ergodic theory

In this article we have been trying to justify the basic reasons behind the equivalence of time and ensemble average. But the lack of proofs for the validity of the ergodic hypothesis had created doubts and debates among many physicists. Lev Landau and Lifshitz in their book statistical physics say that ergodic hypothesis may not be needed to develop SM[10, 11]. They start with a system present

in a bath (equivalent to canonical ensemble) with the argument that the interactions of the system with the surrounding bodies makes the system's phase point to wander in the phase space. In a sufficiently long interval of time τ the phase point spends a time $\Delta\tau$ in the volume $\Delta p\Delta q$ in the phase space. The probability that its phase point will be found in the volume $\Delta p\Delta q$ is

$$w = \lim_{\tau\to\infty} \frac{\Delta\tau}{\tau}$$

If $\rho(p,q)$ is the actual probability distribution in the systems phase space from its temporal behavior then we write, $dw = \rho(p,q)dpdq$. Therefore the mean value of any dynamic function can be written as

$$\langle f \rangle = \int \rho(p,q)f(p,q)dpdq$$

From the definition of probability it is obvious that statistical averaging is exactly equivalent to the infinite time averaging.

$$\langle f \rangle = \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau f(t)\,dt$$

The important point to be noted here is that we have only one system under consideration and no statistical ensemble is considered. Thus in Landau's method, one starts with a system present in a bigger system and one assumes the property of statistical independence and then one can reach the canonical distribution and no ergodic hypothesis is required. In Landau's theory the macro observable are considered as sum functions $f = \sum_i f_i$ based on a very general fact of statistical independence. Landau argue that

predictions of statistical mechanics are very reliable due to the fact that the relevant observables take almost constant value on the energy surface. They prove:

$$\frac{\sqrt{<\Delta f^2>}}{<f>} = \frac{1}{\sqrt{N}}$$

Thus equality of time average and phase space average is imposed due to the fact of temporal constancy of special macroscopic observables in equilibrium. The magnitude of fluctuations goes as $\propto \frac{1}{\sqrt{N}}$ which are very small for macroscopic systems. Detailed descriptions of Landau is a topic for another article which we hope can be done later.

6   Conclusions

In this article we are making an attempt to understand the Ergodic hypothesis. We showed that van der Waals equation cannot be obtained using time average, but can be obtained easily using ensemble average. It is also shown the basic reason for this hypothesis is the equality of probability of occurrence of an event in both cases, which make them equivalent.

References

[1] L. Boltzmann (1871). Wien.Ber. 63, 679-711.

[2] L. Boltzmann (1872). Wien. Ber. 66, 275-370.

[3] L. Boltzmann (1884). Creeles Journal 98, 68-94.

[4] Gibbs, Josiah Willard. Elementary Principles in Statistical Mechanics, (New York: Charles Scribner's Sons, 1902).

[5] Prasanth. P, Reshma. P and . K. M. Udayanandan (2020). Physics Education, 36(3).

[6] Kamerlingh Onnes H. (1902). KNAW Proceedings 4, 1901-1902, Amsterdam, 125-147.

[7] R. Clausius (1870). Dublin Philos. Mag. J. Sci. 40, 122-127.

[8] H. Arslan (2013). Appl. Math. 4, 688.

[9] Tolman, R. C. The Principles of Statistical Mechanics, (Oxford: Clarendon Press, 1938).

[10] Landau,Lev and Lifshitz E. M. Statistical Physics, (Pregmon Press,1958).

[11] Navinder Singh (2011). arXiv:1103.4003 v2[cond-mat.stat-mech].

A Simplistic Expression for the Escape Velocity from a Galaxy

Padmanabh S. Sarpotdar[1]

[1]Department of Physics, Khare Dhere Bhosale College, Guhagar.
(Affiliated to University of Mumbai)
Dist. Ratnagiri (MH) 415703, India.
padphy@gmail.com

Abstract

Astronomical observations show that the galaxies have a flat rotation curve. Clearly, they are supposed to have a different density distribution than the other familiar objects like planets and stars. Here a general discussion and related concepts on 'escape velocity' as taught in higher secondary and introductory undergraduate classes are extended to derive a simplistic formula for the escape velocity of an object from a galaxy.

1 Introduction

In the context of an object under the influence of gravitational force due to other bodies, the term 'escape velocity' refers to its minimum speed necessary to "escape" the gravitational field experienced by it, such that, starting from its initial position (say, a distance r) it eventually reaches to a resting position corresponding to $r \to \infty$.

For two point objects *viz.* with masses M and m ($M \gg m$) separated by a distance r, the formula for escape velocity v_e is trivial, as we equate the kinetic energy of the mass m with the gravitation potential experienced by it *i.e.*

$$\frac{1}{2}m{v_e}^2 = \frac{GMm}{r} \implies v_e = \sqrt{\frac{2GM}{r}} \quad (1)$$

Where G is the universal constant of gravitation ($G = 6.67 \cdot 10^{-11} \ N \cdot m^2/kg^2$) One should note that, this gravitational potential GMm/r is also equal to the work required against the gravitational force $\vec{F}(r) = (GMm/r^2)\hat{r}$ while taking the object from its position at r to infinity. This work done against the gravitational force could be determined by

$$W = \int_A^B \vec{F} \cdot \vec{dr} = \int_A^B F \, dr \quad (2)$$

as the object travels from point A to B.

Equation (1) holds true for all the bodies with a spherical mass distribution with an overall constant density profile such as stars, planets *etc.* owing to the two fundamental theorems by Newton:

Newtons first theorem: *A body that is inside a spherical shell of matter experiences no net gravitational force from that shell.*

Newtons second theorem: *The gravitational force on a body that lies outside a spherical shell of matter is the same as it would be if all of the shells matter were concentrated into a point at its center.*

In case of a galaxy though, which contains the stars, dark matter (putative) and the atoms & molecules constituting the interstellar medium, we need to modify the treatment to derive a formula for v_e.

2   The rotation curve and density distribution for a galaxy

As depicted in Figure (1), the composition of spiral galaxy has three major parts[1]: 1. The 'disk' which primarily contains hot, young stars and lots of gas and dust. 2. The 'bulge' which is the centre of the galaxy populated with old stars and with very little amount of gas and dust 3. The spherical 'halo' consisting of old and low-metallicity stars and globular clusters.

Though the most of the mass in a galaxy is present in chunks (discrete distribution), the gravitational force between these large chunks is dominated by distant objects, and

hence could be approximated by a smooth density. Hence, we need to determine the escape velocity for a mass m with respect to a smoothly varying density distribution $\rho(r)$. A rotation curve could be defined for any
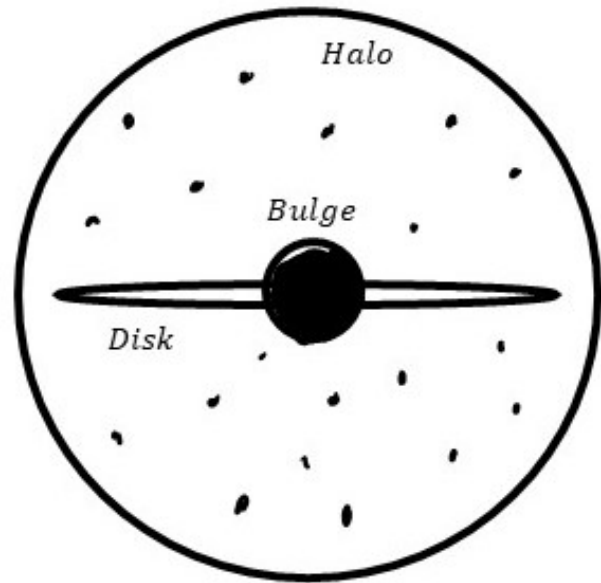

Figure 1: Structure of a galaxy

rotating object, and is understood as a plot showing the variation of orbital speed v_o of an object as a function of distance r from the gravitational centre. Typically, three categories of rotation curves exist as follows:

1. *Keplerian rotation*: (*e.g.* A planet of mass m orbiting around a star of mass M) By equating the centrifugal force with the gravitational force ($mv_o^2/r = GMm/r^2$), we get

$$v_o = \sqrt{GM/r} \implies v_o \propto \frac{1}{\sqrt{r}} \qquad (3)$$

2. *Solid body rotation*: (*e.g.* A particle inside rigid sphere rotating with constant angular speed ω, at a distance r from its centre

of gravity) Here, each particle is performing uniform circular motion with speed

$$v_o = r\omega \implies v_o \propto r \qquad (4)$$

3. *Galactic rotation*: The rotational speeds for a galaxy, as measured through experiments, do not follow the rules mentioned in the above two orbital systems. In fact, there is a discrepancy between the observed rotation curves and the theoretical predictions based on the visible light they emit.

A typical rotation curve for a spiral galaxy is shown in Figure (2), along with the other two cases. It indicates a flat rotation curve (or constant v_o) which implies that the mass of a galaxy would increase as we move away from its centre into the halo. However, much of these mass is not visible. Hence, the name 'Dark matter'. Though there are theories which speculate about the nature of dark matter, none of them is universally accepted as of now. Still, the ubiquitous presence of dark matter in spiral galaxies is well established[2], as it interacts gravitationally.

Let us consider one such galaxy with a flat rotation curve (implying $v_o(r) = v_c = constant$) which extends out to some radius R with a spherically symmetric mass distribution up to the boundary of halo region. Further, we assume that the mass density is zero beyond the radius R. The mass contained within a radius r ($r < R$) is given by the expression

$$M_{<r} = \int_0^r \rho(r')\, dV \qquad (5)$$

Here, the primed coordinate runs from 0 to r. For an object of mass m revolving with
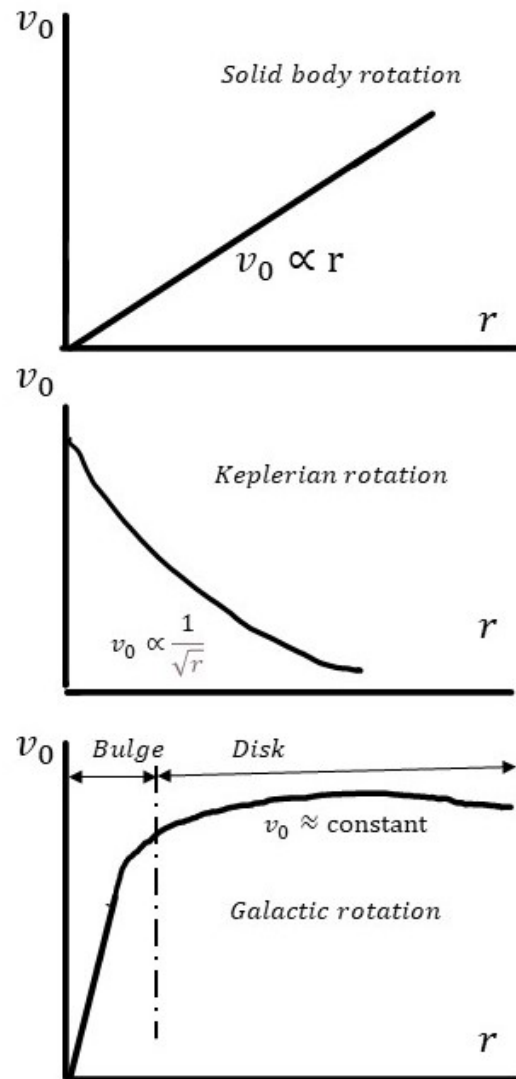

Figure 2: Types of rotation curves

the constant speed v_c at this distance r, using Equation (3), one can write

$$v_c = \sqrt{GM_{<r}/r} = constant \qquad (6)$$

which implies,

$$M_{<r} \propto r \quad Or, \quad \int_0^r \rho(r')\, dV \propto r \qquad (7)$$

Since, for a spherically symmetric distribution the volume element is $dV = 4\pi r'^2 dr'$,

Equation (7) holds true if $\rho(r') \propto 1/r'^2$[3] or $\rho(r') = k/r'^2$ (k being a constant of proportionality). Thus, Equation (5) now gives

$$M_{<r} = \int_0^r \frac{k}{r'^2} 4\pi r'^2 dr' = 4\pi kr \qquad (8)$$

And, the total mass of the galaxy is $M_{<R} = 4\pi kR$. Using Equation (8) in Equation (2), We get

$$v_c{}^2 = 4\pi kG \qquad (9)$$

3 An Expression for the escape velocity

In order to calculate the escape velocity of this object of mass m at a distance r from the galactic centre, we apply equation (2) with the limits over r' from r to ∞ Thus,

$$W = \int_r^\infty F \, dr' = \int_r^R \frac{GM_{<r'}m}{r'^2} \, dr' + \int_R^\infty \frac{GM_{<R}m}{r'^2} \, dr'$$

$$W = 4\pi kGm \left[\int_r^R \frac{1}{r'} \, dr' + R \int_R^\infty \frac{1}{r'^2} \, dr' \right]$$

Or, using Equation (9),

$$W = v_c{}^2 m[ln(\frac{R}{r}) + 1]$$

Now, Equating this work required with the minimum kinetic energy $\frac{1}{2}mv_e{}^2$ the body must have at r, in order to 'escape' from the galaxy's gravitational field, we get[4],

$$v_e{}^2 = 2v_c{}^2[1 + ln(\frac{R}{r})] \qquad (10)$$

Let us try to estimate a value for the escape velocity at the position of our Solar system in the Milky Way Galaxy by considering following parameters as given in [1]: The linear speed of rotation of the Sun around the Galactic centre is $v_c = 250 \, km/s$. The radius of Solar system's orbit is $r = 30,000 \, light \, years$ and, the extent of galactic boundary for the Milky Way goes up to $R = 90,000 \, light \, years$. Substituting these values into Equation (10), we find $v_e \approx 512 \, km/s$, which is within acceptable error bars with a latest value of the escape velocity as determined in[5].

Acknowledgments

References

[1] F. Shu The Physical Universe: An Introduction to Astronomy (University Science Books, 1982), pp. 258, 264

[2] E. Corbelli, P. Salucci. Mon. Not. R. Astron. Soc. 311, 441-447, 2000

[3] J. Bovy. http://www.astro.utoronto.ca/~bovy/teaching.html

[4] P. Armitage. https://jila.colorado.edu/~pja/astr3830/problem_set1.pdf

[5] G. Monari et.al. A&A 616, L9, 2018

Evaluation of Planck's Constant using Light Emitting Diodes: A new apporoach using ExpEYES-17

Durjoy Roy[1] and Bidisha Biswas[2]

[1]Department of Electronic Science, Rishi Bankim Chandra College, Naihati 743165, India.

`roy.durjoy@gmail.com`

[2]Department of Electronic Science, Bhairab Ganguly College, Kolkata 700056, India.

`bidisha005@gmail.com`

Submitted on 06-03-2021

Abstract

Teaching Laboratories in Physics and Electronic Science can be made more interesting place to be using data acquisition techniques and computer interfaced experiments. Using an indigenous tool ExpEYES-17, designed by the IUAC New Delhi, the authors have re-designed the experiment to evaluate Planck's constant using Light Emitting Diodes. This approach, using the programming language python, may find a new way of experimenting, parallel to the traditional hands on experiment approach. The results obtained are in excellent agreement with the standard value of the Planck's Constant h.

1   Introduction

Practical Laboratories in Under-graduate level Physics covers a huge number of experiments. In Electronic Science laboratories, some more experiments, tailor-made for the subject are performed in the laboratories. In the era of the CBCS curriculum, managing teaching laboratories became even more critical, as the number of papers to be studied in UG and PG level has increased drastically, and time constraints are put in to place. Conventionally the experiments involving electronics are carried out using power supplies, analogue or digital meters and oscilloscopes and other related accessories. 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 collection 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, primarily focussing on the experiments related to Particle Physics. ExpEYES-17, a microcontroller based data acquisition system, was designed by a research group in IUAC led by Dr. Ajith Kumar B.P, and the hardware design was published under the CERN-OHL[1]. CERN-OHL is an Open Hardware Licensing method devised by the CERN (*Conseil Europen pour la Recherche Nuclaire*, 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 Licence, popularly known as 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 softwares 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, and expansions. If somebody want to add or remove software and/or hardware features to this device, he can do it. The standard user interface of the software contains about fifty experiments that be performed by the students of various level. The authors, are 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 authors have re-designed a very popular experiment of graduate level, i.e., Evaluation of Planck's constant using Light Emitting Diodes with the ExpEYES-17 hardware as the experiment and data acquisition platform, and writing necessary software code in python.

The first author has published an article on ExpEYES-17 interfaced experiment on Field-Effect Transistor Characteristics[2] earlier, and can be consulted for more details of the device ExpEYES-17.

2 Introduction to ExpEYES-17

The IUAC has conceived the Phoenix Project[3] in 2004. The name of the project was the abbreviation of *Physics* with

*Ho*me-made *E*quipment & *I*nnovative *E*xperiments. ExpEYES-17 once again, is the abbreviation of *Exp*eriments for *Y*oung *E*ngineers and *S*cientists.

The EXPEyes-17 kit is built around a PIC24EP64GP204 Micro-controller, along with other necessary hardware that support the Micro-controller. A schematic of the complete hardware is available under the CERN-OHL at www.expeyes.in[4], and can be constructed by anybody who have moderate level of grasp over electronic circuits, and programming microcontrollers.

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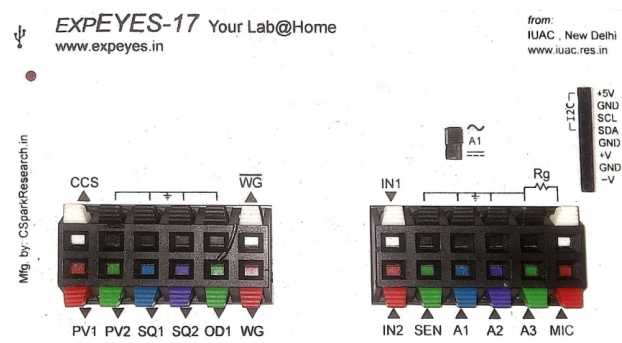
Figure 1: A view of the inputs & outputs of the ExpEYES-17 kit.

ExpEYES-17 kit consists of three main blocks, one is for the generator of various signals and voltages, another for the inputs of measurements and works like digital storage oscilloscope (DSO) and the third block is for the I^2C modules which will not be used here, and thus, will not be discussed here. Figure 1 shows the top end of the kit

designed by the IUAC.

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

Unlike a simulator, ExpEYES-17 kit is a hardware experiment and data acquisition system and can be used with real voltages and currents as the inputs or outputs. The kit is connected to a PC which provides the power to run the kit along with all its components under experiment through a USB 2.0 type A to Type Mini-B connector. The PC runs the ExpEYES-17 user software over Linux or Windows through the USB; though the user can choose to write and execute a programme in the terminal or python shell instead of using the provided software.

The boundary of working of the kit is limited by the USB parameters like data transfer limit of the USB 2.0 interface (480 Mbit/sec) and working voltage of maximum +5V DC and 500 mA DC current etc. Internal PCB design with the input/output terminals of ExpEYES-17 kit is shown in Figure 2.

In our experiment, we will deal with

the characteristics of an LED, and find the Planck's constant from them; the input voltage to the LED will be taken from the Output (Signal Generator) Block of the ExpEYES kit, and the voltage output of the LED will be fed to the Input (Oscilloscope & Sensors) Block of the kit.

2.2 Input (Signal Generator) Block to LED:

Table 1 derived from the first authors previous publication in Physics Education [2], shows the available DC voltage sources, current source, square, sine and triangular wave generators, the inverted generator with all their specifications. The waves are synthesized and using the GUI interface or python command in Linux terminal, we can control the outputs.

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

O/P	Description
CS	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 hen 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.

The microcontroller used in ExpEYES-17 operates at +3.3VDC with 0.3VDC tolerance. This can easily be generated from the +5VDC USB 2.0 connection provided by

the PC and supplied to the microcontroller.

+5VDC is supplied to PV1, and +3.3VDC is supplied to PV2, and both are through charge pumped DC-to-DC converters to generate the negative voltages as well, built around LM324 OP-AMPs. These voltage values are defined by Pulse Width Modulation from the microcontroller outputs, where variation of Duty Cycle in a rectangular wave train is translated to equivalent DC Voltage.

Here in our work, we only have used the PV1 output to bias the LED under test.

2.3 Output (Oscilloscope & Sensors) Block:

The Table 2 summarizes the available inputs of the kit, detailed out in [2].

Table 2: Summary of the inputs of the ExpEYES-17 which senses the voltage and current of the LED

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.

In this article, we will only be using the channel A1 for measuring the LED voltage. The input signal at A1 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, interested reader can consult [2] for more details.

The Input and the Output sections are interfaced with a set of push-to-release terminal blocks as shown in Figures 1 & 2, and the authors have used this connectors to connect the LEDs and the resistors required in this experiment.

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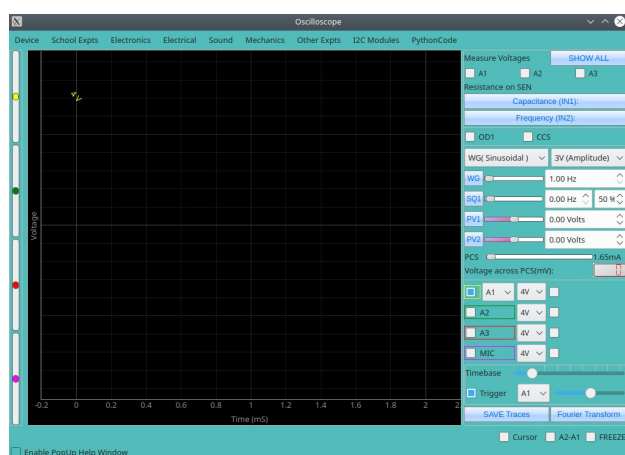
Figure 3: A view of the User Interface (Software) of the ExpEYES-17 kit.

The software of ExpEYES-17, designed in C and Python, controls the hardware. It dictates the microcontroller how to generate signals, interpret the inputs, and send it back to the PC. Here in our work we have used Python code to drive the kit by calling specific libraries to handle the hardware. The Users Manual for the kit is available under GNU-GPL at their website[9].

The interface between the device and the PC connected is developed using C. The Graphical User Interface (GUI) shown in Figure 3, designed with Python mostly comprises of an oscilloscope display . In our article, we will not be using the GUI, and the description about the GUI is skipped. However, interested reader can consult the article published in Physics Education [2].

3   Theoretical Background:

In this article, the authors have designed a program to find the Planck's constant using Light Emitting diodes. The theoretical background of the experiment is elaborated in this section.

3.1   Definition of Planck's Constant:

Max Karl Ernst Ludwig Planck (1858-1947) (Figure 4), Popularly Known as Max Planck, was a German theoretical physicist, who won Nobel Prize in Physics in 1918 for discovery of energy quanta.

Figure 4: Max Karl Ernst Ludwig Planck (1858-1947), in 1918. (Published from `https://www.nobelprize.org/prizes/physics/1918/planck/biographical/[5]`)

The Planck constant, or Planck's constant, is defined as the quantum of electromagnetic action that relates a photon's energy to its frequency. The Planck constant multiplied by a photon's frequency is equal to a photon's energy. The Planck constant is a very fundamental physical constant denoted as h, and of paramount importance in quantum mechanics.

The Planck's relation, also known as the Planck-Einstein relation, tells that the energy of a photon E is proportional to the frequency of the photon ν.

$$E \propto \nu \tag{1}$$

The constant of proportionality is known to as the Planck's constant h, and the equation becomes

$$E = h\nu \tag{2}$$

The Planck's Constant is defined to have a constant value of

$$h = 6.62607015 \times 10^{-34} J.s \tag{3}$$

LED's are the monochromatic incoherent light sources, and they can be used to find the Planck's constant due to their narrowband optical emission properties.

3.2 Working Principle of Light Emitting Diode

Light Emitting Diodes or LEDs are made of compound semiconductors. They follow the principle of Electroluminescence.

In LED, compound semiconductors are used to form a P-N junction diode. Materials such as Gallium Arsenide (GaAs), Indium Phosphide (InP) and others, with various trace doping elements, are used to generate different colors of light. A pn junction emits light when electric current flows through it. This phenomenon is known as

Electroluminescence. Electrons cross from the N-region and recombine with the holes existing in the P-region. Free electrons are in the conduction band of energy levels, while holes are in the valence energy band. Thus the energy level of the holes is lower than that of the electrons. Some amount of the energy thus generated must be dissipated to recombine the electrons and the holes. This energy is emitted in the form of light, and some amount of heat as well.
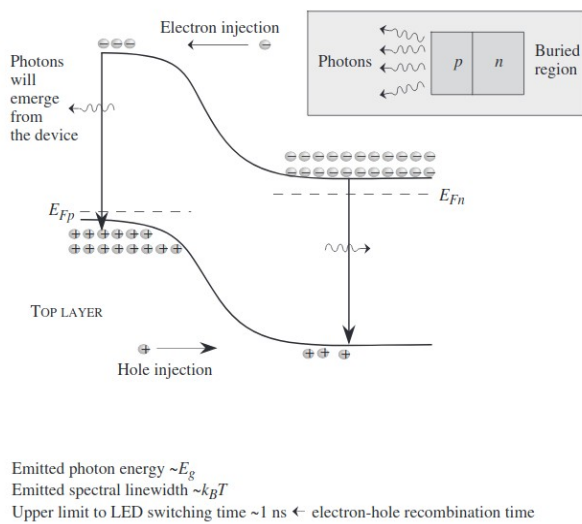
electrons and holes combine at the depletion layer (or active layer), light or photons are emitted or radiated in all directions. Figure 5 is showing the mechanism of light emission due to electron-hole pair combining on forward biasing. The explanation behind the emission of photons in an LED diode lies in the energy band theory of solids.

Emitted photon energy $\sim E_g$
Emitted spectral linewidth $\sim k_B T$
Upper limit to LED switching time ~ 1 ns $\leftarrow$ electron-hole recombination time

Figure 5: Light Emission from GaAs LED junction with band structure (Published from Jasprit Singh, Semiconductor Devices: Basic Principles, pp. 476, ISBN:-9788126511020, Willey India, 2009)[6])
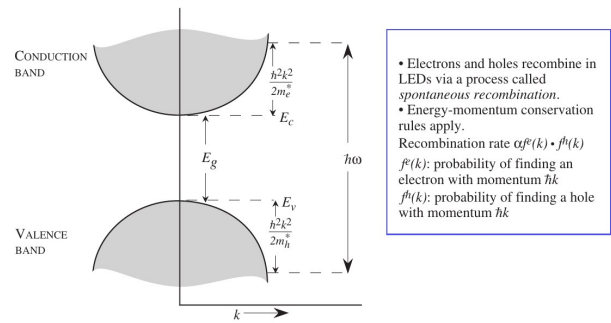

Figure 6: E-k diagram for the conduction and valence bands. Optical transitions are vertical (Published from Jasprit Singh, Semiconductor Devices: Basic Principles, pp. 477, ISBN:-9788126511020, Willey India, 2009)[6])

Like an ordinary diode, the LED diode works when it is forward biased. In this case, the n-type compond semiconductor is heavily doped than the p-type forming the p-n junction. When it is forward biased, the potential barrier gets reduced and the

The minimal-energy state in the conduction band and the maximal-energy state in the valence band are each characterized by a crystal momentum, or wave-vector, called the k-vector, in the Brillouin zone. If the k-vectors are different, the material has an indirect band gap. The band gap is called direct while the k-vevtor of electrons and holes is the same in both the conduction band and the valence band. An electron can directly emit a photon in a direct band gap material. In an indirect gap marerial, a photon cannot be emitted because the

electron must pass through an intermediate state and transfer momentum to the crystal lattice. Amorphous Silicons and Germaniums are indirect band gap materials, and they don't emit photons.

According to this theory, whether the electron-hole combining will give out photons or not depends on whether the material has a direct band gap or indirect band gap. In a direct bandgap material, the bottom of the energy level of conduction band lies directly above the topmost energy level of the valence band on the Energy vs Momentum wave vector k (E-k) diagram. The E-k diagram of a GaAs diode is shown in figure 6. When electrons and holes recombine, a quantum of energy $E = h\nu$ corresponding to the energy gap (eV) is escaped in the form of light energy or photons where h is the Plancks constant and ν is the frequency of light.

Figure 7 shows the schematic of a typical dome type LED, as used in this experiment. The LED is actually on the very small semiconductor die, sat upon the LED frame. Anode and Cathode leads are used to bias the LED.

3.3 Method of finding the Planck's Constant using LED:

When forward biased, depending upon the material composition, forward current starts flowing after a certain value V_{Th},
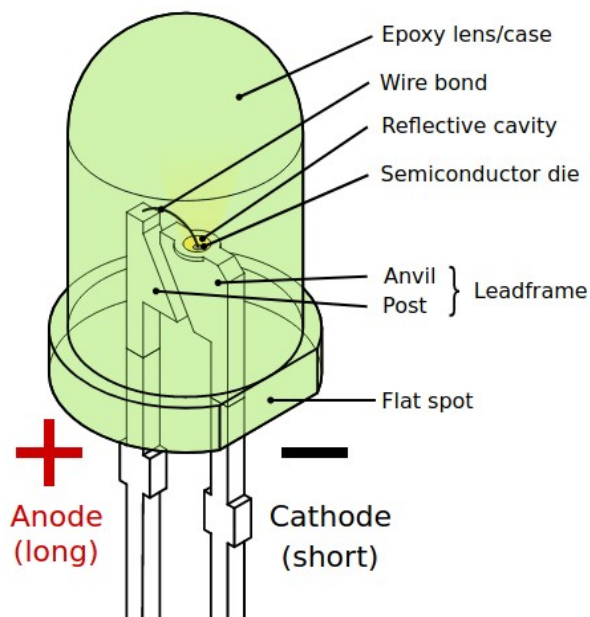
Figure 7: Typical Structure of Dome type LED. The actual LED is on the semiconductor die, and covered with translucent or transparent epoxy resin lens (Published from Wikipedia `https://commons.wikimedia. org/w/index.php?curid=6431789`)[7])

called the Threshold Voltage, of the LED voltage V_D. This voltage is directly related to the Enery band gap E_g of the particular LED, and the relation is given by

$$E_g = e \times V_{Th} \qquad (4)$$

where e=electronic charge.

The light is emitted by LED is due to emission of photon, and the relation between the photon energy and its frequency, ν, can be written from equation (2) as

$$E_g = h\nu \tag{5}$$

From equations (4) and (5), one can write

$$eV_{Th} = h\nu \tag{6}$$

The relation between the wavelength λ and frequency ν of any electromagnetic wave is given by

$$\nu = \frac{c}{\lambda} \tag{7}$$

where c= velocity of light in free space.

Thus, from equations(6) and (7) we get

$$eV_{Th} = h\frac{c}{\lambda} \tag{8}$$

From the above equation, the value of the Planck's constant can be calculated as

$$h = eV_{Th}\frac{\lambda}{c} \tag{9}$$

In the above equation, e and c are known constants. This equation, i.e., equation (9) provides the working formula of our experiment. The value of λ depends upon the colour of the LED, and V_{Th} can be measured by writing a python program and running it on the ExpEYES-17 kit.

4　Experiment Set-up:

The experiment setup included four LEDs, the ExpEYES kit, and a Computer running Ubuntu Linux. The Hardware and software of the experiment are elaborated below.

4.1　Hardware, or Circuit for the experiment:

Three LEDs of primary colours, i.e., Red, Green and Blue, and one having Infrared emission were used in the experiment. More colour can always be used by simply adding the color name and corresponding wavelength values in the program. Spectrometer is not used here, as the LEDs are monochromatic incoherent light sources, i.e., the light emitted by an LED provides pretty narrowband illumination. They are not as narrowband as lasers, but the optical bandwidth of an LED is quite workable without a spectrometer. The values of the wavelength are taken from the data-sheets available at the manufacturers website.[12]

A 100Ω resistance is connected in series with the LED. The primary goal of the resistor is to create a voltage drop, rather than to limit the LED current, as the python program that runs the experiment is designed to drive the LED for a very short period of time. ExpEYES-17 can not measure current directly. So, the 100Ω resistance creates a voltage drop, and that drop is used to calculate the current passing through the resistaor, and the LED. Figure 8 shows the circuit schematic for the experiment.

The Anode of the LED, via the 100Ω resistor, is connected to the programmable voltage source PV1 of the ExpEYEs. The Cathode is directly grounded. The Voltage accross the diode is measured at the

Figure 9: Actual Experiment set-up. The black crocodile clip is biting the junction of the resistor and the LED anode to connect the intersection to A1. The other end of the resistors are connected to PV1 (Red).
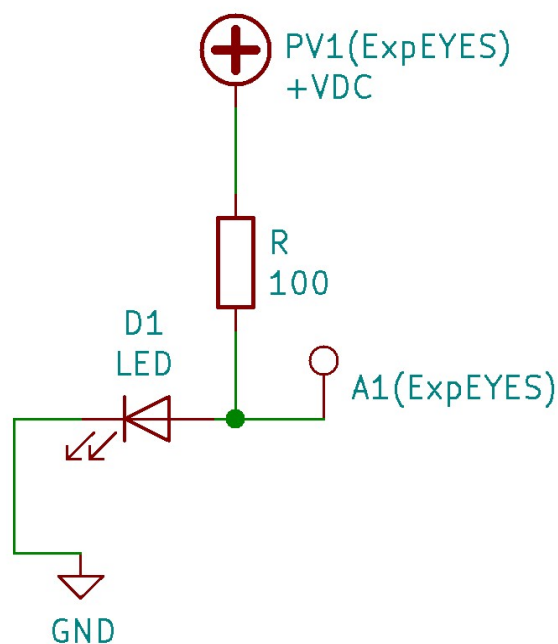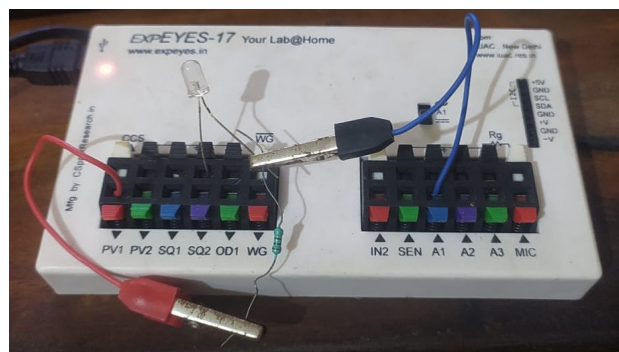
Figure 8: Circuit Schematic of the experiment setup

junction of the Anode and the resistor, by connecting the point to A1, with reference to the ground. A schematie of the actual circuit used in the experiment set-up is shown in Figure 9.

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 voltage drop accross the LED coming in through the channel A1. However, there is an option in the software where one can use python code directly.

The above circuit is implemented on the ExpEYES-17 kit as shown in figure 9. Breadboard is not used here, as the termi-

nals of the kit were capable enough to hold the circuit.

4.2 Software for the experiment:

The software, or the program of this experiment using ExpEYES-17 is written in Python. The authors have written the codes for LED I-V characteristics. In 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.

4.2.1 LED I-V Characteristics:

In this program, PV1 is varied from 0 to + 5.0 volt to provide forward bias to the LED in a while loop. The diode current I_D is found by dividing the voltage drop across the 100 Ω resistor by the expression

$\mathtt{id} = (\mathtt{PV1} - \mathtt{A1}) * 1000/100,$ where the factor 1000 is used to convert the current to milli amperes. The data for Diode Currents and Diode voltages are stored in two lists `ida[]` and `vda[]` respectively, and plotted using `matplotlib` library of python.

The Authors have kept an option to choose LED color for experiment. The program, when executing, asks for input to select the Color by entering one of the letters R, G, B, or I to select Red, Green, Blue or Infrared Color respectively. Upon receiving the Colour from the user, the program selects the corresponding standard value of λ, the colour of plot, and the text that is to printed on the plot.

At the end of the loop, the LED is switched off quickly by assigning 0 to PV1 so as to protect the LED from unnatural death, as the resistor used here is not primarily focussed to limit the current, but to calculate the diode current.

4.2.2 Marking of V_{Th}:

There are many a convention to determine the value of the Threshold Voltage V_{Th}. The authors have chosen the method in which the very start of the forward current is selected by comparing the current values with zero. Then from the list index of the current I_{Th}, The voltage V_{Th} is searched out of the list of the voltage values.

4.2.3 Calculation of the Planck'c Constant h:

At the last part of the program, the value of Planck'c Constant h is determined by the program, and printed on the screen. The values are found using the equation(9). As the LEDs are monochromatic light source, the standard λ values for corresponding color of LED are assigned to a variable in the program, and used depending upon the choice of color selected at the start of the program.

4.2.4 The Python Program:

```
1 #Python Program to plot LED I-V Characteristics and to calculate
     Planck's Constant
2
3 #Initialization of device
4 import eyes17.eyes
5 p = eyes17.eyes.open()
6 from pylab import *
7 #import libraries
```

```python
 8 import time
 9 import matplotlib.pyplot as plt
10
11 #Get the Color of the LED being used
12 print('Select the Colour of the LED being used')
13 col=input ("Enter R for RED, G for Green, B for Blue, or I for
      Infrared, and press ENTER:")
14 if (col=='R'):
15     lambin = 650 #wavelength in nm
16     colname='red'
17     plotcol=colname
18 elif (col=='G'):
19     lambin = 560 #wavelength in nm
20     colname='green'
21     plotcol=colname
22 elif (col=='B'):
23     lambin = 470 #wavelength in nm
24     colname='blue'
25     plotcol=colname
26 elif (col=='I'):
27     lambin = 940 #wavelength in nm
28     colname='infrared'
29     plotcol='black'
30 else:
31     print('LED Colour is not defined')
32     exit()
33
34 lamb = lambin * pow(10, -9 )#convert wavelentgh to SI unit
35
36 #Plot Title and labels
37 plt.title('Forward I-V Characteristics of LED')#Plot Title
38 plt.xlabel('LED Voltage Vd (Volt)') #x axis label
39 plt.ylabel('LED Current Id (MilliAmp)') #y axis label
40 plt.text(0.1,1.5, 'LED color='+str(colname)) #LED Color
41 plt.text(0.1,0.5, 'Wavelength (nm)='+str(lambin)) #LED wavelength
42 #values of constants
43 e = 1.60217662 * pow(10, -19) #value of electronic charge in SI system
44 c = 2.99792458 * pow(10, 8) #value of speed of light in SI system
45
```

```python
46  #Finding the data for plot and calculation
47  pv1=0.0
48  vda = [] #initiate LED voltage list
49  ida = [] #initiate LED current list
50  ion()#enable interactive plotting
51  print ('vs','vd', 'id', sep='\t') #print header
52  while (pv1 <= 5.00):
53      p.set_pv1(pv1) #set source voltage
54      vs = pv1 # check source voltage
55      vd = p.get_voltage('A1') # get LED voltage
56      id = (vs-vd)*1000/100 #find LED current in mA for r=100 Ohm
57      vda.append(vd) #append LED voltage in list
58      ida.append(id) #append LED current in list
59      print (vs,vd, id, sep='\t') #print data
60      pause(0.2)#insert delay to give the microcontroller a breathing
        time
61      pv1 += 0.2 #next value of source voltage
62  p.set_pv1(0.0) #Switch off the LED
63
64  #Finding the value of vth from the lists
65  ith = list(filter(lambda i: i > 0.00, ida))[0] #Find the LED Current
        when it becomes more than zero
66  ind=ida.index(ith) #Find the index of the threshold current in the
        list
67  vth = vda[ind] #Find the threshold voltage corresponding to ith
68  print('') #print blank line
69  #print the value of the elements for calculation
70  print('LED wavelength :-', lambin, 'Nanometer')
71  print('Measured Value of Threshold Voltage for', colname, 'LED is:-',
        vth, 'volt')
72  print('Value of Electronic Charge', e, 'coulombs')
73  print('Value of Speed of Light', c, 'meter/seconds')
74  print('') #print blank line
75
76  #calculating the value of Planck's constants
77  h = (e*vth*lamb)/c
78  print('Value of Plancks Constant for ', colname, ' LED is:-', h,
        'J.s')
79
```

```
80  #plotting the graph
81
82  plt.plot(vda, ida, color= plotcol) #plot graph
83  plt.grid(True) #show gridlines
84  plt.plot(vth,ith, 'ko') #add a black marker at the Vth
85  plt.annotate("      Vth", (vth, ith)) # Mark Vth
86  plt.show() #show graph
```

4.2.5 Program Output:

The output of the program shows two voltage values, namely the sourche voltage v_s, the diode voltage v_d and one current value, the diode current i_d. After displaying all the values, it displays the Calculated value of the Planck's constant for the color of the LED used. A screenshot of the output shown for Red LED in the figure 10.
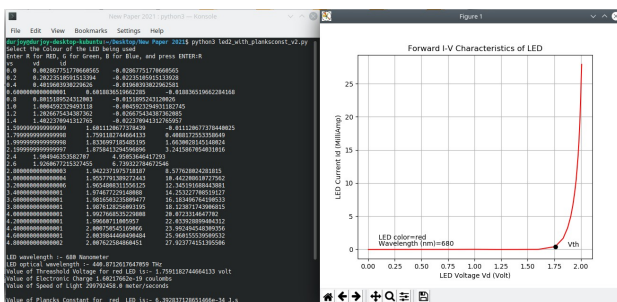
Figure 10: Screenshot of typical output for RED LED

The authors have designed the program to calculate the h values for each of the LED. The following section elaborates the values obtained running the python code.

5 Results:

5.1 Direct Calculation from the LED forward characteristics:

The results are obtained as plots as shown, in the Figures 11, 12, 13 and 14 for Red, Green, Blue and Infrared LEDs respectively.
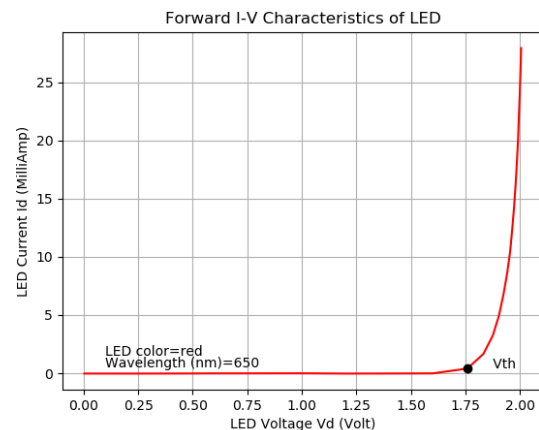
Figure 11: Forward Characteristics for Red LED ($\lambda = 650nm$)

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 to plot the curves.
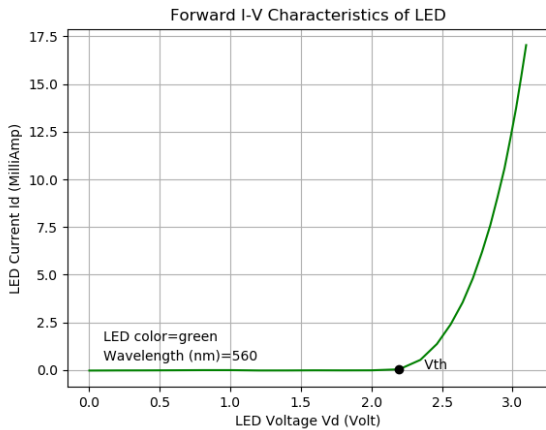
Figure 12: Forward Characteristics for Green LED ($\lambda = 560nm$)

The program calculates the value of the Planck's constant using equation (9), and prints the result as well. A screenshot of the actual program output, with a value of $\lambda = 680nm$ is shown in figure 10.
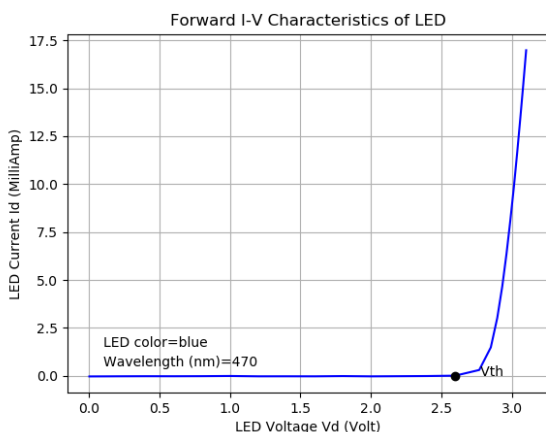

Figure 13: Forward Characteristics for Blue LED ($\lambda = 470nm$)

It is clearly visible from the plots that these Light Emitting Diodes are obeying the diode characteristics we generally see with normal PN junction diodes. In cases of LEDs, however, the threshold voltages are different.
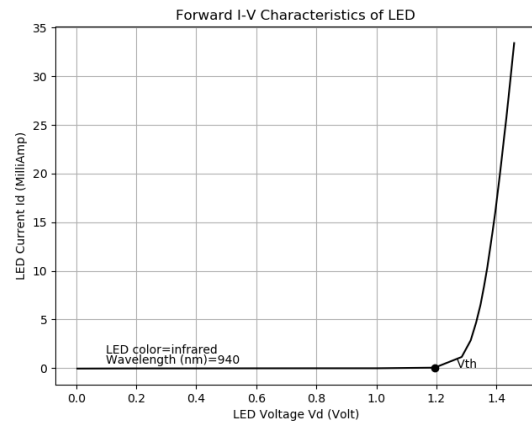
Figure 14: Forward Characteristics for Infrared LED ($\lambda = 940nm$)

Except the Infrared LED, all the LEDs used in this experiment are in visible range of light. During the experiment, the visible LEDs emit light with increasing intensities. Nothing is visible when the Infrared LED is put under test.

The LEDs are typically more fragile than their PN junction ordinary diode counterparts while reverse biased. In this article, reverse characteristics are not required, and thus, not plotted.

A screenshot of the Program Output is given in figure 15. An Infrared LED is used in this case.

Table 3 lists the values of h obtained from the program. These values will be used to find an average value of the Planck's Constant and to compare that with the actual value.

Table 3: Experimental Data

Obs. No.	Colour of LED	Wavelength (λ) in nm	Value of Planck's Constant h in $J.s$
1	Infrared	940	$6.000208322582497 \times 10^{-34}$
2	Red	650	$6.112555479157524 \times 10^{-34}$
3	Green	560	$6.573997191749779 \times 10^{-34}$
4	Blue	470	$6.527765197052901 \times 10^{-34}$
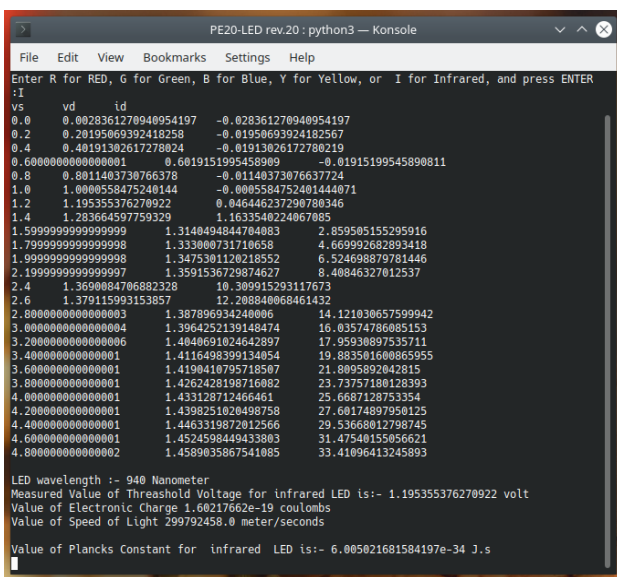
Figure 15: Program output only for Infrared LED ($\lambda = 940nm$)

5.2 Calculation of Average Value & $\%error$:

From the Table (3), it is found that the average value of the Planck's Constant h, calculated directly from the program is $6.20249047262 \times 10^{-34} J.s$.

The standard value of the Planck's constant is $h = 6.62607015 \times 10^{-34} J.s$. The following section calculates the amount of error in average of the measured values, with reference to the theoretical value. The percentage error $\%error$ is given by

$$\%error = \left| \frac{h_{experimental} - h_{theoretical}}{h_{theoretical}} \right| \times 100\% \tag{10}$$

Putting the average value of the Planck's constant in the equation(10), we have

$$\%error = \left| \frac{6.20249047 - 6.62607015}{6.62607015} \right| \times 100\%$$

$$\%error = \left| \frac{-0.42357968}{6.62607015} \right| \times 100\%$$

$$\%error = 6.39\% \tag{11}$$

From the calculated value of the Planck's Constant h from the table (3), it is found that the error margin (calculated in eq 11) is about 6.4%, which is in excellent agreement with the standard value.

6 Discussions

The experiment, when performed in a conventional way, includes steps like noting down the readings manually, and drawing the plots. Here, using the ExpEYES-17 kit, and the python programming lanugage, 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 very small amount of current as it draws the required power through the USB connection, and the current generation is limited by the USB 2.0. One more downside of this experiment is that the authors did not use any spectrometer to measure the actual wavelength of the LED lights. There are some amount of incoherence in the LEDs, and thus, the LEDs are not truly monochromatic. Authors have used the standard values from the manufacturers data-sheets for λ available at [12].

Conventionally, in this experiment when performed in Laboratory, the Threshhold voltage of an LED is measured just when the LED starts glowing. As the sensitivity of human eyes varies from person to person, The value of V_{Th} may be read differently. To minimize the ambiguity, the LED currents are measured by the program at the point the current starts flowing.

There is another convention of measuring the threshold voltage from the plot, is by drawing a tangent over the graph, and take the value of V_{Th} at the point the tangent intercepts the V_D axis. This can be done in python, but the complexity of curve fitting in the python program is out of the scope of discussion in this article.

The authors opine that 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

References

[1] CERN Open Hardware Licence `https://ohwr.org/cernohl`

[2] Durjoy Roy, "Junction Field Effect Transistor Characteristics: A new approach using ExpEYES-17", Physics Education, ISSN 0970-5953, Vol. 36, No. 3, April-June, 2020

[3] IUAC Phoenix Project `https://www.iuac.res.in/phoenix/`

[4] ExpEYES-17 Website `http://www.expeyes.in`

[5] MAx Planck Biography, Nobel Prize Website `https://www.nobelprize.org/prizes/physics/1918/planck/biographical/`

[6] Jasprit Singh, "Semiconductor Devices: Basic Principles", pp. 476-477, ISBN 9788126511020, , Willey India, 2009

[7] Wikipedia `https://commons.wikimedia.org/w/index.php?curid=643178`

[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

[12] LED Datasheets `https://www.vishay.com/leds/`

Suggestive pedagogy for learning quantum confinement effects through online simulator Quantum Dot Lab

Amit Kumar

Department of Electronics
Bhaskaracharya College of Applied Sciences, University of Delhi
South West Delhi 110075, India.

Abstract

Nanotechnology is prerequisite for senior school, science, and engineering courses throughout the world. The basic concepts, however, are difficult for young students to comprehend. As a complement to current teaching-learning procedure, we present a pedagogical approach using the online simulator Quantum Dot Lab. In this paper, we show students how they can visulize wavefunctions, energy state diagram, and absorption spectra for typical ZnS quantum dots while learning quantum confinement effects. It is elaborated that size, shape, and temperature dependencies get explained in an appropriate educational design. In addition, the two dimensional quantum harmonic oscillator is discussed through 2D plotting demonstrations of typical wavefunctions.

1.   Introduction

The study of quantum physics is crucial to understanding the mechanisms involved in the applications of nano-scale low dimensional materials. The recent development of precision fabrication techniques have allowed low-dimensional structures to be fabricated. An entirely new field of technology has been enabled by this, known as nanotechnology. Nanotechnology is not limited to physics, but consists of physics, mathematics, biology, engineering, design, and computation. The ever-growing knowledge in this area, especially during the last decade, has made quantum physics compulsory for senior secondary school students or for students entering undergraduate courses across the globe [1]. The primer courses generally incorporate mathematical derivations with applications from nearly all areas of sciences and engineering.

However, studies show a lack of understanding among young students [2]. The factors that contributed to this are possible inefficient pedagogies during teaching/learning, a lack of experiments to supplement theoretical concepts, poor visualization tools during learning procedures, a lack of direct experience with quantum mechanics in daily life, *etc.* It is essential that learners be provided with proper visualization in order to understand basic concepts.

This paper proposes a pedagogy that will help students gain a basic understanding of the physics of quantum dots and nanostructures. The quantum world of quantum dots can be easily accessed even at room temperature, which makes them a great laboratory for exploring quantum physics. Therefore, quantum dot simulations could serve as a useful tool in explaining quantum mechanics. It

is challenging for elementary school students to conceptualize quantum physics as it describes particles (duals of waves) mechanics in terms of probabilistic approaches which contrast with our everyday experiences. The best way to visualize is to plot in three-dimensional (3D) regions. That is missing from the curriculum of higher secondary or lower undergraduate levels. The plotted 3D curves are provided in reference texts, especially as atomic orbitals of the H-atom, but they are not linked to theoretical concepts. If a little time is dedicated to plotting the 3D plots of quantum probabilistic curves, a more thorough understanding could be gained.

1. Theoretical background

In most preliminary quantum physics syllabuses, we discuss a particle inside a one-dimensional box to introduce quantum mechanics. The basic mechanics' problem involves finding out wave functions and eigen energy values for particle by solving the Schrodinger's wave equation [3]. However, for the best understanding, it is necessary to translate this notion to the three-dimensional box, where the particle has three coordinate freedoms. Let us consider the case where a particle with mass m restricted to move in a three-dimensional box with dimesions (L_x, L_y, L_z) in (x, y, z) axes, respectively. Typical dimensions are approximately 5-10 nm. The basics of the subject involve obtaining the normalized wavefunction, ψ(x, y, z) and energy eigen values, E_{n_x, n_y, n_z}. It is either possible to derive this or to simply extrapolate it from the 1D problem and write it as equation (1) and (2).

$$\psi(x, y, z) = \sqrt{\frac{8}{L_x\, L_y\, L_z}} \, sin\left(\frac{n_x\pi}{L_x}x\right) sin\left(\frac{n_y\pi}{L_y}y\right) sin\left(\frac{n_z\pi}{L_z}z\right)$$

$$(1)$$

$$E_{n_x, n_y, n_z} = \frac{\hbar^2\pi^2}{2m}\left\{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2\right\}$$

$$(2)$$

where $< n_x, n_y, n_z >$ set of quantum numbers corresponding to (x, y, z) directions respectively, that is defined to initiate from <1, 1, 1>. Let the box be a cuboid, $L_x = L_y = L_z = L$, then in this case, equation (2) can be written as follows

$$E_{n_x, n_y, n_z} = k(n_x{}^2 + n_y{}^2 + n_z{}^2)$$

$$(3)$$

where, $k = \frac{\hbar^2\pi^2}{2mL^2}$, is a constant for the given particle in the box. There are the different set of values of $<n_x, n_y, n_z>$ that gives same value of energy E. The term 'degenerate state' describes these conditions *i.e.* states having the same value of energy, *e.g.* $<2, 1, 1>, <1, 2, 1>,$ and $<1, 1, 2>$ all have $E = 6k$, so they represent the three degenerate states.

The next step is to visualize these equations through 3D plots. The modulus of wavefunction, $|\psi| = \sqrt{\psi\,\psi^*}$, offers a meaningful interpretation about probability density (*i.e.* the regions for finding particle).

2. Computational details

The plotting of 3D curves, as described above, has become so easy these days that even a smartphone could be used for this. The young learners could do this even if they had no experience in computer programming. In this work, we propose to obtain a 3D structure plot for $|\psi| = \sqrt{\psi\,\psi^*}$ using an online software tool, *Quantum Dot Lab,* that gives geometrical solutions to Schrodinger's equation for a particle confined in a nanometric cube. This tool is easily used for obtaining simulations of quantum dots of different sizes by tuning the material properties and the dimensions. It is integrated with the NanoHub, the simulation and modeling platform operated by Network for Computational

Nanotechnology (NCN) [4]. Generally, optical absorption spectrum detection for quantum dots of different shapes, materials of different effective masses is performed by this technique. A student would find this a useful resource for learning fundamentals of quantum dot physics through the study of confined electronic systems. In the present study, we demonstrate how to plot the orbitals of nontoxic ZnS quantum dots on a cubic-shaped box with *Quantum Dot Lab*. The application was tested on mobile devices and supported, so it could easily be integrated into classroom lectures.

3. Results and discussions

The young learners struggle with understanding spatial quantum confinement, but are well-versed in classical mechanics. The reason for this is that children generally see classical object motion in daily life at a young age. It is difficult to conceptualize the quantum mechanical world at the subatomic level at first. It has been observed that plotting 3D plots is an excellent tool in helping to develop imagination. *Quantum Dot Lab* provides an intuitive way to engage students during a class, so it is suggested that these can be used as pedagogical tools. In order to compute the eigenstates of the system, this tool solves Schrödinger's equation with the Lanczos method [5].

The II-VI group material ZnS is used in optoelectronic applications. This compound semiconductor has a wide bandgap of 3.68 eV, an effective mass of electron $0.34m_0$ with a lattice constant of 0.541 nm [6]. By setting these structure parameters, *the Quantum Dot Lab* simulator is used to obtain the permitted orbital shapes for electrons in the cuboid quantum dot of dimensions $10 \times 10 \times 10 \ nm^3$ for the first 8 energy states as shown in figure 1. The blue-colored zones point to the regions with the highest probability to find particle (electron), while the red-colored zones point to those with the lowest probability. This

spatial variation of particle cloud allows students to gain an insight into how particles are spatially quantized in quantum mechanics. The plots they see in the text are generally similar to the self-plotting exercises. It help them gain a better understanding of difficult concepts and jargon, such as quantum spatial confinement, probability distributions, electron clouds and delocalization.
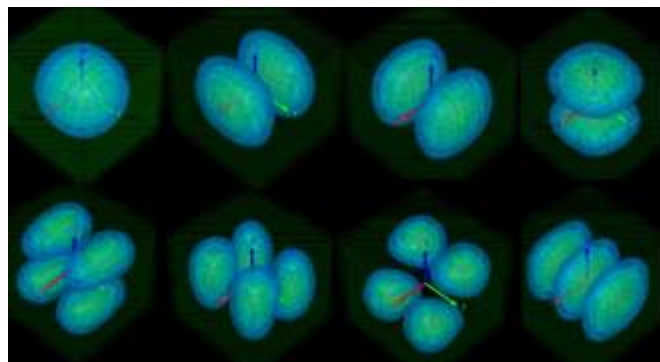

FIG 1. 3D plots for $|\psi| = \sqrt{\psi \psi^*}$. First 8 states, for quantum number set $<n_x, n_y, n_z>$: $<1,1,1>$, $<2,1,1>$, $<1,2,1>$, $<1,1,2>$, $<1,2,2>$, $<2,2,1>$, $<2,1,2>$, $<1,3,1>$ respectively, for cuboid shaped quantum dot of ZnS semiconductor.
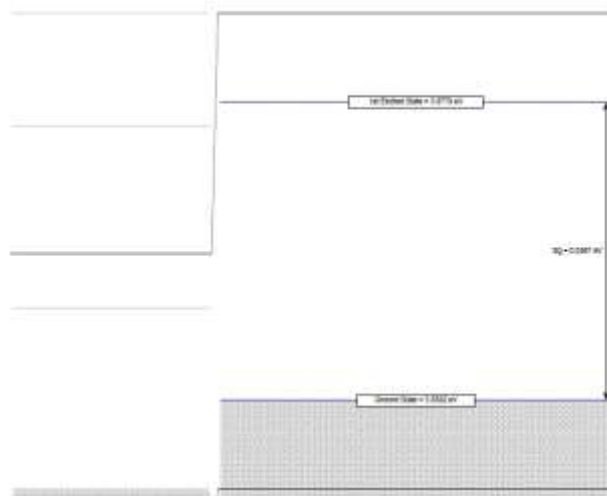

FIG 2. Energy state diagram for cuboid of ZnS having dimension $10 \times 10 \times 10 \ nm^3$.

It can be challenging to convey discretization of energy bands and students tend to be more curious

about it at first. It is felt that they do not quite understand why energy bands in solids are discretized at low dimensions. The energy state diagram obtained through simulation can be seen in figure 2. This implies that the ZnS cuboid quantum dots have a ground state energy level of 3.7192 eV and a first excited state of 3.7579 eV, resulting in an energy band gap of 0.0387 eV. It shows that as the dimensions of the bulk material are reduced, the band gap energy also changes. Hence, easy tailoring the bandgap could be accomplished using the quantum dots' size. For instance, if the energy band diagram for a cuboid of dimesions $5 \times 5 \times 5 \ nm^3$ is simulated then ground state energy increases to 3.8550 eV and the first excited state to 4.0211 eV. This makes the quantum dot bandgap increased to 0.1661 eV as compared to that cuboid of $10 \times 10 \times 10 \ nm^3$. If we simulate for $20 \times 20 \times 20 \ nm^3$ cuboid, the ground state energy becomes 3.6888 eV and the first excited state at 3.6975 eV making bandgap 0.0087 eV. Students may be able to understand how quantum confinement effect increase the energy bandgap for quantum dots (Quantum size effect). This is why quantum dots find application in optical devices. In addition, this illustrates the concept of energy dispersion due to the size dispersion of quantum dots. Since quantum dots of different sizes have different energy level diagrams, each energy level gets dispersed as a result of size variation, as is the case in real samples.

The absorption spectrum for quantum dots provides very useful information about optical devices. The prior simulations helped to understand the overall behavior of optical devices such as solar cells. These quantum dots absorb photons, but since the energy of the electrons is discrete, hence the absorption will also take place in a discrete manner. Thus, only photons with energies that are in accordance with state energy differences will be absorbed by quantum dots. This property of quantum dots is used to design LEDs,

LASERs, photodetectors, and solar cells. The quantum dot should have pre-state occupied and excited state empty for absorption when light with proper transition energy is illuminated. The quantum dot lab could be accessed for the measurement of optical transition rates and absorption strengths. This is done by shining the light source on the quantum dot with the setting of two angles **P** and **ϕ** as depicted in figure 3.
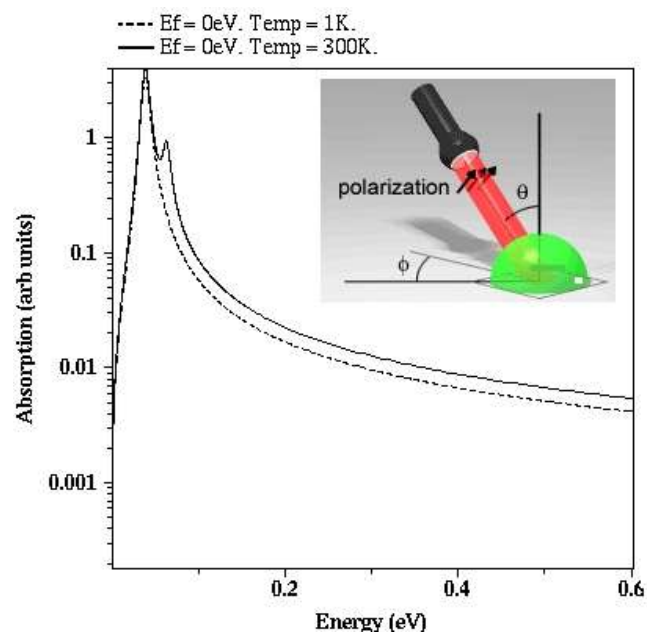

FIG 3. Absorption spectrum for polarized light at 1 K and 300 K for ZnS cuboid quantum dot of size $10 \times 10 \times 10 \ nm^3$.

The optical parameters are set to **P**=45°, **ϕ**=0°, a fermi level 0 eV for electron, a temperature 300 K, and state broadening 0.01 eV. The absolute Fermi level is set as NO. This makes the Fermi level relative to the lowest eigenstate, otherwise this level would have been absolute. A sweep parameter is taken as temperature with a minimum of 1 K and a maximum of 300 K. Figure 3 shows the simulated temperature dependence of the quantum dot's absorption spectrum.

It is observed that at low temperatures when electrons are at ground state, only the first peak

exists at about the bandgap energy, ΔE = 0.04 eV. As the temperature increased to room temperature, yet another peak appeared at around 0.06 eV, but its strength was low. Typically, such peaks are observed in semiconductor materials when absorbing a photon of higher energy than the bandgap at low temperatures (< 30 K) [7]. In quantum dots, however, such peaks can even be detected at room temperature. This contrast is due to the quantum confinement of charge carriers even at room temperature. This causes electron-hole pair (known as exciton) to be bound together by Coulombic force and act as atoms in free space. The presence of these excitonic peaks in the absorption spectrum of a sample can be considered as confirmation of quantum confinement. The number of excitonic peaks increases when there are more closely spaced energy states available near the first excitation level. If the dimensions of a quantum dot are taken to be $10 \times 8 \times 5 \ nm^3$, then there will be nearly three energy states at the initial level of excitation. From the geometrical interpretations *i.e.* inverse order to space dimensions, it is expected that p$_x$-orbital corresponding to quantum number set <2, 1, 1> will have lower energy than p$_y$, <1, 2, 1> and then p$_z$, <1, 1, 2> orbitals (s-orbital being ground state, <1, 1, 1>). In figure 4, three excitonic peaks can be seen in the absorption spectrum as a result.

It is also important to note that quantum confinement effects depend on the shape of the quantum dots. Figure 5 illustrates the absorption excitonic peaks of dome, cone, pyramid, cuboid, and cylindrical-shaped ZnS quantum dots having dimesions, $10 \times 8 \times 5 \ nm^3$. Consequently, this indicates that the dispersion of energy in quantum dots is closely related to their shape. The surface area of the quantum dots influences the intensity of the absorption peaks, higher the surface area more will be the absorption. Among the five shapes

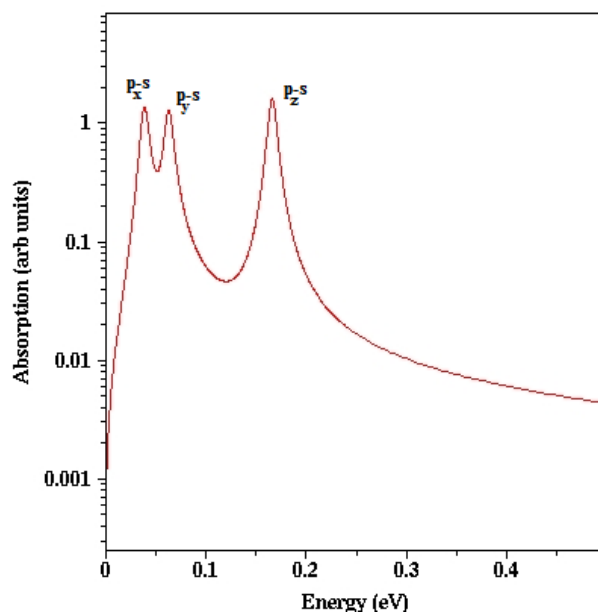listed, cuboid shows the highest absorption, while the cone shows the lowest.

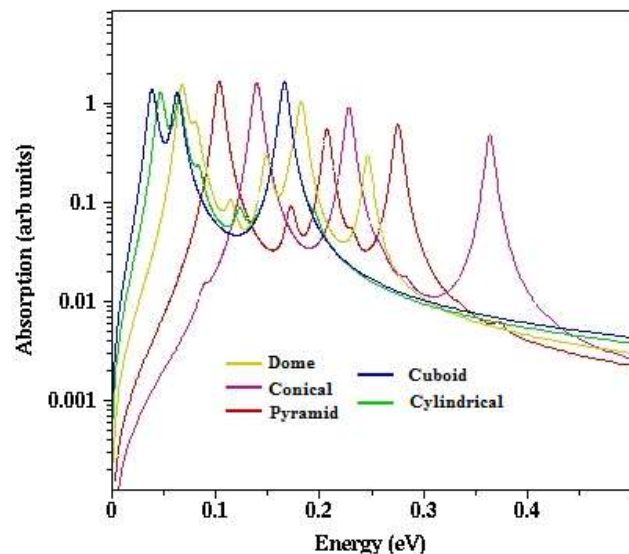FIG 4. Absorption spectrum for polarized light at 300 K for ZnS quantum dot of size $10 \times 8 \times 5 \ nm^3$.

FIG 5. Absorption spectrum for polarized light at 300 K for different shaped ZnS quantum dot of size $10 \times 8 \times 5 \ nm^3$

4. Simple Harmonic Oscillator

The simple harmonic oscillator is another example that is discussed at the beginning of our quantum physics education as a one-dimension problem. Schrodinger's solutions for the quantum harmonic oscillator include Hermite polynomials $H_n(x)$ multiplied by a Gaussian function [8]. A particle executing one-dimensional harmonic oscillations can be extended to two-dimensional wavefunctions as given by equation (4).

$$\psi(x,y) = N_{n_x} N_{n_y} \, e^{-(\alpha_x^2 x^2 + \alpha_y^2 y^2)/2} \, H_{n_x}(\alpha_x x) \, H_{n_y}(\alpha_y y)$$
(4)

where

$$N_{n_x} N_{n_y} = \left(\frac{\alpha_x \alpha_y}{2^{(n_x+n_y)} n_x! \, n_y! \, \pi} \right)^{1/2} \text{and } \alpha_x = \frac{m\omega_x}{\hbar}, \alpha_y = \frac{m\omega_y}{\hbar}$$

A set of two quantum numbers $<n_x, \, n_y>$ quantizes the energy eigenvalues, as shown in equation (5).

$$E_{n_x n_y} = \left(n_x + \tfrac{1}{2}\right)\hbar\omega_x + \left(n_y + \tfrac{1}{2}\right)\hbar\omega_y$$
(5)

The simple case when $\omega_x = \omega_y = \omega$, has degeneracy in the energy states, for different sets of quantum numbers. For example, $< 2, \, 0 >, < 0, \, 2 >, and \ < 1, \, 1 >$ have the same energy

$$E = E_{20} = E_{02} = E_{11} = 3\,\hbar\omega.$$

A simple plot of the probabilistic cloud for particles executing harmonic oscillations contains the essential information for learners. The main problem is that classical oscillatory behavior (for example, the motion of a pendulum) cannot be easily realized by equation (4). The 2D plots for $\psi(x,y)$ however, add value to equation (4) when learners become aware of the 2-dimensional oscillatory behavior of particles. The free online simulation exercise available at https://www.st-andrews.ac.uk/ [9] can be used to obtain such 2D plots as shown in figure 6 for the case when $E = 3\hbar\omega$.
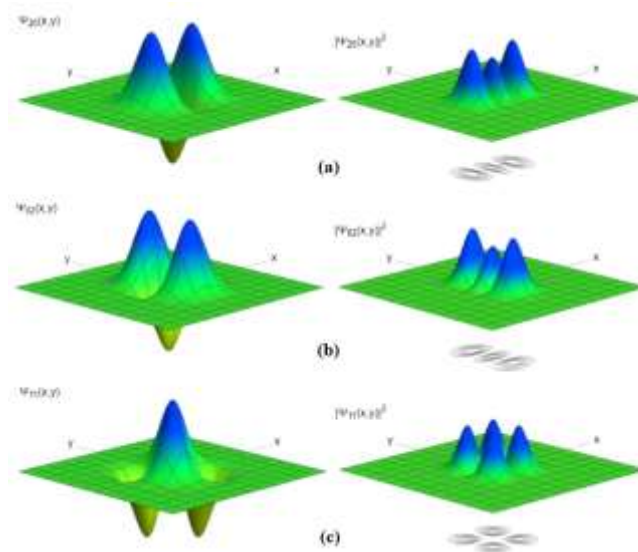

FIG 6. 2D plots for ψ and $|\psi| = \sqrt{\psi\,\psi^*}$ of three degenerate states for quantum number set $<n_x, n_y>$: (a) <2,0> (b) <0,2> (c) <1,1> for particle executing simple harmonic oscillations.

The learners get the idea of the non-oscillatory behavior of a particle in the ground state <0,0> by plotting ψ $_{00}$ and oscillatory behaviors in excited states by plotting subsequent wavefunctions.

Conclusions

The initial challenges associated with teaching quantum mechanics or nanotechnology to young learners can be eased by utilizing interactive simulation software such as Quantum Dot Simulator. This easy-to-use online simulator is an effective method for teaching-learning. It gives students an intuitive understanding of quantum mechanics. Additional benefits could be to utilize quantum dot lab in optimising the shape and size for optical applications. The further suggestion is to improvise 2D and 3D plotting of wavefunctions to get the shape and orientations for probabilistic cloud for the particles under quantum confinement, as is presented for particle executing simple harmonic oscillations. A positive attitude to

the subject can be provoked among young learners through such pedagogy.

Acknowledgement

References

[1] Stadermann H K E, Berg Evan den B and Goedhart M J 2019 *Phys. Rev. Phys. Educ. Res.* **15** 010130

[2] Krijtenburn-Lewerissa K, Pol H J, Binkman A, and van Joolinggen W R 2020 *Phys. Rev. Phys. Educ. Res.* **16** 010132

[3] Beiser A 2003 *Concepts of Modern Physics* (McGraw-Hill)

[4] Prasad Sarangapani, James Fonseca, Daniel F Mejia, *et. al.* 2018, Quantum Dot Lab, https://nanohub.org/resources/qdot

[5] Prasad Sarangapani, *et al.* "Quantum dot lab: an online platform for quantum dot simulations." *2015* International Workshop on Computational Electronics (IWCE). IEEE, 2015.

[6] Chukwuocha E O, Onyeaju M C, and Harry Taylor S T 2012 *World J. Cond. Mat. Phys.* **2** 96

[7] Wolfe J P 1982 *Physics today* **35** 46

[8] Binder S M, 2021 *Introduction to Quantum Mechanics,* 103 (Academic Press)

[9] Energy eigenfunctions of the two-dimensional quantum oscillator, 2D plotting exercises available at https://www.st-andrews.ac.uk/ (University of St Andrews)