PHYSICS EDUCATION

www.physedu.in

Volume 35, Number 4

In this Issue

A Dozen Beautiful Equations

R. Ramachandran

44, Gopal Puram . IIT Co-op Housing Society, Kanpur 208016, India

Formerly of Institute of Mathematical Sciences, Taramani,

Chennai 600116, India,

Dept. of Physics, Indian Institute of Technology Kanpur,

Kanpur 208016, India

rr_1940@yahoo.co.in, rr@imsc.res.in

Submitted on 12-Nov-2019

Abstract

A Dozen Beautiful Equations that has helped the Evolution of Physics and Mathematics are compiled. Each is a gem and gave rise to entire tower of applications in many fields beyond where it originated. The names associated with each one of them defines the genius of their time and celebrated for all subsequent eternity.
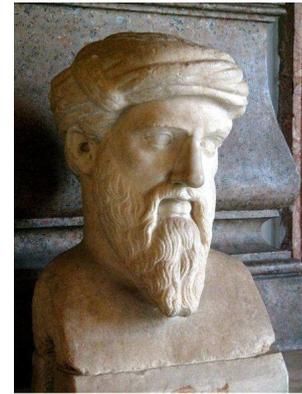
Figure 1: Pythagoras

1 Baudhayana - Pythagoras Theorem

The equation

$$x^2 + y^2 = z^2$$

where x, y and z are the lengths of base, height and hypotenuse for every right angled triangle is well known to all high school students. This theorem is usually attributed to Greek scholar Pythagoras[1](c. 570 - c. 475 BC). We now know that Baudhayana (around 800 BC), a Vedic scholar had a formulation of the same theorem in his *Sulba Sutra*, a manuscript with several mathematical gems. It says: "A rope stretched along the length of the diagonal (of a rectangle) produces an area which the vertical and horizontal sides make together."[2]

If x, y and z are, in addition integers, they form what is known as a Pythagorean

triple. The set $\{3, 4, 5\}$ and $\{5, 12, 13\}$ are two well known examples. Note that if a, b and c form a triple, it is easy to see that ka, kb and kc with any integer value k is also one. There are infinitely many such distinct sets and each will generate additional sets. *Sulba Sutra* gives prescriptions to identify a class of such triples.[1]

Incidentally, the French mathematician Pierre de Fermat, in 1637, conjectured that there are no integer valued solution for x, y and z to the equation

$$x^n + y^n = z^n$$

for any integer $n > 2$. The proof for the correctness of the conjecture remained elusive, even though Fermat himself had remarked in his copy of *Arithmetica*, an Ancient Greek text on mathematics written by the mathematician Diophantus in the 3rd century *CE* "that he has found a proof", but he had added that "it can not fit in the margins of his note book with him to jot down". He never published the proof. Most likely the alleged proof, that he had, was flawed. It needed several sophisticated developments in the field of Differential Geometry for the British Mathematician Andrew Wiles to provide an unequivocal proof for the validity of Fermat's conjecture in 1995, earning for him the laurels of Abel Prize in the year 2016.

2 Napiers Theorem

Next example of remarkable Equations is:

$$\log x + \log y = \log xy; \quad x > 0, \quad y > 0$$

Logarithm is a a mathematical function that maps the operation *Multiplication* into *Addition*. The logarithm of the multiplicative product of two numbers, is a sum of their logarithms. This opens up a beautiful plethora of possibilities in very many fields. A Scottish Mathematician John Napier is identified with discovering and developing logarithms[3]. He developed a tool set for multiplications - which were referred to as Napier's Logs, which are the logarithm tables, that we used in our school days. They formed the principle for the *Slide Rule* that most engineers and surveyors used to carry before the advent of modern electronic calculators.

This is in a sense, inverse of the operation of 'raising a number to some power'. Let us take a number 10, raise it to its 3rd power, meaning $10 \times 10 \times 10$ denoted as $10^3 = 1000$. Indeed it is easy to see that $10^3 \times 10^2 = 10^{3+2} = 10^5$. Generalising,

$$10^x \times 10^y = 10^{x+y} \quad x, y \text{ any real number.}$$

Further generalising, the base 10 can be replaced by any positive number; a particularly interesting choice is the transcendental number called Euler's constant $e = 2.718281828459.....$

$$e^x \times e^y = e^{x+y} \quad x, y \text{ any real number.}$$

Let us get back to logarithms. The function has an attribute called base, which

1. All images are downloaded from Wikipedia, mostly copies of portraits from some archive.

Figure 2: John Napier, Scottish National Gallery

could be taken as either 10, e, 2 or any positive number; $\log_{10} x$ when the base is 10 and $\log_e x$ if the base is Euler constant e. If we denote $y = \log_{10} x$ then $x = 10^y$. Similarly if $y = \log_e x$ then $x = e^y$, which is referred to as the exponential function.

The tables of logarithms with base 10 computed upto required decimals (tables of 4 decimal places was in common usage in schools during our time $\sim$ 1950's) used to serve as a devise to quickly carry out multiplication of numbers. When two numbers are to be multiplied, instead of carrying out usual multi-step algorithm, one may more simply add their logarithms to get the logarithm of their product and look at the table of what is referred to as anti-logarithms to read off the number correct upto desired number of decimals. Often when we write $\log x$ it is implied that the base is 10.

When the base is e, the logarithm is referred to as 'natural logarithm' and denoted as $\ln x$. It is also often referred to as 'Napierian logarithms'. Since the exponential function has a representation as a power series :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

as x increases, the function increases faster than any polynomial. The functions $\{x^n, \quad 1 > n > 0\}$, like polynomials (for example x^k any $k \geq 1$), are also monotonically increasing function, however growing slower than linear functions. The logarithmic functions are also increasing function as x changes but with rate even slower than those of fractional power. We may note that the above expression for e^x also provides us a definition of the Euler constant e through

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

The base 2, called bit is used, naturally, in information science.

3   Leibniz Differential Calculus

The equation that was arrived at, independently by Leibniz and Newton, formed the foundation of classical analysis[4][5]:

$$\frac{dy}{dx} = \lim_{\delta \to 0} \frac{y(x + \delta) - y(x)}{\delta}$$

If there are two variables, say, position of an object x and time t, x vs t or $x(t)$ provides us

Figure 3: Gottfried W Leibniz

a trajectory. The rate at which x varies at any instant is given by $\frac{dx}{dt}$ and hence $v(t) = \frac{dx}{dt}$ is a derivative of $x(t)$. We may picture the trajectory in a graph with ordinates x and t. We can visualise the derivative v as the tangent drawn at every point on the x vs t curve as illustrated on the figure below.
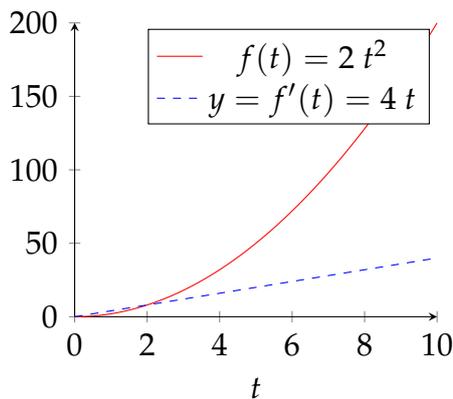
The process of the differentiation brings about the notion of continuity and the limiting process. We recognise that the speed when defined as a (finite) distance travelled in a (finite) interval of time as being average speed in that duration. The notion of velocity at an instant requires this to be computed as a limit process of infinitesimal displacement over an infinitesimal interval of time. We begin the era of precision analysis.

We may here add a remark that Differ-

ential and Integral Calculus and many pioneering mathematical theorems indeed originated in India from about mid 14th century (Madhava School) to about 1620. *Yuktibhasa*, mainly based on Nilakantha's *Tantra Samgraha in Sanskrit* is considered, possibly the first text on the foundations of Calculus and pre-dates those of European mathematicians. Madhava's treatment is more precise and reflects erudite Mathematics it embodied, a fact not fully appreciated by the modern Historians of Mathematics. [6]
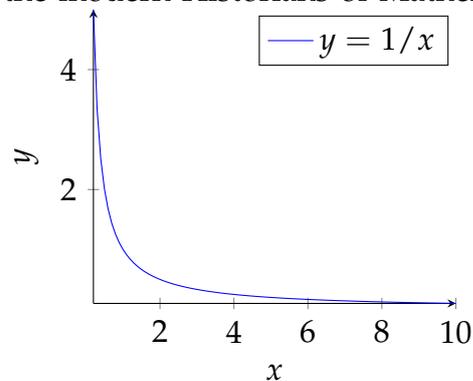
We perceive that physical properties can be described by continuously varying functions. Their derivatives enabled through differentiation are also physical observables. Physical law in view of its smooth behaviour are described by Differential Equations. The inverse process of Differentiation is Integration. Graphically differentiation is to determine the tangent to the curve and the process of integration is the measure of the area swept by the curve over an interval. If $y = f(x)$ then $y' = \frac{df}{dx}$ is the derivative function. The inverse process is $f(x) = \int dx y' = \int dx \frac{df}{dx}$. The derivative of x^n is nx^{n-1} Naturally $\int dx x^n = \frac{x^{n+1}}{n+1}$.

The derivative of $\ln x$ is $\frac{1}{x}$ and hence we

have a neat definition for the function $\ln x$:

$$\ln x = \int dx \, \frac{1}{x} \, ,$$

$$\ln x - \ln x_0 \equiv \ln \frac{x}{x_0} = \int_{x_0}^{x} dx \, \frac{1}{x} \, .$$

The area under the curve in the second graph is geometric way to get the logarithmic function.

4 Newton - Laws of Motion, Law of Gravity

British Scientist Isaac Newton is identified with two different iconic equations, one which lays the foundation for classical dynamics and the other spells out Universal Law of Gravitation.

The three laws of motion by Newton are[7]:

1. A body continues to be in a state of rest or uniform motion unless acted upon by a force on it.

2. A system of mass m accelerates on the application of a force on it such that

$$m\frac{d^2\mathbf{x}}{dt^2} = \mathbf{F},$$

or equivalently

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}.$$

where $\mathbf{x}$ and $\mathbf{p}$ are respectively the position and momentum attribute of the body or system.

3. In an isolated system, for every force acted upon a body there is an equal amount
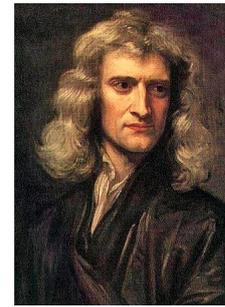

Figure 4: Isaac Newton

and oppositely directed reaction by the body on the agency that exerts force. These sum up the firm foundation of the classical dynamics and lets us understand the nature of inertia. Note that the acceleration of the body is proportional to the force whatever be the nature of force (gravitational, electrical or frictional and whatever) and the quantum of acceleration is proportional to its mass, which measures its inertia. It is also referred to as inertial mass m_{in}. Forces come in many forms. Electric Forces act on bodies with electric charge, magnetic forces on electric current. Gravitational forces are exerted by all bodies, the magnitude of the force proportional to its mass content. Two bodies of mass m and M exert mutual attraction to each other and follows Newton's Universal Law of Gravitation[7]:

$$F = \frac{G_N m M}{r^2} \qquad \mathbf{F}_{gravity} = -\frac{GmM}{r^2}\hat{\mathbf{r}}.$$

G_N is Newton's gravitation Constant

It is remarkable that this equation is responsible for all planets to exhibit elliptical tra-

jectory about Sun; and a near circular, but really elliptical orbit of moon about earth; - as well as how the legendary apple from the tree allegedly fell on the head of Newton. That the forces are universal and do not depend on the detailed composition of the body is important to note. The magnitude of the gravitational force is indeed proportional to its mass that thus quantifies its gravitational content. To distinguish from the earlier definition, we may term this as gravitational mass m_{gr}.

Observe that the two definitions of mass, the first signifying the amount of inertia and the second arising from it being the source of gravitational force field need not be the same. That they indeed are, is due to universal nature of gravity, which we become first aware through a remarkable experiment carried out by Galileo; who found that the time taken by two different bodies, say one heavy and the other light or one of stone and another some metal, from the top of leaning tower of Pisa to the ground was exactly identical. The force experienced by the body is proportional to m_{gr} and the acceleration it induces on it is proportional to $1/m_{in}$, implying for the time taken to differ if the ratio m_{gr}/m_{in} for different bodies vary. The absence of such variation allows us to identify both masses to be the same. This classic experiment, rather profound, was repeated, with enhanced precision by Hungarian Physicist Eötvös with the use of an ingenious torsion balance at the start of last century[8], firmly establishing universality

Figure 5: Leonhard Euler

of gravity.

Thus begins the glorious era of Classical Mechanics.

5 Euler's Identity - The Beautiful Equation

$$e^{i\pi} + 1 = 0$$

Euler's identity is beautiful because it combines five of the most important constants (numbers) in mathematics into a single elegant equation.[9]

1 – the basis of all numbers in the set and the multiplicative identity

0 – the symbolic concept of nothingness and the additive identity

π – the ratio of a circle's circumference to its diameter ($\pi = 3.14159\ldots$)

e – the base of natural logarithms which occurs widely in mathematical analysis ($e = 2.718\ldots$).

i – the "imaginary" number, representing square root of -1, $i = \sqrt{-1}$.

Interestingly a combination of two entirely different transcendental quantities can be combined together to become an integer!

The identity brings five of the super-heroes of mathematics $(e, i, \pi, 1, 0)$ together with three of the basic arithmetic operations (addition, multiplication, and exponentiation).

We quote Benjamin Pierce, a noted American 19th-century philosopher, mathematician, and a professor at Harvard: "$\cdots$ is absolutely paradoxical; we cannot understand it, and we don't know what it means, but we have proved it, and therefore we know it must be the truth."

A study has also, it seems, shown that Euler Identity is so beautiful that it excites the same areas of the brain as a great piece of music or art would.

In a sense this is not that much surprising because the exponential of a complex variable z is a product of two parts: $e^z = e^{x+iy} = e^x e^{iy} = e^x(\cos y + i \sin y)$. e^x is either rapidly growing function when $x > 0$ or decaying one if $x < 0$. Both real and imaginary parts of e^{iy} oscillate between ± 1. It is clear that when $z = in\pi$, e^z is integer valued ± 1, positive when n is even and negative when n is odd.

6 D'Alembert Wave Equation & Fourier Heat Equation

$$\frac{\partial^2}{\partial t^2} u(x,t) = c^2 \frac{\partial^2}{\partial x^2} u(x,t)$$

is a partial differential equation derived by a French Mathematician Jean le Rond d'Alembert, in 1747 as the equation to solve the problem of a vibrating string[10]. It is

Figure 6: Jean le Rond d'Alembert

a prototype equation with a wide range of applications, describing the phenomena of all kinds of wave motions - acoustic waves, temperature (heat) waves, electromagnetic waves are some of the classic examples.

A generic solution of the equation is $u(x,t) = u_0 \ e^{i(kx - \omega t + \phi)}$; $k = \frac{2\pi}{\lambda}$, $\frac{\omega}{c} = \frac{2\pi \nu}{c}$. It can also be written as $u(x,t) = u_0 \ \sin\left(\frac{2\pi x}{\lambda} - 2\pi\nu t + \phi\right)$ with the condition $\nu\lambda = c$, oscillatory functions period λ in length and period T in time (or frequency $\nu = 1/T$).

Partial differential equations find applications in many branches of Physics and Engineering and the d'Alembert equation and the analogue Laplace equation together with boundary conditions defines many phenomena. For the one dimensional second order linear equation above, the solution has two constants, such as u_0, ϕ are fixed by two boundary condition, say by specifying $u(x,0)$ and $\frac{du}{dt}\big|_{t=0}$.

In three dimensions the analogous equation is $\left(\frac{\partial^2}{\partial t^2} - c^2 \nabla^2\right) u(\mathbf{x}, \mathbf{t}) = \mathbf{0}$.

In physics and mathematics, the heat equation is a partial differential equation

Figure 7: Joseph Fourier

Figure 8: Claude-Louis Navier

that describes how the distribution of some quantity (such as heat) evolves over time in a medium, as it flows from places where the temperature is higher towards places where it is lower. It is similar to diffusion equation, relevant for flow of fluid from higher pressure region to one that is lower. This equation was first developed and solved by Joseph Fourier in 1822 to describe heat flow and hence is the name[11]. It is of fundamental importance in diverse scientific fields. With D as the diffusion constant, it takes the form

$$\frac{\partial}{\partial t} u(t, \mathbf{x}) = D \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial x_j^2} + \frac{\partial^2}{\partial x_k^2} \right) u(t, \mathbf{x})$$

The heat equation is the prototype example of a parabolic partial differential equation. If the solution is separable in variables, say time and space in the form $u(t, x) = A(t)f(x)$ it is seen that the partial differential equation reduces to coupled ordinary differential equations. For example, the equation

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}$$

becomes

$$\frac{dA}{dt} f = DA \frac{d^2 f}{dx^2}$$

$$\frac{dA}{dt} = -k^2 DA \text{ and } \frac{d^2 f}{dx^2} = -k^2 f$$

$$A \sim e^{-k^2 Dt} \text{ and } f(x) \sim \sin(kx + \phi).$$

If the boundary condition is such that $u(t = 0, x) = f(x)$ and $u(t, 0) = u(t, L) = 0$ then $kL = n\pi$, leading to most general solution in the form:

$$u(t, x) \sim e^{-\frac{n^2 \pi^2}{L^2} Dt} \sin \left(\frac{n\pi x}{L} + \phi \right).$$

7 Navier - Stokes Equation

The Navier-Stokes equation, in modern notation, is

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{\nabla P}{\rho} + \nu \nabla^2 \mathbf{u}$$

where $\mathbf{u}$ is the fluid velocity vector, P is the pressure, ρ is the density, ν is the kinematic viscosity of the fluid, and ∇^2 is the Laplace operator. Navier-Stokes equation, in fluid mechanics, is a partial differential

Figure 9: George Stokes

equation that describes the flow of incompressible fluids. The equation is a generalization of the equation by Swiss mathematician Leonhard Euler in the 18th century obtained to describe the flow of incompressible and frictionless fluids. In 1821 French engineer Claude-Louis Navier introduced the element of viscosity (friction) for the more realistic and vastly more difficult problem. Throughout the middle of the 19th century, British physicist and mathematician Sir George Gabriel Stokes improved on this work, though complete solutions were analytically possible only for a few simple two-dimensional flows. In view of the non linear nature, closed analytical solutions are not an available option. The complex vortices and turbulence and chaos, that occur in three-dimensional fluid (liquid or gas) flows as velocities increase have proven intractable to any but approximate numerical analysis methods. We look for solutions in different regimes, characterised by a dimensionless Reynold number, which, in effect, is the ratio between the inertial forces in a fluid and the viscous forces. A fluid in motion tends to behave as sheets or layers of

infinitely small thickness, sliding relative to each other, forming a laminar flow. Indeed this equation is the foundation for classical fluid dynamics, the basis for many phenomena in almost all branches of Science and Technology, such as Aeronautical Engineering, Chemical Technology, Atmospheric Sciences and Oceanography. [12]

8   Maxwells Equations of Electrodynamics

James Clark Maxwell unified Electricity and Magnetism and summarised the underlying laws governing Electrodynamics. Maxwell's equations[13] are a set of four equations that form the theoretical basis for describing classical electromagnetism. Electromagnetic interactions govern most of Physics and all of Chemistry and biology.
The Lorentz law, where q and $\mathbf{v}$ are respectively the electric charge and velocity of a test particle, defines the electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ by specifying the total electromagnetic force $\mathbf{F}$ as

$$\mathbf{F} = q[\mathbf{E} + \mathbf{v} \times \mathbf{B}].$$

• Gauss's law: Electric charges produce an electric field.
• There are no magnetic monopoles, which could be magnetic analogue of electric charges.
• Ampere's Law: Steady electric current is the source of Magnetic fields.
• Faraday's law: Time-varying magnetic fields produce an electric field and hence an

Figure 10: James Clerk Maxwell

electric current.

These four laws constitute Maxwell Equations that govern Classical Electrodynamics

Gauss Law:

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

$$\oint_{surface} \mathbf{E} \cdot d\mathbf{s} = \frac{1}{\epsilon_0} \int_{volume} \rho dV$$

are the differential and integral forms. The surface integral is over a closed surface $\mathbf{s}$ enclosing the volume V with charge distribution ρ.

No Monopole: $\Rightarrow$

$$\nabla \cdot \mathbf{B} = 0,$$

$$\oint_{surface} \mathbf{B} \cdot d\mathbf{s} = 0$$

Ampere's Law:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t},$$

$$\oint_{loop} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_s \mathbf{j} \cdot d\mathbf{a} + \mu_0 \epsilon_0 \frac{d}{dt} \int_s \mathbf{E} \cdot d\mathbf{a}$$

are the differential and integral forms. In the latter the integration over a closed loop l that encloses the flux of electric field across the surface $\mathbf{a}$.

Faraday's Law:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t},$$

$$\oint_{loop} \mathbf{E} \cdot d\mathbf{l} = -\epsilon_0 \frac{d}{dt} \int_s \mathbf{B} \cdot d\mathbf{a}$$

are analogous differential and integral forms. The induced EMF over the closed loop l is the negative rate of change of magnetic flux across the surface $\mathbf{a}$ enclosed by the loop.

The charge density $\rho(t, \mathbf{x})$ and current density $\mathbf{j}(t, \mathbf{x})$ obey continuity constraint on account of conservation of electric charge:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0,$$

$$Q \equiv \int_{volume} \rho dV = -\oint_{surface} \mathbf{j} \cdot d\mathbf{s}.$$

The laws of electrodynamics are succinct expression behind the technology developed in the 20th century. It enables us to use energy in the most convenient form. The preceding century relied on heat as pri-

mary source of energy and mechanical energy it can generate through steam turbines as a means to turn the wheels of technology. We now proceed further and, thanks to Faraday, have the turbines work electrical generators to produce electrical energy, indeed a more convenient form to transmit. Same law is responsible for it to run an electric motor to result in myriad technical applications.

We may further note that if in a charge free region $\mathbf{E} = E(x)\hat{\mathbf{k}}$ and $\mathbf{B} = B(x)\hat{\mathbf{j}}$, both fields are functions of only x variable, then Maxwell's Equations implies:

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t} \qquad \text{and} \qquad \frac{\partial B}{\partial x} = -\frac{1}{c^2}\frac{\partial E}{\partial t}$$

where we have used $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$. Taking partial derivative of the first equation with respect to x, the second with respect to t and equating the pair we can conclude that

$$\frac{\partial^2 E}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}$$

Similarly we can derive

$$\frac{\partial^2 B}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 B}{\partial t^2}$$

The orthogonally oriented E and B obey one dimensional wave equation representing an electromagnetic wave propagating along x direction. Indeed, Maxwell was the first to provide a theoretical explanation for Light to be a classical electromagnetic wave and, in doing so, compute the speed of light c in terms of known constants: magnetic and electric permittivity μ_0 and ϵ_0.

Thus Electricity, Magnetism and Optics get fully integrated.

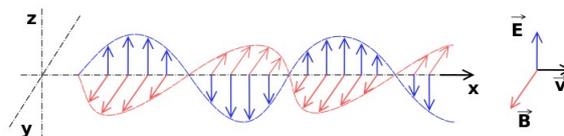

Figure 11: Electromagnetic wave

9 Boltzmann and Entropy

The iconic equation

$$S = k_B \ln W$$

$$k_B = 1.380649 \times 10^{-23} J/K$$

referred as Boltzmann constant

is an absolute Gem; engraved as an epitaph on the grave of Ludwig Boltzmann. Newtonian mechanics deals with the dynamics of a few bodies. However in reality we deal with system made up of large number of bodies. How do these laws translate? We define macro-variables such as Volume V, Pressure P, Temperature T, Internal Energy E etc. One enigmatic variable is Entropy S, which is sometime defined through $\delta S = \frac{\delta Q}{T}$. It is related to the probable distribution of the micro-states, the system corresponds to. It is referred as the quantitative measure of an inherent *disorder* when there is a large collection of bodies.

For an isolated system, say gas in a chamber of Volume V at some Pressure P and Temperature T. we can associate its heat content as Q or equivalently its internal Energy E and its Entropy as S. How does one measure Entropy? How will S change as the volume of the system is doubled? Model

the gas as a collection of N identical distinguishable objects (gas of molecules) distributed in the two halves of a chamber, N_1 in chamber 1 and N_2 in chamber 2 and a barrier separating the two. Number of micro-states or distinct configurations is given by $W = \frac{N!}{N_1! N_2!}$, with $N = N_1 + N_2$. Initially when the gas is contained in chamber 1 and chamber 2 empty, $W_{in} = \frac{N!}{N!0!} = 1$, implying $S_{in} = k_B \ln W_{in} = 0$. On removing the barrier the N objects gets divided into most probable distribution with $N_1 = N_2 = N/2$ in each half. The number of distinct distributions finally is given by the expression $W_f = \frac{N!}{(N/2)!(N/2)!}$. Using Stirling Formula $\ln N! \sim N \ln N - N$ for large N, we can express:

$$S_f = k_B \left[\ln N! - 2 \ln \frac{N}{2}! \right]$$

$$= k_B \left[N \ln N - N - 2\frac{N}{2} \ln \frac{N}{2} + 2\frac{N}{2} \right]$$

$$= k_B N \ln 2.$$

$$\Delta S = S_f - S_{in} = k_B N \ln 2$$

We may now use an alternate way to compute the change in Entropy for the (ideal) gas, expanding isothermally to double its volume, using what we learnt in high school. Recall the First law of Thermodynamics

$$dQ \equiv TdS = dE + PdV = dE + RT\frac{dV}{V}.$$

It is seen that (since for an isothermal expansion there is no change in E) this leads to

$$\Delta S \sim R \ln \frac{V_f}{V_{in}} = k_B N \ln 2$$

when we identify gas constant $R = k_B N$.

Classical Physics starts from Galileo, Newton and Maxwell and needs Boltzmann's statistical processes to reach the pinnacle. Kinetic Theory of Heat links the macroscopic state of an ideal gas (no interaction) with the associated micro-states that makes up the macro-state. Heat energy is absorbed by a system at source temperature T, works an ideal Carnot's engine and discards unspent energy to a sink at $T' < T$. Efficiency of the Engine is derived to be $\frac{T-T'}{T}$, always less than unity. All realistic engines have lesser efficiency than the *ideal* Carnot engine. With this perspective, we may give an alternate understanding of Entropy as the *realisable* content of the Energy of the system.

The collection of objects can have a range of energy values, discrete as well as continuous; what is the relative probability that an object has energy E? Boltzmann factor $e^{-\beta E}$ is the 'Occupancy rate' for a classical particle where $\beta = 1/k_B T$. If the density of energy levels is $\rho(E)$, the Maxwell - Boltzmann distribution, when the system has reached a thermal equilibrium, is given by

$$W(E) = \eta(E)\rho(E) \quad \eta(E) = e^{-\beta E}.$$

For an ideal gas of particles of (molecular) mass m with $E = \frac{1}{2}mv^2$ and $\rho = 4\pi v^2 dv$ we have a Maxwellian velocity distribution

$$W \propto v^2 e^{-\frac{mv^2}{2k_B T}}.$$

Quantum mechanics adds a variation to the occupancy rate η. Identical particles are, in

Figure 12: Ludwig Boltzmann - Epitaph

principle, not distinguishable, since the notion of trajectory of a particle is no longer possible. In effect this causes a quantum mechanical repulsion among identical fermions (such as electrons, protons etc., when Pauli's exclusion principle is operative) and a tendency of attraction for bosons (such as photons, Helium nuclei etc., capable to form Bose condensates) and is expressed by the formula

$$\eta = \frac{1}{e^{\beta E} + \alpha},$$

α assumes a value +1 for fermions, -1 for bosons and 0 for a classical particle.

The concept of information entropy was picked up by Claude Shannon in his 1948 paper "A Mathematical Theory of Communication"[15], where he uses Entropy as a measure of the unpredictability of the state, or equivalently, of its average information content and a parallel formula

$$H = - \sum_{i=1}^{n} P(x_i) \log_2 P(x_i).$$

where H is Information Entropy and $P(x_i)$ denotes the probability measure of a ran-

dom variable shows up. Information entropy is the average rate at which information is produced by a stochastic source of data.

10   Einstein Relativity

At the beginning of the 20th century we witness two revolutionary changes in our physical perception - Theory of Relativity and Quantum Mechanics. In the first, Newtonian concept of absolute space and absolute time needed to be given up. In its place Einstein, treating the concepts of Space and Time as inter-dependent and measurements by all observers as being *relative*, made two postulates[14]:

1. Physics appears the same in all inertial frames of reference (All observers in relative uniform motion constitute inertial frames for reference)

2. Velocity of light c, as observed in any inertial frame is the same, numerically about 3×10^8 m/s.

This summarises Principle of Special Relativity.

Two frames of references, with Cartesian co-ordinates labelled as (x, y, z, t) in frame S and (x', y', z', t') in frame S' with S' moving relative to frame S with velocity v along direction of unit vector $\hat{\mathbf{i}}$ along x-axis are related by Lorentz transformation equations:
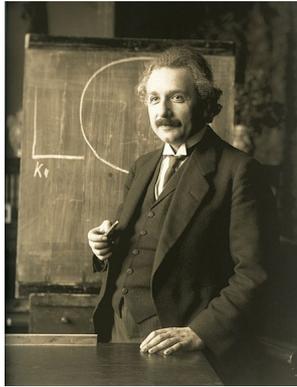
Figure 13: Albert Einstein-1921

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}$$

$$y' = y \qquad z' = z$$

$$ct' = \frac{ct - vx/c}{\sqrt{1 - v^2/c^2}}$$

There are many profound consequences: Space and Time get interlinked; moving observer will find the lengths contracted by a factor $\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$ and time dilated by $\frac{1}{\gamma} = \sqrt{1 - v^2/c^2}$. Simultaneity is not preserved by Lorentz transformation. Events that are simultaneous in one frame need not be so in the other. What was identified as inertial mass by Newton needs a revisit, since Energy, Momentum are now frame dependant conserved quantities and mass is a derived quantity. The term mass gets identified with an entity that is invariant in all inertial frames of reference, related by relevant Lorentz transformation. Key equation for a system is

$$E^2 = \mathbf{p}^2 c^2 + m_0^2 c^4.$$

m_0 is known as the invariant mass. When it is at rest, momentum $\mathbf{p} = 0$ and the energy content of a mass at rest is $m_0 c^2$. Both energy and momentum increase as the particle is set in motion. For small velocities $|\mathbf{v}| \ll c$, the momentum increases linearly $\mathbf{p} = m_0 \mathbf{v}$. For larger velocities, it grows as $\mathbf{p} = \frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}}$. Substituting for $\mathbf{p}$ we get

$$E^2 = \frac{m_0^2 v^2 c^2}{(1 - v^2/c^2)} + m_0^2 c^4 = \left(\frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} \right)^2$$

$$\Rightarrow E = mc^2, \text{with } m = \frac{m_0}{\sqrt{1 - v^2/c^2}}.$$

We thus arrive at the iconic expression

$$E = mc^2$$

declaring the equivalence of mass as its energy content.

Just as formulated by Newton, mass or equivalently Energy, is the source of Gravity. However, the General Theory of Relativity (proposed by Albert Einstein in 1915[16]) which describes gravity not as a force, but as a consequence of the curvature of space-time caused by the distribution of mass. A flat space - time co-ordinates are given by a set of Cartesian co-ordinates x^μ; $\mu = 0, 1, 2, 3$ and the distance interval between two neighbouring points differing by dx^μ is given by an invariant

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2.$$

More generally for a general curvilinear co-ordinates we may define a second rank symmetric tensor $g_{\mu\nu}(x^0 = ct, x^i; i = 1, 2, 3)$ as

metric tensor, which has the information of the curvature in its functional dependence. Indeed the derivative of a scalar field $S(x^\mu)$: $V_\lambda(x) = \frac{\partial S}{\partial x^\lambda}$ is a Vector field. However the derivative of Vector field is not a tensor field. We need to account for the curvature and parallel transport of the vector displacement needs another factor, namely affine connection:

$$T_\lambda^\mu(x) = \frac{\partial V^\mu(x)}{\partial x^\lambda} + \Gamma_{\lambda\sigma}^\mu(x)V^\sigma(x)$$

Christoffel symbol $\Gamma_{\lambda\sigma}^\mu$ can be expressed in terms of the metric tensor:

$$\Gamma_{\lambda\sigma}^\mu = \frac{1}{2}g^{\mu\nu}\left(\frac{\partial g_{\nu\sigma}}{\partial x^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x^\sigma} - \frac{\partial g_{\lambda\sigma}}{\partial x^\nu}\right)$$

The curvature is characterised by the non-zero value for the elements of a fourth rank Riemann Christoffel Tensor written as

$$R_{\alpha\beta\gamma}^\mu = \left(\frac{\partial \Gamma_{\alpha\gamma}^\mu}{\partial x^\beta} - \frac{\partial \Gamma_{\alpha\beta}^\mu}{\partial x^\gamma} + \Gamma_{\sigma\beta}^\mu\Gamma_{\gamma\alpha}^\sigma - \Gamma_{\sigma\gamma}^\mu\Gamma_{\gamma\beta}^\sigma\right)$$

Define the related Ricci tensor $R_{\mu\nu}$ and Ricci Scalar curvature by contracting the curvature tensor to get:

$$R_{\mu\nu} = R_{\mu\sigma\nu}^\sigma, \qquad R = g^{\mu\nu}R_{\mu\nu}.$$

We arrive at the celebrated Einstein equation of Gravitation

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G_N}{c^4}T_{\mu\nu}$$

where $G_{\mu\nu}$ is known as Einstein tensor, G_N is Newton's Gravitational constant and $T_{\mu\nu}$ as the Energy Momentum tensor. The left hand side of the equation describes the geometry of space-time that results from its source of Energy-Momentum distribution that is on the right hand side.

11 Schrödinger Equation

Classical Mechanics relied on Newton's equation that asserted that once we know the initial position and initial momentum/velocity of the system and the force it is subjected to, we can trace its trajectory. Quantum uncertainty principle, that constraints our simultaneous knowledge of the both: $\Delta x \Delta p \geq \hbar/2$, renders the notion of trajectory fuzzy and puts an end to the determinacy, that was the norm for solutions to dynamics of the classical era. The paradigm shift of probabilistic nature of dynamics turns all observables into linear operators in a Hilbert space that describes the state of the system, represented through a complex valued wavefunction Ψ. The operator associated with the momentum is $p = \frac{\hbar}{i}\frac{\partial}{\partial x}$ [In 3 d space it takes the form $\mathbf{p} = \frac{\hbar}{i}\nabla$] and similarly Energy $E = \frac{\hbar}{i}\frac{\partial}{\partial t}$. Indeed total Energy is made up of Kinetic Energy ($\mathbf{p}^2/2m$) and Potential Energy ($V(x)$) and so we may treat $E = \frac{\mathbf{p}^2}{2m} + V(x)$ as an operator equation. Ernest Schrödinger, thus derived the equation[17] for the wave-function $\Psi(\mathbf{x}, t)$, a complex valued function in the configuration space of the system (here the position of the particle $\mathbf{x}$ at an instant t):

Figure 14: Erwin Schrödinger 1933

Figure 15: P A M Dirac

$$\frac{\hbar}{i}\frac{\partial \Psi}{\partial t} = -\hbar^2 \nabla^2 \Psi + V(\mathbf{x})\Psi;$$

$\Psi(\mathbf{x},t)^*\Psi(\mathbf{x},t)$ is the probabilistic density is normalised so that $\int d^3\mathbf{x}|\Psi|^2 = 1$.

Non-relativistic Quantum Mechanics makes a smooth transition from the classical dynamics to the quantum analogues. The wave functions for Hydrogen atom and for Harmonic Oscillator could be solved exactly by the relevant Schrödinger Equations. We find many features in Atomic and molecular Physics, condensed matter Physics, Optics well explained by non relativistic Quantum Mechanics. When the equations are not exactly solvable, we are able to get approximate solutions to the desired level of precision by using perturbation and variational techniques.

12 Dirac Equation

The final beautiful equation we list is the famous Dirac Equation for electron[18]. Indeed it is the correct equation of all spin 1/2 particle, such as protons, neutrons, all

quarks and leptons (e, μ, τ) etc. that form the basis for all matter. Amusingly it was referred to as the equation that which describes most of physics and ALL of Chemistry and Biology. That is a bit of hyperbole - but it does emphasise the special role played by the valence electrons (that are relatively free in atoms and molecules0 in shaping the interactions that are aspects of the structure and dynamics in Physics and Chemistry. The interesting feature is that it represents free electron with arbitrary energy, spreading to its relativistic domain as well. Dramatically, it necessitated and predicted the existence of positron the conjugate antiparticle of electron before its discovery in the laboratory. In the non-relativistic Schrödinger wave-functions are represented by a complex valued square integrable wave function, say $\phi(\mathbf{x},t)$, the electron is represented by a (4 - component) 'Spinors' $\psi_\alpha(x^\mu)$; $\mu = 0, 1, 2, 3$ the space-time indices, $\alpha = 1, 2, 3, 4$ are the indices giving the row number of a coloumn matrix. An electron state may be labelled with its Momentum, Energy and its two fold Spin state. We will give the orig-

inal version Dirac proposed and evolve to presently used form.

We start with the set of three 2×2 Pauli matrices σ_i:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and note that the components of the spin angular momentum $S_i = \hbar\sigma_i/2$ satisfies the algebra of angular momentum $[S_i, S_j] = i\hbar\epsilon_{ijk}S_k$, where ϵ_{ijk} is the Levi-Civita symbol non zero only when $i \neq j \neq k$ and is +1 for $i, j, k = 1, 2, 3$; $= 2, 3, 1$ and $= 3, 1, 2$ and is -1 otherwise. Define a set of mutually anti-commuting 4×4 hermitian matrices β, α_i as

$$\beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \quad \alpha_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}$$

$$\text{i = 1,2,3 and with } I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The matrices are mutually anti-commuting and their squares are I_4, identity matrix in 4 dimension. [$\{\beta, \alpha_i\} = 0$, $\beta^2 = I_4$ and $\{\alpha_i, \alpha_j\} = 2\delta_{ij}I_4$]. The Dirac equation in the form originally proposed by Dirac is:

$$\left(\beta mc^2 + c\frac{\hbar}{i}\vec{\alpha} \cdot \vec{\nabla} \right) \psi(\mathbf{x}, t) = -\frac{\hbar}{i}\frac{\partial\psi(\mathbf{x}, t)}{\partial t}$$

where ψ is a 4-component spinor representing a free electron of mass m. The spin state of the electron is encoded in the spinor wave function.

The covariant form of Dirac equation is obtained with following identification: with

space-time co-ordinates x^μ, define $\gamma^0 = \beta$ and $\gamma^i = \beta\alpha_i$ to provide us with four Dirac matrices γ^μ, $\mu = 0, 1, 2, 3$ that satisfy the anti-commutation relations:

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}I_4,$$

The metric $\eta_{\mu\nu}$ is of the form:

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Dirac Equation assumes in compact covariant form:

$$\boxed{i\hbar\gamma^\mu\frac{\partial\psi}{\partial x^\mu} = mc\psi}$$

We may also write it as $(\gamma^\mu p_\mu - mc)\psi = 0$, where p_μ is the four-momentum operator, further abbreviated as $(\not{p} - mc)\psi = 0$ using the shorthand notation for $\gamma^\mu p_\mu$ as $\not{p}$.

13 Epilogue

Indeed, the choice of Equations for 'Beautiful Dozen' is subjective. You may make a different set. A discerning reader may notice that I really have a couple more than a dozen by squeezing in one section both creations of Newton on one hand and got Wave Equation and Heat Equation together in another section, notwithstanding a tower of phenomena that are linked with each. Similarly both revolutions of Einstein are grouped under Relativity. You may prefer

more modern (more contemporary) gems, such as Yang Mills Equation that uncovers the 'gauge principle' as unifying ALL fundamental forces (save gravity) and/or London- Anderson- Englert- Brout- HIGGS- Guralnik- Hagen- Kibble- Weinberg mechanism that brings out the subtle principle of hidden symmetry as a beautiful dynamical concept, relevant in many different areas. I could draw a line at the turn of the last millennium, but that will deprive us of the last two or even three entries. I may have missed many!

References

[1] W K C Guthrie, (1978) *A history of Greek philosophy, Volume 1: The earlier Presocratics and the Pythagoreans*, p. 173. Cambridge University Press.

[2] George Gheverghese Joseph (2000) *The Crest of the Peacock: Non-European Roots of Mathematics*, 2nd Edition. Penguin Books, 2000. ISBN 0-14-027778-1.

[3] Diploudis, Alexandros (1997) *Undusting Napier's Bones*. Heriot-Watt University, 1997 "Napier, John". Dictionary of National Biography. London: Smith, Elder and Co. 1885–1900.

[4] Bodemann, Eduard (1966) *Der Briefwechsel des Gottfried Wilhelm Leibniz in der Königlichen öffentliche Bibliothek zu Hannover* 1895, (anastatic reprint: Hildesheim, Georg Olms, 1966).

[5] Boyer, Carl B. (1959) *"III. Medieval Contributions". A History of the Calculus and Its Conceptual Development.* Dover. pp. 79–89. ISBN 978-0-486-60509-8.

[6] Divakaran, P. P. (2007, 2018) *The First Textbook of Calculus: 'Yuktibhasha'* Journal of Indian Philosophy, vol. 35, no. 5/6, 2007, pp. 417–443;
*The Mathematics of India:Concepts, Methods, Connections*Published by Hindustan Book Agency & Springer ISBN 978-981-13-1774-3.

[7] Isaac Newton (1687, 1728(in English)) *Philosophiæ Naturalis Principia Mathematica* 3 Volumes London.

[8] R. v. Eötvös (1890) *Mathematische und Naturwissenschaftliche Berichte aus Ungarn*, 8, 65.

[9] L Euler (1748) *Introductio in analysin infinitorum* Euler's identity appears in his monumental work of mathematical analysis.

[10] D'Alembert (1747) *Recherches sur la courbe que forme une corde tenduë mise en vibration* (Researches on the curve that a tense cord [string] forms [when] set into vibration), Histoire de l'académie royale des sciences et belles lettres de Berlin, vol. 3, pages 214-219.

[11] Fourier, Joseph. (1878) *The Analytical Theory of Heat.* Cambridge University Press (reissued by Cambridge University Press, 2009; ISBN 978-1-108-00178-6).

[12] Landau, L. D.; Lifshitz, E. M. (1987) *Fluid mechanics, Course of Theoretical Physics, 6 (2nd revised ed.)*, Pergamon Press, ISBN 978-0-08-033932-0, OCLC 15017127.

[13] Jackson J.D. (1999) *Classical. Electrodynamics*. Third Edition. John Wiley.

[14] Einstein, Albert (1905) *Zur Elektrodynamik bewegter Körper*, Annalen der Physik 322 (10): 891–921. English Translation 'On the Electrodynamics of Moving Bodies' by Meghnad Saha available on the web.

[15] Shannon, Claude (1948) *A Mathematical Theory of Communication Bell System Technical Journal. 27 (3): 379–423.*

[16] Einstein, Albert (1915) *Zur allgemeinen Relativitätstheorie* [On the General Theory of Relativity] Preussische Akademie der Wissenschaften, Sitzungsberichte, 1915 (part 1), 315; (part 2), 778–786, 799–801, 831–839, 844–847.

[17] Schrödinger Erwin (1926) *An Undulatory Theory of the Mechanics of Atoms and Molecules.* Physical Review. 28 (6): 1049–1070.

[18] Dirac P A M (1928) *The Quantum Theory of the Electron* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 117 (778): 610–624.

Darboux transformations and New exactly solvable potentials

C. Sivakumar[1], P. T. Varghese[2] and K. Babu Joseph[3]

[1] Department of Physics, Maharaja's college, Ernakulam,

Pin-682011, Kerala, India

thrisivc@yahoo.com

[2] Department of Physics, Rajagiri school of Engineering and Technology,

Kerala, India.

varghese@rajagiritech.edu.in

[3] Cochin University of Science and Technology,

Kochi, Kerala, India.

bb.jsph@gmail.com

Submitted on 07-Oct-2019

Abstract

A large number of exactly solvable quantum mechanical potential problems are formed using a technique called Darboux transformations(DT). The similarity between the eigen value spectra of the new and original Hamiltonian is shown. We generated exactly solvable potentials from harmonic oscillator potential and Coulomb potential and for the latter we developed a modified DT to get new potential problems.

1   Introduction

Solving the Schrodinger equation exactly for different potentials is one of the fundamental problems of quantum mechanics, some of which are studied at the undergraduate level. The intent of this paper is to present new potentials other than harmonic oscillator potential, Coulomb potential etc., for which the Schrodinger equation has exact solutions. The Darboux transformations(DT) is a classical method for generating new solutions from given solutions. Here by means of DT, we enlarge the class of potentials for which the Schrodinger equation can be exactly solved.

We know that one can construct new exactly solvable potentials of one dimensional Schrodinger equation using a method that is derived from the inverse scattering problems(based on Gelfand-Levitan equation) and represents an integral transformation

of the solutions of the initial Schrodinger equation[1, 2, 3, 4]. Here we propose another method, which is to use a differential transformation suggested by a French mathematician, Gaston Darboux, a century ago.

The fundamental results of Darboux remained unrecognized for a long time. In 1979, Matveev applied Darboux's theorem to a large class of linear partial differential equations as well as differential-difference and difference-difference linear evolution equations [5]. DT also became popular in the study of nonlinear partial differential equations [6]. In fact there is some connection between the two ways of constructing new exactly solvable models. This has been studied by Schnizer and Leeb and Samsanov in their papers [7, 8, 9]. Mielnik pointed out the use of a factorization method to construct new Hamiltonians with eigen value spectrum coinciding with original Hamiltonian [10]. This method is however a special case of DT [6, 11]. Some authors proposed a connection between DT and super-symmetric quantum mechanics [12, 13, 14, 15]. We are proposing how DT can be effectively used to generate potentials which have exact solutions other than known ones.

The plan of the paper is as follows. In section 2, we present a brief discussion of DT and show that it leads to new exactly solvable potentials for the Schrodinger equation. Sections 3 and 4 deal with two potential problems which are used for generating new solvable models. Section 4 also presents the modification needed for DT in order to apply it to the radial problem. Section 5 contains the conclusions.

2 Darboux Transformations

To introduce the technique of DT, consider one dimensional Schrodinger equation in the form

$$-\frac{\partial^2 \Psi(x)}{\partial x^2} + v(x)\Psi(x) = \lambda \Psi(x) \quad (1)$$

where $v = \frac{2mV}{\hbar^2}$ and $\lambda = \frac{2mE}{\hbar^2}$

Suppose Eq. (1) is exactly solved for a given potential and we know all of its eigen functions and eigen values for the potential $v(x)$. Let $\Psi_1(x, \lambda_1)$ be a solution of Eq. (1). Its logarithmic derivative is defined as

$$\sigma = \frac{1}{\Psi_1}\frac{\partial \Psi_1}{\partial x} \quad (2)$$

Now a DT permits one to obtain the general solution of another Schrodinger equation

$$-\frac{\partial^2 \Phi(x)}{\partial x^2} + u(x)\Phi(x) = \lambda \Phi(x) \quad (3)$$

under the DT,

$$\Phi(x) = [\frac{d}{dx} - \sigma]\Psi(x) \quad (4)$$

$$u(x) = v(x) - 2\frac{\partial \sigma}{\partial x} \quad (5)$$

where Ψ is any arbitrary solution of Eq. (1).

We can state Darboux's theorem in the following way: the Schrodinger equation, Eq. (1) is covariant with respect to the action of the DT, $\Psi(x) \rightarrow \Phi(x)$ and $v(x) \rightarrow u(x)$. Thus Eqs. (1) and (3) are exactly solvable Schrodinger equations for two different potentials $v(x)$ and $u(x)$ respectively, having

exactly the same eigen values. But it is obvious that for Eq. (3), the eigen value spectrum exactly coincides with that of Eq. (1) except for the state Ψ_1. The examples treated in the following sections show that the coincidence between the eigen values of $u(x)$ and $v(x)$ is only partial.

3 New exactly solvable models from the H.O. potential

We know that Schrodinger equation gives exact solutions for the H.O. potential, $V(x) = \frac{1}{2}kx^2 = \frac{1}{2}m\omega^2 x^2$ where k is the force constant and ω is the angular frequency of the particle of mass m. The equation for

H.O. is

$$-\frac{\partial^2 \Psi(x)}{\partial x^2} + [\frac{m^2\omega^2 x^2}{\hbar^2}]\Psi(x) = \frac{2mE}{\hbar^2}\Psi(x)$$
(6)

where E satisfies $(n + \frac{1}{2})\hbar\omega$, $(n = 0, 1, 2, ...)$ for the harmonic oscillator potential. Let us now show that DT leads new exactly solvable Schrodinger equation from Eq. (6). The new equation is,

$$-\frac{\partial^2 \Phi(x)}{\partial x^2} + u(x)\Phi(x) = \lambda\Phi(x)$$
(7)

where $u(x)$ and $\Phi(x)$ are defined by DT (4) and (5).

To define logarithmic derivative we choose, first excited state $(n = 1)$ wave function [16] of Eq. (6)

$$\Psi_1(x, \lambda = \frac{3m\omega}{\hbar}) = [\frac{m\omega}{\pi\hbar}]^{\frac{1}{4}}\sqrt{\frac{2m\omega}{\hbar}}x \exp{(-\frac{m\omega x^2}{2\hbar})}$$
(8)

Therefore,

$$\sigma = \frac{1}{\Psi_1}\frac{\partial \Psi_1}{\partial x} = \frac{1}{x} - \frac{m\omega x}{\hbar}$$
(9)

It immediately follows from DT that,

$$u(x) = \frac{m^2\omega^2 x^2}{\hbar^2} + \frac{2m\omega}{\hbar} + \frac{2}{x^2}$$
(10)

This is an anharmonic potential and DT gives exact solutions to this potential, provided various $\Psi(x)$ of Eq. (6) are given. Now we will define $\Phi(x)$, using Eq. (4) and $\Psi(x)$ is arbitrarily chosen from Eq. (6). Let us choose $\Psi(x)$ as the ground state $(n = 0)$ wave function of the harmonic oscillator

potential,

$$\Psi(x) = [\frac{m\omega}{\pi\hbar}]^{\frac{1}{4}}\exp{(-\frac{m\omega x^2}{2\hbar})}$$
(11)

Hence,

$$\Phi(x) = -[\frac{m\omega}{\pi\hbar}]^{\frac{1}{4}}\frac{1}{x}\exp{(-\frac{m\omega x^2}{2\hbar})}$$
(12)

This represents an exact solution of Eq. (7) and corresponding eigen value is $\lambda = \frac{m\omega}{\hbar}$. To check the admissibility condition on the new wave function, we use the normalization condition of the wave function, $\int_{-\infty}^{\infty}\Phi^*\Phi dx = N$. It is seen that The new wave function $\Phi(x)$ is non-normalizable

because of the essential singularity of this wave function at $x = 0$. This suggests that eventhough Eq. (7) is exactly solved, the new wave function $\Phi(x)$ is quantum mechanically unacceptable. This rules out the existence of the corresponding eigen state

$(\lambda = \frac{m\omega}{\hbar})$ from the new exactly solvable potential Eq. (10). A similar fate befalls other even order eigen states $(n = 2, 4, 6, ...)$ as well.

Now let us choose $\Psi(x)$ corresponding to third excited state $(n = 3)$,

$$\Psi(x) = [\frac{m\omega}{\pi\hbar}]^{\frac{1}{4}} \sqrt{\frac{3m\omega}{\hbar}} [x - \frac{2m\omega x^2}{3\hbar}] \exp{(-\frac{m\omega x^2}{2\hbar})} \tag{13}$$

which yields

$$\Phi(x) = [\frac{m\omega}{\pi\hbar}]^{\frac{1}{4}} \sqrt{\frac{3m\omega}{\hbar}} [-\frac{4m\omega x^2}{3\hbar}] \exp{(-\frac{m\omega x^2}{2\hbar})} \tag{14}$$

This $\Phi(x)$ is normalizable. The corresponding eigen value (λ) can be obtained from Eq. (7) on substitution of $\Phi(x)$ and $u(x)$ into that equation, which gives $\lambda = \frac{7m\omega}{\hbar}$. Similarly it can be shown that if $\Psi(x)$ chosen to be any solution of Eq. (6) corresponding to $n = 3, 5, 7, ...$, the new solutions $\Phi(x)$ are normalizable.

Thus the initial Schrodinger equation (6) for the H.O. potential leads to a new exactly solvable equation (7) for an anharmonic potential $u(x)$ given by Eq. (10). When we assume solutions, $\Psi(x)$ of initial equation, we can solve Eq. (7) to give new solutions $\Phi(x)$. Thus we have two exactly solvable potentials instead of one. Eventhough the eigen value spectrum of $u(x)$ coincides with that of H.O. potential, only those eigen states corresponding to $n = 3, 5, 7, ...$ are quantum mechanically accept-

able. Thus the coincidence of eigenvalues is only partial. In the above example we have taken first excited state wave function of Eq. (6) to form a new solvable potential $u(x)$. Similarly we can take any solution of Eq. (6) to define the new exactly solvable potential. This leads to a large class of exactly solvable potentials for the Schrodinger equation. We should mention one more point regarding the above idea. We got a new exactly solvable equation (7) from (6) provided its wave functions are known. Similarly we can form another exactly solvable model from eq. (7), when all of its solutions are known. This procedure is called the iteration of DT. This can be further extended. Thus starting from a single Schrodinger equation exactly solvable potential, we can generate a wide class of exactly solvable potentials by DT.

4 Coulomb potential (Hydrogen atom)

For a two particle central potential problem, the general solution of Schrodinger equation [17] has a radial and an angular part such as $\Psi(r,\theta,\phi) = R_{nl}(r)Y_{lm}(\theta,\phi)$. The radial part $R(r)$ satisfies radial wave equation,

$$-\frac{d^2R}{dr^2} - \frac{2}{r}\frac{dR(r)}{dr} + vR(r) = \lambda R(r) \quad (15)$$

where $v = \frac{2\mu V(r)}{\hbar^2} + \frac{l(l+1)}{r^2}$ and $\lambda = \frac{2\mu E}{\hbar^2}$; μ is the reduced mass of the two particle system and $\frac{l(l+1)}{r^2}$ is the centrifugal potential. For Hydrogen atom, Coulomb potential(in cgs system of units) is $V(r) = -\frac{Ze^2}{r}$ and $E_n = -\frac{\mu Z^2 e^4}{2\hbar^2 n^2}$, $n = 1,2,3,...$(Z is the atomic number). If we use the expression for the classical Bohr radius of the atom $a = \frac{\hbar^2}{\mu e^2}$, then

$$v(r) = -\frac{2Z}{ar} + \frac{l(l+1)}{r^2} \quad (16)$$

and

$$\lambda = -\frac{Z^2}{a^2 n^2} \quad (17)$$

4.1 Modification for Darboux transformation

In order to apply DT to Eq. (15), the usual transformation in the case of Schrodinger equation needs a slight modification. Darboux transformed exactly solvable radial equation can be written as,

$$-\frac{d^2\chi(r)}{dr^2} - \frac{2}{r}\frac{d\chi(r)}{dr} + u(r)\chi(r) = \lambda\chi(r) \quad (18)$$

where

$$\chi(r) = [\frac{d}{dr} - \sigma]R(r) \quad (19)$$

by the Darboux transformation. Where the logarithmic derivative σ is defined as

$$\sigma = \frac{1}{R_1}\frac{dR_1(r)}{dr} \quad (20)$$

$R_1(r)$ is a particular solution of Eq. (15) for $\lambda = \lambda_1$. $R(r)$ is arbitrarily chosen from the solution set of Eq. (15). ($R(r)$ can be any radial wave function of Eq. (15) except that used to define logarithmic derivative).

It can be seen that radial wave equation is covariant under the transformations $R(r) \rightarrow \chi(r)$ and $v(r) \rightarrow u(r)$ only if

$$u(r) = v(r) - 2\frac{d\sigma}{dr} + \frac{2}{r^2} \quad (21)$$

To arrive at this, substitute $\chi(r)$ from Eq. (19) into Eq. (18) and making use of Eq. (15) we get,

$$[u(r) - v(r) + 2\frac{d\sigma}{dr} - \frac{2}{r^2}]\frac{dR(r)}{dr} + [-\frac{dv(r)}{dr} + \frac{d^2\sigma}{dr^2} + \sigma v(r) + \frac{2}{r}\frac{d\sigma}{dr} - \sigma u(r)]R(r) = 0 \quad (22)$$

This equation is meaningful only if the coefficients of $\frac{dR}{dr}$ and R are identically zero, which yields the Darboux transformed potential as,

$$u(r) = v(r) - 2\frac{d\sigma}{dr} + \frac{2}{r^2} \qquad (23)$$

Thus Eq. (17) gives a new exactly solvable radial wave equation for a potential $u(r)$ under the transformations Eqs. (19) and (21). If the l values in the definition of the logarithmic derivative and in the expression for $v(r)$ are different, the potential transformation becomes inadmissible, because any intermediate modification to the potential other than that suggested by Eq. (21) does not come under the purview of Darboux theorem. This means that we cannot apply DT to an s-state wave function ($l = 0$) using the logarithmic derivative defined by another p-state wave function ($l = 1$) and vice versa.

4.2 New exactly solvable models using Coulomb potential

suppose initial radial equation (15) is exactly solved for Hydrogen atom and assume that its solutions [17] are known. Using this reference system we can construct a large number of radial equations exactly solvable potentials by using Eq. (21). Thus Eq. (18) above is an exactly solvable one and substituting for $u(r)$ and $\chi(r)$ in this equation, we can also find the value of λ. $u(r)$ and $v(r)$ have got identical eigen value spectra.

To define the logarithmic derivative, we choose $1s$ wave function ($l = 0$, $n = 1$) of hydrogen atom, which is a solution of Eq. (15) ie.,

$$R_{nl}(r, \lambda_1) = R_{1s}(r, \lambda_1) = R_{1s}(r, -\frac{Z^2}{a^2}) = 2(\frac{Z}{a})^{\frac{3}{2}} \exp\left(-\frac{Zr}{a}\right) \qquad (24)$$

where a is Bohr radius of the atom. Since $l = 0$, $v(r) = -\frac{2Z}{ar}$

$$\sigma = \frac{1}{R_{1s}}\frac{dR_{1s}}{dr} = -\frac{Z}{a} \qquad (25)$$

$$u(r) = -\frac{2Z}{ar} + \frac{2}{r^2} \qquad (26)$$

Thus DT provides a new exactly solvable radial equation Eq. (18) for a new potential Eq. (26). This can be exactly solved knowing various radial functions of Eq. (15). The solutions $\chi(r)$ of new radial equation are given

by DT Eq. (19). Let us arbitrarily choose $R(r)$ as $2s$ wave function ($n = 2$, $l = 0$) of hydrogen atom radial equation,

$$R(r) = \frac{1}{\sqrt{2}}(\frac{Z}{a})^{\frac{3}{2}}(1 - \frac{Zr}{2a}) \exp\left(-\frac{Zr}{2a}\right) \quad (27)$$

It immediately follows that

$$\chi(r) = -\frac{1}{4\sqrt{2}}(\frac{Z}{a})^{\frac{3}{2}}\frac{Z^2r}{a^2} \exp\left(-\frac{Zr}{2a}\right) \quad (28)$$

From the new radial equation Eq. (18), we will get the corresponding eigen value $\lambda =$

$-\frac{Z^2}{a^2 2^2}$. It is clear from Eq. (17) that this eigen value coincides with that of hydrogen atom radial equation. Other solutions of Eq. (18) can be obtained in a similar way, by using $R(r)$ as 3*s*, 4*s*... wave functions of hydrogen atom.

Thus a wide class of radial equations with exactly solvable potentials can be deduced from hydrogen atom radial equation. In some cases the eigen values of the new and old potentials exactly coincide and in some other cases the coincidence is only partial.

5   Conclusions

We attempted to formulate a large class of one dimensional exactly solvable potentials for the Schrodinger equation by applying Darboux transformations.   By applying DT to a reference problem, the class of potentials for which Schrodinger equation can be exactly solved can be enlarged. We can summarise the results as follows:

1. DT can be effectively used to construct an infinite number of exactly solvable systems from a known system.

2. Not all systems obtained by the Darboux procedure are quantum mechanically admissible, but a large number of valid systems can be generated.

3. Darboux transformed Hamiltonian share the eigen value spectrum of the original Hamiltonian only partially.

4. Physical systems representing new solvable potentials to be studied further.

References

[1] H E Moses (1979). J. Math. Phys. **20**, 2047

[2] P B Abraham and H E Moses (1980). Phys. Rev. A **22**, 1333

[3] M Luban and D L Pursey (1986). Phys. Rev. D **33**, 431

[4] D L Pursey (1986). Phys. Rev. D **33**, 1048

[5] V B Matveev (1979). Lett. Math. Phys. **3**, 213

[6] V B Matveev and M A Salle. *Darboux transformations and solitons*, (Springer-Verlag, New York, 1991)

[7] W A Schnizer and H Leeb (1994). J. Phys. A: Math. Gen. **27**, 2605

[8] B F Samsanov (1995). J. Phys. A **27**, 6989

[9] W A Schnizer and H Leeb (1993). J. Phys. A: Math. Gen. **26**, 5145

[10] B Mielnik (1984). J. Math. Phys. **25**, 3387

[11] A A Andrianov, N V Borisov and M J Ioffe (1984). Phys. Lett. A **105**, 19

[12] R W Haymaker and A R P Rau (1986). Am. J. Phys. **54**, 928

[13] C V Sukumar (1985). J. Phys. A:Math. Gen. **18**, 2917

[14] N Nag and R Roychoudhary (1995). J. Phys. A:Math. Gen. **28**, 3525

[15] A Valance, T J Morgan and H Bergeson (1990). Am. J. Phys. **58**, 487

[16] S Flugge. *Practical quantum Mechanics*, (Springer-Verlag, New York, 1971)

[17] P M Mathews and G Venkatesan. *A text book of quantum mechanics*, (Tata McGraw-Hill: New York, 1976)

Simulating Boundary Behaviour of Transverse Waves in Strings using SCILAB

Pragati Ashdhir, Nishant Goyal and Mritunjay Tyagi

Department of Physics, Hindu College, University of Delhi, Delhi-110007, India

pragatiashdhir@hinducollege.ac.in

Submitted on 14-09-2019

Abstract

The boundary effects are integral to any kind of wave motion whenever travelling waves encounter an obstacle or a change in the medium of propagation. They are known to be responsible for some of the well known phenomena like echoes, radar detection or sound production in musical instruments. In this work simulation of the boundary behaviour of transverse waves along a string is done. The effects arising due to rigid boundary conditions, free boundary conditions, real (neither free nor rigid) boundary conditions and impedance discontinuity (change in medium) conditions have been discussed.These simulations can be used as visual aids or tool kits to make the traditional methods of teaching and learning of the underlying physics concepts more effective. SCILAB, an open source computational software has been used for numerical calculation, simulation and animation as needed. The target group of the present work are the students and

teachers of the introductory undergraduate level Physics.

---

1 Introduction

The pedagogical merits of making computer programming and modelling an integral part of physics education has been recognized long back [1]. By integrating computational physics techniques with the traditional teaching-learning methods of physics, a major improvement in the conceptual understanding and problem solving skills of students can be expected [1,2,3].It it is needed to introduce programming and numerical techniques early in the course of physics education.

To begin with, the students can be made to solve problems using both analytical as well as computational techniques. These problems can be picked from anywhere in the curriculum of the physics course being pursued by them. The next step can be

to generalise these problems in such a way that solving them analytically becomes either very tedious or almost impossible and the only option left with students is to adopt a computational approach. Consequently, they would be able to understand the limitations of analytical methods and appreciate the importance of computational methods. Such an insight would motivate the students to probe deeper into the physics concepts being taught to them and train them to solve more realistic and research oriented problems. The adoption of this kind of teaching-learning approach could go a long way in inspiring the students to choose a future in physics.

In this work it is attempted to illustrate how one can use a numerical approach for a better visualization and understanding of the boundary behaviour of transverse waves in strings as compared to a purely analytical treatment of the phenomena that is mostly found in textbooks. A simulation is done of how the propagation of transverse waves along a string would be affected by the conditions that exist at is two ends.

The paper begins with a discussion on modelling of the physical phenomenon of wave propagation along a string and the computational approach used in developing the relevant computer codes. Subsequently, the reflection of transverse waves at rigid and free boundaries of the string are discussed. In real strings such as those found on musical instruments, the waves encounter boundaries that are neither per-

fectly rigid nor perfectly free. The interaction of waves with such 'real' boundaries form the next part of the discussion. It is followed by a discussion on the response of transverse waves along a string to an impedance discontinuity , i.e. , when the waves see a change of medium .One of the most important phenomenon arising due to boundary conditions in strings, the formation of stationary waves is simulated in the last section of the paper. The simulation is done using SCILAB, an open source numerical computational package. Programming in SCILAB is much easier and concise in comparison to other high level languages like FORTRAN, C or C++ used in scientific computing.The simplicity of programming in SCILAB is attributed to its matrix based computation, dynamic typing and automatic memory management features. It is the ease of using SCILAB even for people without prior programming experience that enable the computer codes developed in this work to be used as tool kits for visualization and an enhanced understanding of transverse wave propagation on a string as a bounded medium. Minor modifications in the codes can be done quite easily to explore those aspects of the phenomenon whose analytical treatment can become challenging.

2 Modelling Transverse Wave Propagation along a String

Wave propagation in a medium is attributed to the cohesive forces binding the neigh-

bouring particles of the medium and to the elastic forces developed in the medium by the moving disturbance. The mathematical description of a wave is given by the classical wave equation, the one-dimensional version of which can be written as

$$\frac{\partial^2 u(x,t)}{\partial t^2} = v^2 \frac{\partial^2 u(x,t)}{\partial x^2}, \qquad (1)$$

where the wave propagation takes place along x-axis. u(x,t) is the instantaneous displacement of the medium particle located at x and v is the velocity of wave propagation or the wave velocity. Equation (1) is a linear partial differential equation.

Whatever the waves, the wave velocity can always be expressed as a function of "elasticity" or potential energy storing mechanism in the medium and the "inertia" of the medium through which its kinetic energy is stored. When a continuous string stretching from x=0 to infinity offers itself as the medium of wave propagation, the wave velocity is given by

$$v = \sqrt{\frac{T}{m}}, \qquad (2)$$

where T is the tension in the string and m is its linear density , i.e. mass per unit length of the string. If the string is homogeneous, then both T and m are constants in space and time. It can be easily shown that any function of the form

$$u(x,t) = f_1(vt - x), \qquad (3)$$

$$\text{or } u(x,t) = f_2(vt + x) \qquad (4)$$

will be a solution of the wave Eq. (1). Equation (3) represents a wave travelling along the positive x-axis, while Eq. (4) is for a wave moving in the negative x-direction. Further, since Eq. (1) is linear, the principle of superposition holds good for all its solutions. Thus, the complete solution of the classical wave equation can be written as,

$$u(x,t) = f_1(vt - x) + f_2(vt + x). \qquad (5)$$

The one-dimensional propagation of waves along the string under different physical situations is modelled by the wave Eq. (1) subject to appropriate boundary and initial conditions.

3   The Computational Approach

The Finite Difference Method has been used to solve the given wave equation. In this method the derivatives occurring in the equation and in the boundary/initial conditions are replaced by their finite difference approximations, thereby, transforming the differential equation to an equivalent difference equation.

To set up the difference equation, a (N+2)×(M+2) dimensional rectangular grid in space and time is chosen. The boundaries of the grid are decided by the range of x and t variables over which the solution to the problem is sought. If x ϵ [x_0,x_f] and t ϵ [t_0,t_f], then the grid points along the space and time axes are given by

$$x_i = x_o + (i-1)h, \quad \text{where } i = 1, 2, \ldots, N+2 \qquad (6)$$

$$t_j = t_o + (j-1)l, \quad \text{where } j = 1, 2, \ldots, M+2 \qquad (7)$$

such that $x_{N+2} = x_f$ and $t_{M+2} = t_f$. The lines $x = x_o$, $x = x_f$, $t = t_o$ and $t = t_f$ form the grid boundaries, h and l are the step sizes for defining the grid points along the x-axis and t-axis respectively. Using the forward difference approximation, the finite difference representation of the given one-dimensional wave equation can be easily shown to be

$$\frac{u(i,j+1) - 2u(i,j) + u(i,j-1)}{l^2} = v^2 \left[\frac{u(i+1,j) - 2u(i,j) + u(i-1,j)}{h^2} \right] \tag{8}$$

where, u(i,j) = u(x_i, t_j). Solving Eq. (8) for u(i,j+1), we get,

$$u(i,j+1) = \frac{l^2 v^2}{h^2} \left[u(i+1,j) + u(i-1,j) \right] + 2 \left[1 - \frac{l^2 v^2}{h^2} \right] u(i,j) - u(i,j-1). \tag{9}$$

From Eq. (9) it can be seen that if u is known at all x_i at the times t_j and t_{j-1} then u at all x_i at the next time step, t_{j+1}, can be determined. This is an explicit method for determining the solution.

It is to be mentioned that the size of finite differences determine the accuracy of a solution. If these differences are made smaller the accuracy of the solution is expected to improve. However, the extent to which the accuracy of a solution can be improved needs to take into account its "stability". The "stability" of a solution implies its meaningfulness, i.e. the solution should be physically significant. In the finite difference approximation of the given wave Eq. (1), the condition for stability of its solution can be shown to be

$$\left| \frac{vl}{h} \right| \leq 1 . \tag{10}$$

Thus, to ensure solution stability the step sizes h and l are chosen such that they satisfy the condition

$$\frac{v^2 l^2}{h^2} = 1 . \tag{11}$$

Substituting Eq.(11) in Eq.(9), we get

$$u(i,j+1) = u(i+1,j) + u(i-1,j) - u(i,j-1) \tag{12}$$

as the required recursion relation.

Now, since we have $i = 1, 2, 3, ..., (N+$

2) and $j = 1, 2, 3, ..., (M + 2)$, the solution of difference Eq.(12) is equivalent to solving a system of algebraic linear equations. To solve the system of equations in the present case, Gauss-Siedel method has been chosen because of its iterative approach and faster convergence.The difference Eq.(12) along with appropriate boundary conditions and initial values are thus transformed into a suitable SCILAB code.

The algorithm used for writing the code is:

An initial guess is made for $u(i, j)$ at all interior points of the grid.

Equation (12) is used to compute $u^m(i, j)$ at all interior points. The index m is the iteration number.

If the prescribed convergence threshold is reached, iteration is stopped, else it is continued.

For the given iteration, the new value of u is updated.

The control is transferred to step 2.

In the SCILAB code for Gauss-Siedel method as seen in the screenshot in Fig.1, a variable "done" is declared to be a logical variable and is set to be TRUE (done=%t) at the start of every iteration. As the iteration proceeds, the accuracy of computed $F(i, j)$ [variable used in the code for $u(i, j)$] at each grid point is tested. If the error exceeds the specified accuracy or tolerance, the variable "done" is set to FALSE (done=%f) and the next iteration starts. A provision is also made to exit the iterative WHILE LOOP if the number of iterations become too large.

The screenshot in Fig.2 shows a portion of the code used for animating the simulations.These animations have also been converted to .avi and .gif formats.There are 11 animations in all that depict the results of the present work. These files have been referenced as Animation 1, Animation 2...Animation 11 at appropriate places through out the text. The .gif format of these animations can be accessed at the following link.

https://drive.google.com/drive/folders/1KDhGVYoCIPpy4eWLR5-k6Xwm5Zz1IX4w

These gif files have been created using the open source Imagemagick software.

4   Rigid and Free Boundary

In our aim to aid the understanding of the phenomenon of reflection of waves at rigid and free boundaries we simulate the following physical situation : A meter long string having a mass of 1g is under a tension of 10 N with its one end rigidly fixed and the other end left free. Initially the string is deformed so that it has a "bump" in the middle and is motionless at time t=0. The subsequent motion of the string is to be observed.

The first step in simulation is to write

Figure 1. Screenshot of SCILAB code for Gauss-Seidel method

Figure 2. Screenshot of SCILAB code uesd for animation

the mathematical expression for the initial "bump" function. Let us take the "bump" to be having a Gaussian waveform centred half way along the length of the string at x= L/2 = 0.5 m. Thus, the required expression can be written as,

$$u(x,t) = \begin{cases} 0 & , x = 0, \\ e^{-A\,(x-0.5)^2} & , 0 < x < L, \\ 0 & , x \geq L, \end{cases} \quad (13)$$

where A is some constant and L is the length of the string equal to 1 m in this case. Next, Eq. (1) is solved using Eq. (13) as one of the initial conditions. The other initial condition is arrived at by the consideration of the string being motionless at t=0. This manifests itself in the time derivative of the string's transverse displacement to be zero at t=0, i.e.,

$$\frac{\partial u(x,t)}{\partial t}t = 0 = 0 . \quad (14)$$

The discretization of the above initial condition is done by using central difference approximation as follows

$$\frac{\partial u(x_i,t_j)}{\partial t} = \frac{u(x_i,t_{j+1}) - u(x_i,t_{j-1})}{l} . \quad (15)$$

For t=0, we have j=1 and therefore

$$\frac{\partial u(x_i,t_j)}{\partial t}t = 0 = \frac{u(i,2) - u(i,0)}{l}. \quad (16)$$

Substituting Eq. (14) in Eq. (16), we get

$$u(i,2) = u(i,0). \quad (17)$$

Rewriting Eq. (12) for j = 1, we get

$$u(i,2) = u(i+1,1) + u(i-1,1) - u(i,0). \quad (18)$$

By eliminating u(i,0) between Eq.(17) and Eq.(18), we can write

$$u(i,2) = \frac{u(i+1,1) + u(i-1,1)}{2}. \quad (19)$$

Equation (19) is used to compute transverse displacement at various i values for j=2. For $j > 2$, Eq.(12) computes the transverse displacement at the remaining grid points.

The next step is to incorporate the boundary conditions. A rigid boundary is characterized by zero displacement at all times while a free boundary always remains unstrained. In the simulation, the ends x=0 and x=L of the string are taken to be free and rigid boundaries respectively. As the wave pulse approaches the rigidly fixed end at x = L, i.e., u(L,t)=0, the internal restoring forces which allow the wave pulse to propagate exert an upward force on this end of the string. But since the end is rigidly fixed, it does not move. According to Newton's third law, the string support at this end must be exerting an equal downward force on the end of the string. This force creates a wave pulse that propagates from right to left, with the same speed and amplitude as the incident pulse but with opposite polarity. Now, for the wave pulse moving leftwards towards the free end at $x = 0$, the net vertical force at this end of the string must be zero, i.e., we should have

$$T\frac{\partial u(x,t)}{\partial x}x = 0 = 0 . \quad (20)$$

This boundary condition is thus mathematically equivalent to requiring that the slope of the string displacement be zero at the free

end. Hence, in our case

$$\frac{\partial u(x,t)}{\partial x} x = 0 = 0 \qquad (21)$$

is the required boundary condition. The discrete formulation of Eq.(21) using forward difference approximation is

$$\frac{u(2,j) - u(1,j)}{h} = 0 \qquad (22)$$

giving

$$u(1,j) = u(2,j) \qquad (23)$$

for all values of j corresponding to all instants of time. The reflected wave pulse propagates from left to right with the same speed and amplitude as the incident wave and also with the same polarity.

Animation 1 is the result of executing the corresponding SCILAB code. The initial waveform can be seen to dissociate into two identical Gaussian pulses moving in opposite directions towards the two ends of the string.Physically, the observation can be explained on the basis of the law of conservation of linear momentum. Since the initial momentum of the string is zero, the "bump" would evolve with time in a manner that the total momentum of the string remains zero at all later instants. It can also be seen that a crest is reflected as a trough at the rigid boundary while it remains a crest after reflection at the free boundary.

Figure 3 depicts the instantaneous configuration of the string as the two Gaussian pulses propagate along the string and get reflected at it ends. The reflected pulses then superimpose almost destroying each other

at $t \simeq 0.01s$. At $t \simeq 0.02s$, they exhibit complete constructive interference yielding a pulse of the same amplitude as the initial pulse but with opposite polarity. In this simulation the values of the various parameters like the simulation time, tension in the string, mass per unit length and the length of the string can be easily changed in the given SCILAB code. Such changes can be used for further exploration of the phenomenon, for example, if at all then how do the instants of complete constructive and destructive interference of pulses get affected by any of these parameters. To view the configuration of the string at instants other than those shown in Fig.3, modifications can be made in the plotting part of the code.By changing the values of tension and/or mass per unit length the pulses can be made to propagate at different speeds.

5   Real Boundary

In real strings such as those found on musical instruments, the supports lie somewhere between being perfectly rigid and being perfectly free. Most stringed instruments produce sound through the application of energy to the strings either by striking, plucking or bowing them, which sets them into vibratory motion creating musical sounds. The strings alone however, produce only a faint sound that needs to be amplified to be heard. This is accomplished by a device called "bridge" that essentially supports the strings on a stringed musical instrument. It
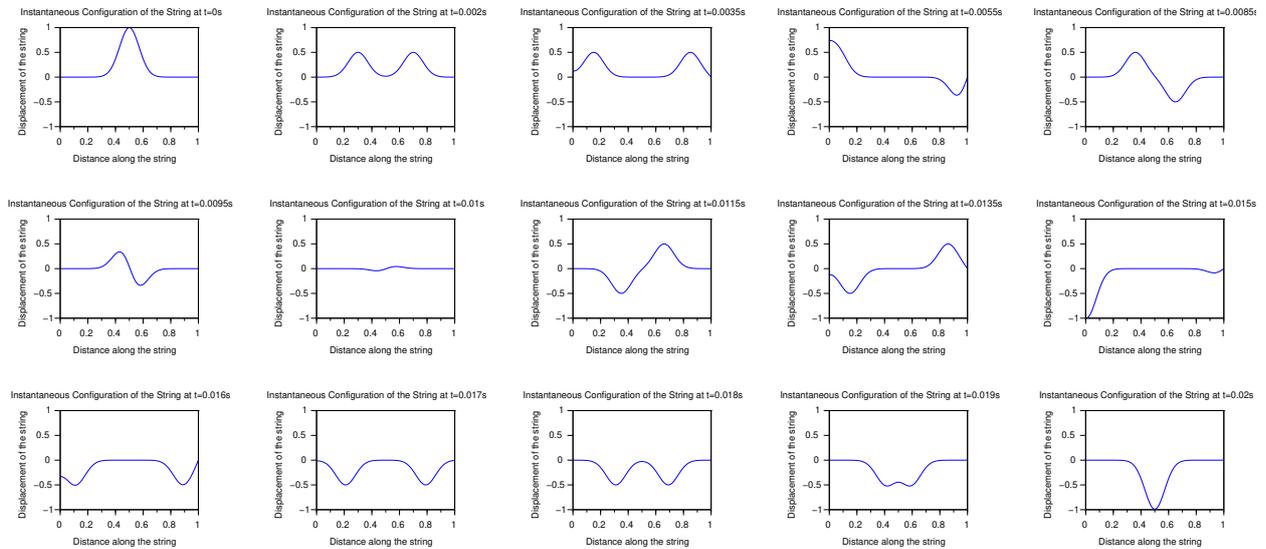
Figure 3. Reflection at Free Boundary at x = 0 and a Rigid Boundary at x = L : T = 10N, L = 1m, $m = 0.001 Kgm^{-1}$. The free end gets displaced with the string and the rigid end remains fixed. String Configuration at t = 0s, 0.002s, 0.0035s, 0.0055s, 0.0085s, 0.0095s, 0.01s, 0.0115s, 0.0135s, 0.015s, 0.016s, 0.017s, 0.018s, 0.019s & 0.02s. At t = 0.0085s, a crest remains a crest on reflection at x = 0 which is a free boundary and a crest becomes a trough on reflection at x = L which is a rigid boundary.

transmits the vibration of those strings to another structural component of the instrument, typically a "sound board" that produces louder sounds. Thus, the bridge is a kind of support that not only allows the strings to vibrate freely but also conducts those vibrations efficiently to the larger surface of the sound board. If the string support were truly rigid, the sound board would not vibrate and we would not be able to hear the sound of the stringed instrument being played.

To visualize the above behaviour of real strings, we model the support as having mass, M_s. The free end would then be cor-

respond to a support having zero mass and the rigid end would correspond to an infinitely massive support. In the simulation, we take the end x=0 of the string to be a real boundary and the end x=L continues to be rigid. The initial deformation of the string is once again assumed to be in the form of Gaussian pulse given by Eq.(13). If M_s is taken to be the effective mass of the support at x=0, then, applying Newton's second law to the support, we get

$$M_s \frac{\partial^2 u(x,t)}{\partial t^2} x = 0 = T \frac{\partial u(x,t)}{\partial x} x = 0 . \quad (24)$$

The finite difference approximation used for the above condition at the boundary is
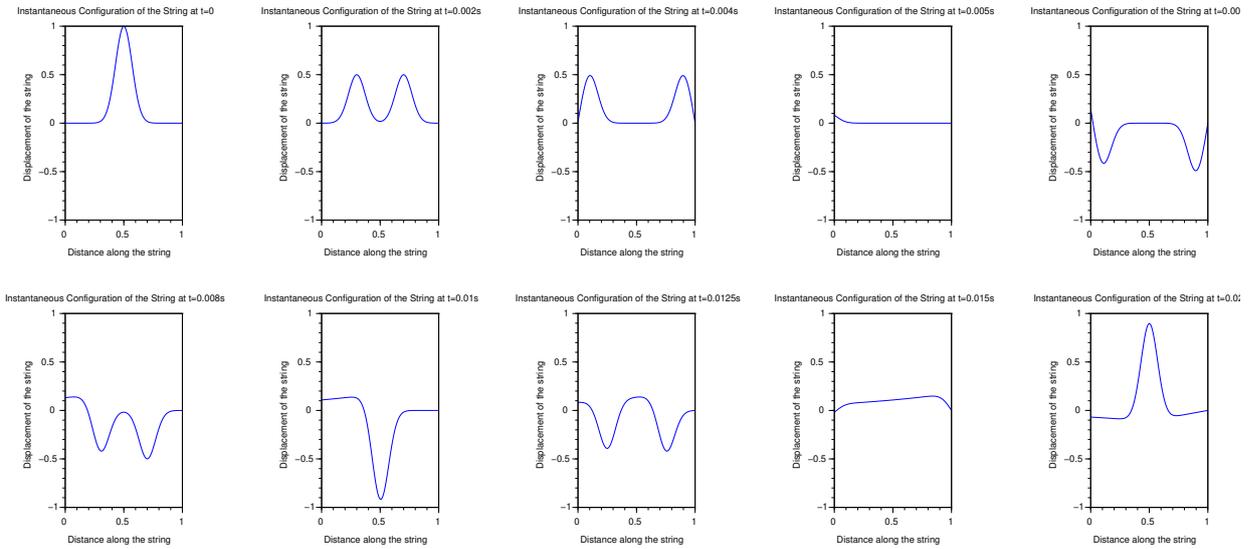
Figure 4. Reflection at Real Boundary at x = 0 and a Rigid Boundary at x = L : T = 10N, L = 1m, m = $0.001 Kgm^{-1}$, Support mass, $M_s = 1g$. The support is equal to the mass of the string so the real boundary oscillates about its mean position. String Configuration at t = 0, 0.002s, 0.004s, 0.005s, 0.006s, 0.008s, 0.01s, 0.0125s, 0.015s, 0.02s.
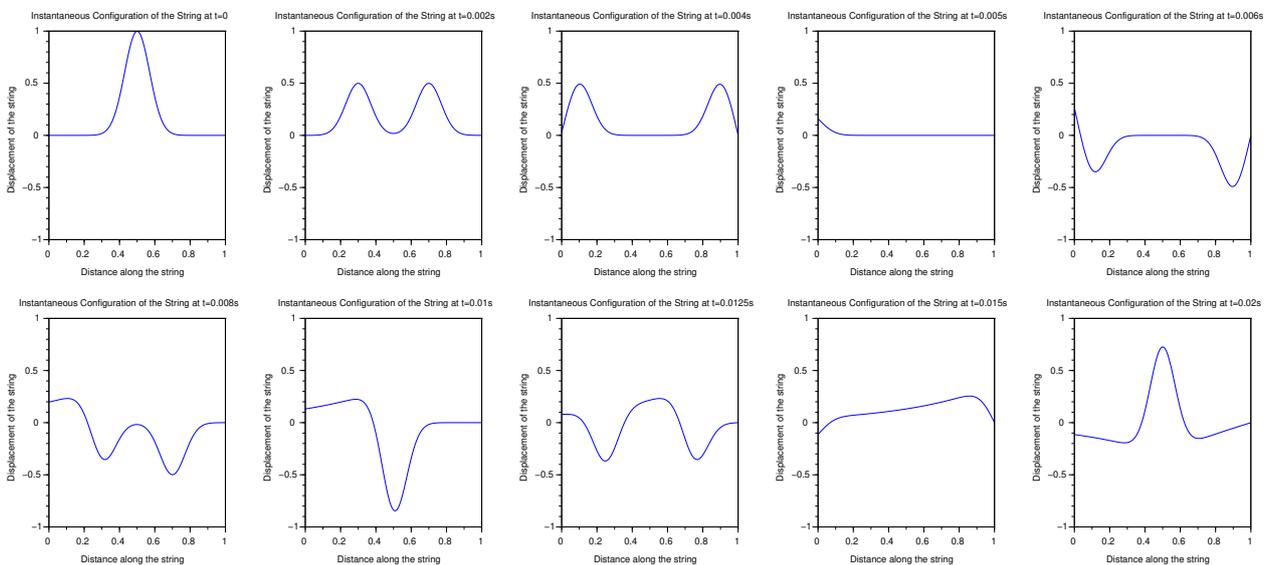
Figure 5. Reflection at Real Boundary at x = 0 and a Rigid Boundary at x = L : T = 10N, L = 1m, m = $0.001 Kgm^{-1}$, Support mass, $M_s = 0.5g$. The support is of mass comparable to the mass of the string so the real boundary oscillates about its mean position. String Configuration at t = 0, 0.002s, 0.004s, 0.005s, 0.006s, 0.008s, 0.01s, 0.0125s, 0.015s, 0.02s.

$$u(1,j+1) = \frac{Tl^2}{M_s h} u(2,j) + \left[2 - \frac{Tl^2}{M_s h}\right] u(1,j) - u(1,j+1) \qquad (25)$$

From Eq. (11), the condition for stability of the solution can be written as

$$\frac{Tl^2}{mh^2} = 1 , \qquad (26)$$

which is re-arranged as,

$$\frac{Tl^2}{h} = hm . \qquad (27)$$

Substituting Eq. (27) in Eq. (25), we get,

$$u(1,j+1) = \frac{mh}{M_s} u(2,j) + \left[2 - \frac{mh}{M_s}\right] u(1,j) - u(1,j-1). \qquad (28)$$

Equation (28) is the required discretized version of the boundary condition for a "real" support at x=0. This equation is looped through j=2 to j=M+1 to compute the displacement of the support at instants of time corresponding to j=3 upto j=M+2. To compute u(1,2), the initial condition corresponding to the time derivative of transverse displacement is used. This condition is given by

$$\frac{\partial u(x,t)}{\partial t} t = 0 = 0 . \qquad (29)$$

Using the forward difference approximation for discretizing Eq.(29), we get

$$\frac{u(i,j+1) - u(i,j)}{l} = 0 . \qquad (30)$$

Putting j=1 for t=0 and i=1 for the boundary at x=0, we get the required expression as

$$u(1,2) = u(1,1) . \qquad (31)$$

In the present simulation of a real support the mass of the support (M_s) is given different values 50g, 5g, 1g, 0.5g, 0.05g and 0.005g. The length of the string is taken to be 1 m with its mass per unit length equal to 0.001 Kg/m. When the mass of the support is 1g or 0.5g , i.e., comparable to the total mass of the string, it can be seen in Animation 2 and Animation 3 respectively that the support vibrates transversely about its equilibrium position at x=0. Thus, the support behaves like a real boundary. In Animation 4 when the mass of the support is 5g, the amplitude of its transverse vibration decreases and the support starts behaving like a rigid boundary. For a further increase in the support mass to 50g in Animation 5, the support or the end x=0 of the string does not move at all and hence behaves like a perfectly rigid boundary. On the other hand, in Animation 6 when the support mass is equal

to 0.05g so that it becomes quite lighter than the string itself, the support is seen to get displaced along with the string and the end x=0 starts behaving like a free boundary. In Animation 7, the support is made even more lighter with mass equal to 0.005g and it behaves like a completely free boundary. Figures 4 to 9 depict the instantaneous configurations of the string in each of the above cases respectively.

6 Impedance Discontinuity

There is another kind of boundary effect that arises due to the "Characteristic Impedance" offered by any medium to waves travelling across it. For a lossless medium, this impedance is determined by the two energy storing parameters, inertia and elasticity. It can easily be shown that the characteristic impedance of a string is given by

$$Z = (Tm)^{1/2} . \qquad (32)$$

To illustrate how the travelling waves would respond to a sudden change of impedance we simulate a physical situation in which a string with a certain tension (T) and mass per unit length (m_1) is smoothly fastened to another string under the same tension (T) but with a different mas per unit length (m_2). The two strings are taken to be of equal length of 1 m each. The junction between the two dissimilar strings is taken at x=0. The first string extends from x= - L to x=0 and the second string from x=0 to x=L, where $| L | = 1m$. The non-junction ends of the two strings at x= -L and x= L respectively are assumed to be rigidly fixed. The specific impedances of the two strings are

$$Z_1 = (Tm_1)^{1/2} \qquad (33)$$

$$and \ Z_2 = (Tm_2)^{1/2} \qquad (34)$$

respectively. Thus, a wave travelling across the given arrangement of strings would see an impedance discontinuity at the junction, x=0.

In the present simulation, a Gaussian waveform centred in the middle of the first string (at x=-0.5 m) and moving right towards the junction (at x=0) is taken. Mathematically, such a waveform can be expressed as

$$u(x,t) = A[exp^{-100(x-v_1t+0.5)^2}] , \qquad (35)$$

where A is the amplitude of the Gaussian waveform and $v_1 = \sqrt{\frac{T}{m_1}}$ is the velocity of wave propagation along the first string. Using Eq. (35) we can write

$$u(x,0) = A[exp^{-100(x+0.5)^2}] = R(x) \ (say) , \qquad (36)$$

$$\frac{\partial u(x,t)}{\partial t} t = 0 = -200Av_1(x+0.5)exp^{-100(x+0.5)^2} = H(x) \ (say) , \qquad (37)$$

$$\frac{\partial u(x,t)}{\partial x} x = 0 = -200A(-v_1t+0.5)exp^{-100(-v_1t+0.5)^2} = G(t) \ (say) , \qquad (38)$$

Figure 6. Reflection at Real Boundary at x = 0 and a Rigid Boundary at x = L : T = 10N, L = 1m, m = $0.001 Kgm^{-1}$, Support mass, $M_s = 5g$. The support mass is quite large as compared to the mass of the string and the real boundary starts behaving like a rigid boundary. String Configuration at t = 0, 0.002s, 0.004s, 0.005s, 0.006s, 0.008s, 0.01s, 0.0125s, 0.015s, 0.02s. At t = 0.008s, it can be seen that the pulse reflected at x = 0 has opposite polarity w.r.t. the incident pulse. The support vibrates about its mean position with small displacement.

and

$$\frac{\partial^2 u(x,t)}{\partial x^2} x = 0 = -200 A exp^{-100(-v_1 t + 0.5)^2}[1 - 200(-v_1 t + 0.5)^2] = Q(t) \; (say) \,. \qquad (39)$$

The initial conditions given by Eqn.(36) and Eqn.(37) are valid for any point along the first string excluding its end at x=-L i.e., $-L < x <= 0$. The discrete formulation of Eqn.(36) can be written as

$$u(i,1) = R(i), \qquad (40)$$

where j=1 corresponds to t=0 and the index i takes up values corresponding to all grid points in the range $-L < x <= 0$. The other initial condition given by Eqn.(37) is

discretized using Eqn.(16) as follows,

$$\frac{u(i,2) - u(i,0)}{2l} = H(i). \qquad (41)$$

Eliminating u(i,0) between Eqn.(18) and Eqn.(41), we get

$$u(i,2) = \frac{u(i+1,1) + u(i-1,1)}{2} + lH(i), \qquad (42)$$

where the index i takes up all values corresponding to the grid points in the range $-L < x <= 0$. The boundary conditions
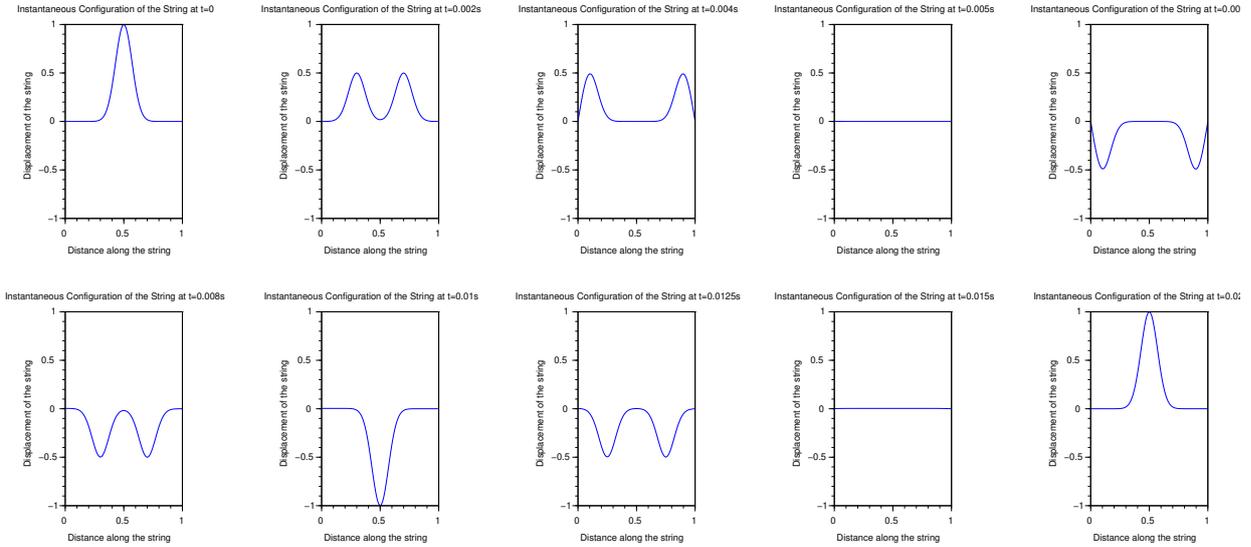
Figure 7. Reflection at Real Boundary at x = 0 and a Rigid Boundary at x = L : T = 10N, L = 1m, m = $0.001 Kgm^{-1}$, Support mass, $M_s = 50g$. The support is very massive as compared to the string so the real boundary behaves like a rigid boundary. String Configuration at t = 0, 0.002s, 0.004s, 0.005s, 0.006s, 0.008s, 0.01s, 0.0125s, 0.015s, 0.02s. At t = 0.008s, it can be seen that the pulse reflected at x = 0 has opposite polarity w.r.t. the incident pulse.

that must be satisfied at the impedance discontinuity, x=0, are:

1. A geometrical condition that the transverse displacement is continuous at the junction for all time. This condition is incorporated by using Eq.(38) to ensure the existence of first order derivative of u(x,t) w.r.t. x at the junction (x=0), hence implying the continuity of u(x,t) across the junction. The forward difference approximation of Eq.(38) is

$$\frac{u(i+1,j) - u(i,j)}{h} = G(j), \qquad (43)$$

for all values of j and the value of index i corresponding to the junction at x=0.

2. A dynamical condition that the trans-

verse force $T(\frac{\partial u}{\partial x})$ is continuous at x=0. This must always hold, else, a finite difference in the force acting on an infinitesimally small mass of the string at the junction would accelerate it infinitely, which cannot be allowed. The incorporation of this condition is done by using Eq. (39) to ensure the existence of second order derivative of u(x,t) w.r.t. x at the junction (x=0), thereby implying the continuity of $\frac{\partial u(x,t)}{\partial x}$ or $T \frac{\partial u(x,t)}{\partial x}$ across the junction. The finite difference approximation of Eq. (39) is

$$\frac{u(i+1,j) - 2u(i,j) + u(i-1,)}{h^2} = Q(j), \tag{44}$$

for all values of j and the value of
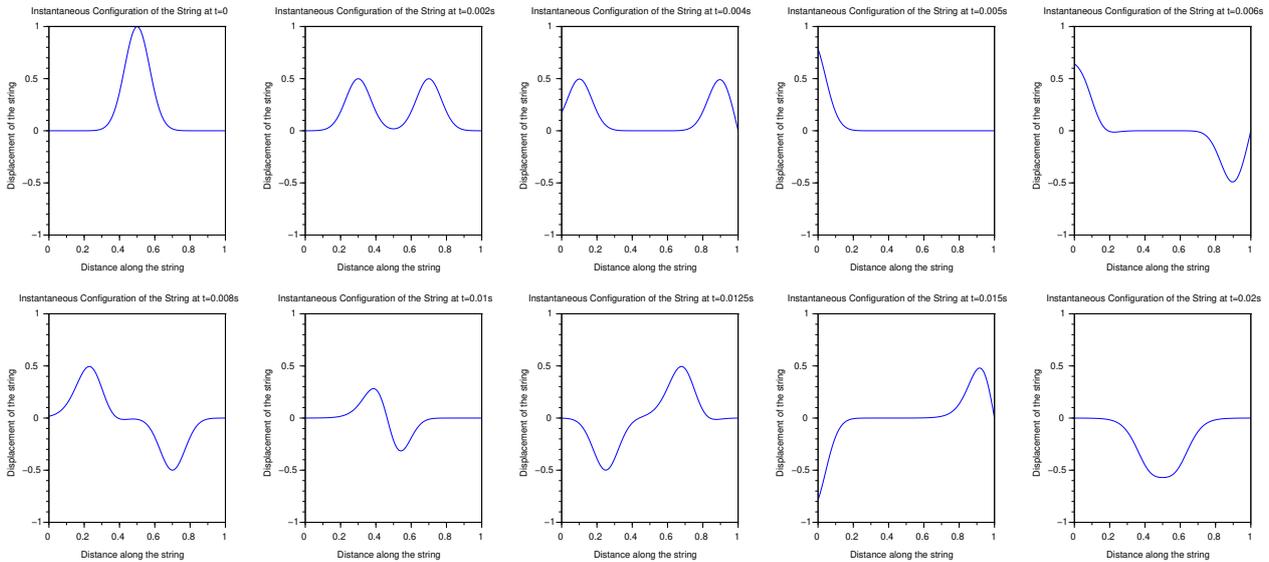
Figure 8. Reflection at Real Boundary at x = 0 and a Rigid Boundary at x = L : T = 10N, L = 1m, $m = 0.001 Kgm^{-1}$, Support mass, $M_s = 0.05g$. The support mass is comparable to the total mass of the string and the real boundary oscillates about its equilibrium position. String Configuration at t = 0, 0.002s, 0.004s, 0.005s, 0.006s, 0.008s, 0.01s, 0.0125s, 0.015s, 0.02s. At t = 0.008s, it can be seen that the pulse reflected at x = 0 has the same polarity as the incident pulse.

index i corresponding to the junction at x=0. There are two more boundary conditions that correspond to the non-junction ends of the two strings to be rigidly fixed, i.e., u(-L,t)=0 and u(L,t)=0. Animation 8, where T=10 N, m_1=0.001Kg/m, m_2=0.004Kg/m and Animation 9, where T=10 N, m_1=0.002Kg/m, m_2=0.001Kg/m depict the behaviour of the wave pulse as it approaches the junction of the strings. In both the cases it can be seen that the incident pulse is partly reflected and partly transmitted. If A, A_r and A_t are incident, reflected and transmitted amplitudes respectively, then it can be

shown analytically that

$$\frac{A_r}{A} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \qquad (45)$$

and

$$\frac{A_t}{A} = \frac{2Z_1}{Z_1 + Z_2} . \qquad (46)$$

In Figure 10 the instantaneous configurations of the string corresponding to Animation 8 are shown. Using Eq.(33) and Eq.(34), we get $Z_1 = 0.1 Kgs^{-1}$ and $Z_2 = 0.2 Kgs^{-1}$. Substituting these values of Z_1 and Z_2 in Eq.(45) and Eq.(46), the reflection coefficient of amplitude is calculated to be

$$\frac{A_r}{A} = -\frac{0.1}{0.3} = -\frac{1}{3} \qquad (47)$$
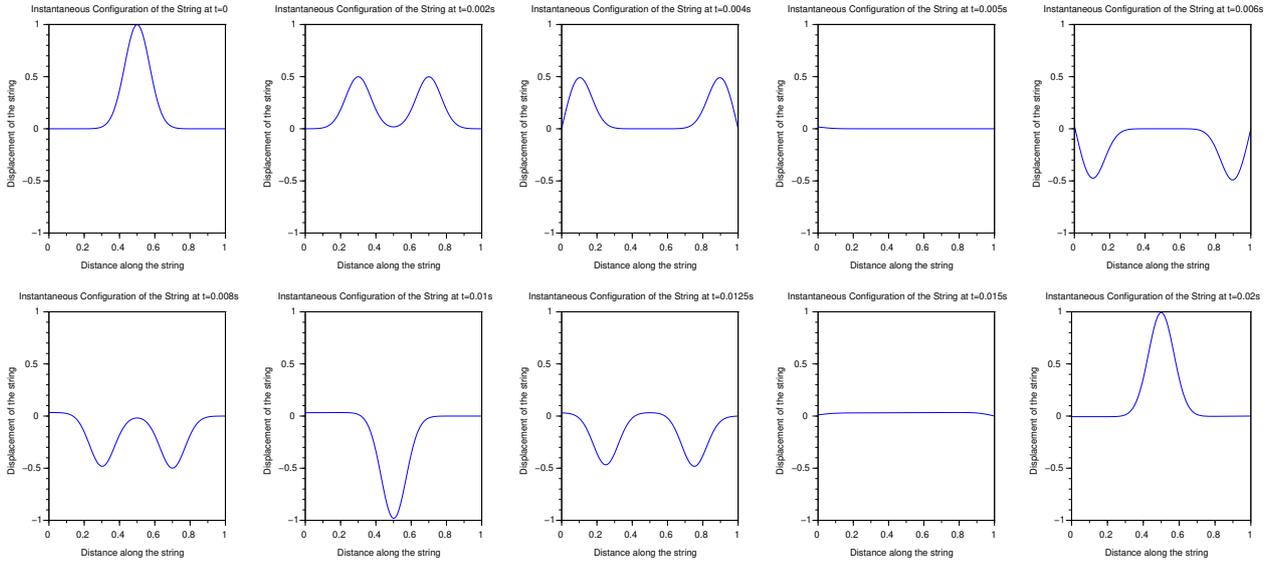
and the transmission coefficient of ampli-
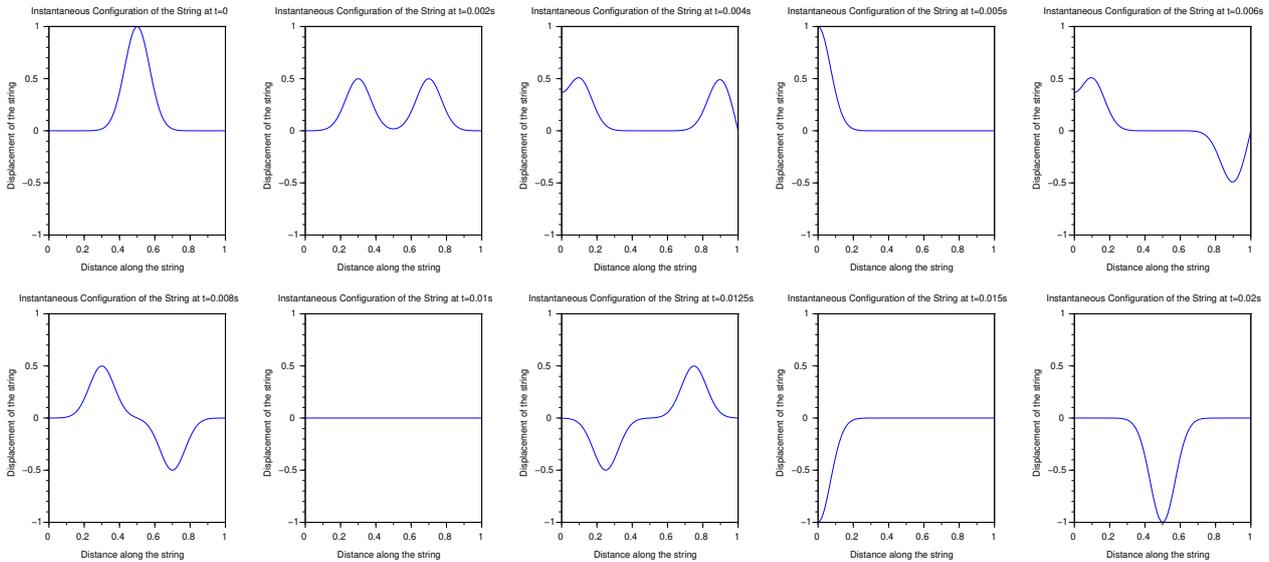
Figure 9. Reflection at Real Boundary at x = 0 and a Rigid Boundary at x = L : T = 10N, L = 1m, m = 0.001Kgm^{-1}, Support mass, M_s = 0.005g. The support mass is much less than the total mass of the string so the real boundary behaves like a free boundary. String Configuration at t= 0, 0.002s, 0.004s, 0.005s, 0.006s, 0.008s, 0.01s, 0.0125s, 0.015s, 0.02s. At t = 0.008s, it can be seen that the pulse reflected at x = 0 has the same polarity as the incident pulse.

tude comes out to be

$$\frac{A_t}{A} = \frac{0.2}{0.3} = \frac{2}{3}. \qquad (48)$$

The configurations of the string at $t = 0.008s$ and $0.01s$ in Fig. 10 verify the above calculated values of reflection and transmission coefficients. The negative sign of the reflection coefficient is verified by the observed change in polarity of the reflected waveform with respect to the incident waveform. Further, since $m_1 < m_2$, we have $v_1 = (T/m_1)^{1/2} > v_2 = (T/m_2)^{1/2}$, i.e., the reflected pulse moves faster than the transmitted one. The exact speeds of propagation of the reflected and transmitted wave forms can also be verified by the plots in Fig. 10. v_1 can be computed from the plots correspond-

ing to $t = 0.008s$ and $0.01s$. In these plots, the x-coordinates of the left end of the reflected pulse are $x_1 = -0.532m$ and $x_2 = -0.727m$. Using these we get, $v_1 = \frac{x_2-x_1}{t_2-t_1} = \frac{-0.0.727+0.532}{0.01-0.008} = \frac{-0.195}{0.002} = -97.5ms^{-1}$. So there is a fairly good agreement between the computed v_1 and the theoretical value of $v_1 = \sqrt{\frac{T}{m_1}} = \sqrt{\frac{10}{0.001}} = 100ms^{-1}$. The negative sign in the computed value is indicative of the reflected pulse moving leftwards along the negative x-axis. To compute v_2, we use the plots corresponding to $t_1 = 0.01s$ and $t_2 = 0.0125s$ in Fig. 10. The x-coordinates of the right end of the transmitted pulse at these instants are $x_1 = 0.412m$ and $x_2 = 0.527m$. So, we get, $v_2 = \frac{0.527-0.412}{0.0125-0.01} = \frac{0.115}{0.0025} = 46ms^{-1}$, which is also
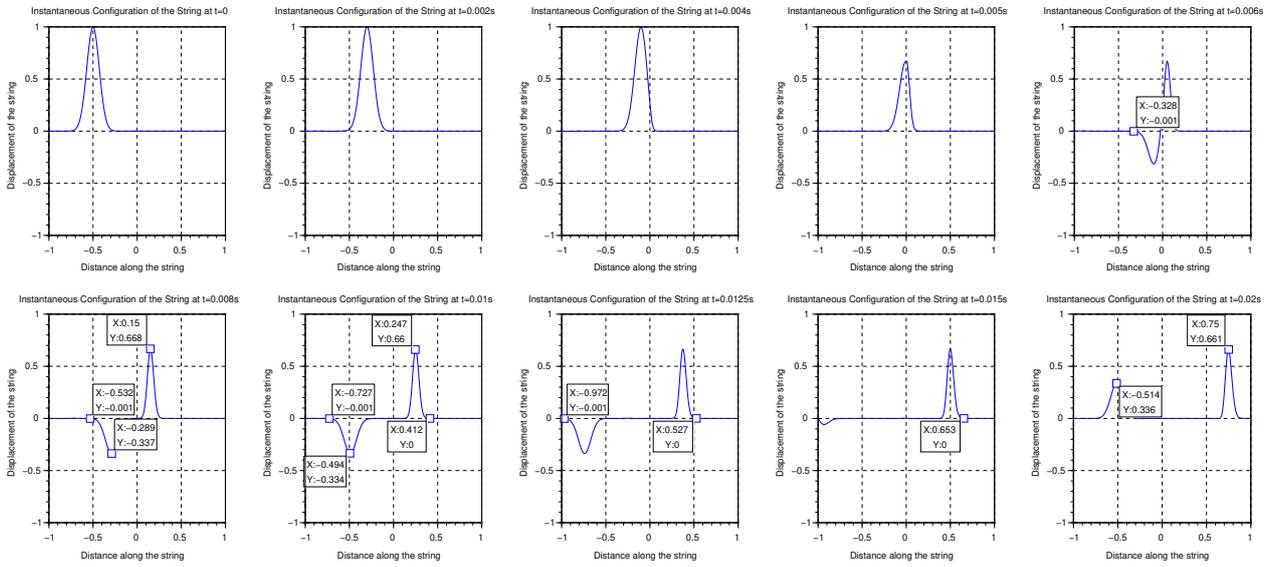
Figure 10. Reflection and Transmission of a Wave at an Impedance Discontinuity at x = 0 : String 1 extends from x = -1 to x = 0, $L_1 = 1m$, $m_11 = 0.001 Kgm^{-1}$, T=10N. String 2 extends from x = 0 to x = 1, $L_2 = 1m$, $m_2 = 0.004 Kgm^{-1}$, T = 10N. Reflected pulse has opposite polarity w.r.t. the incident pulse.
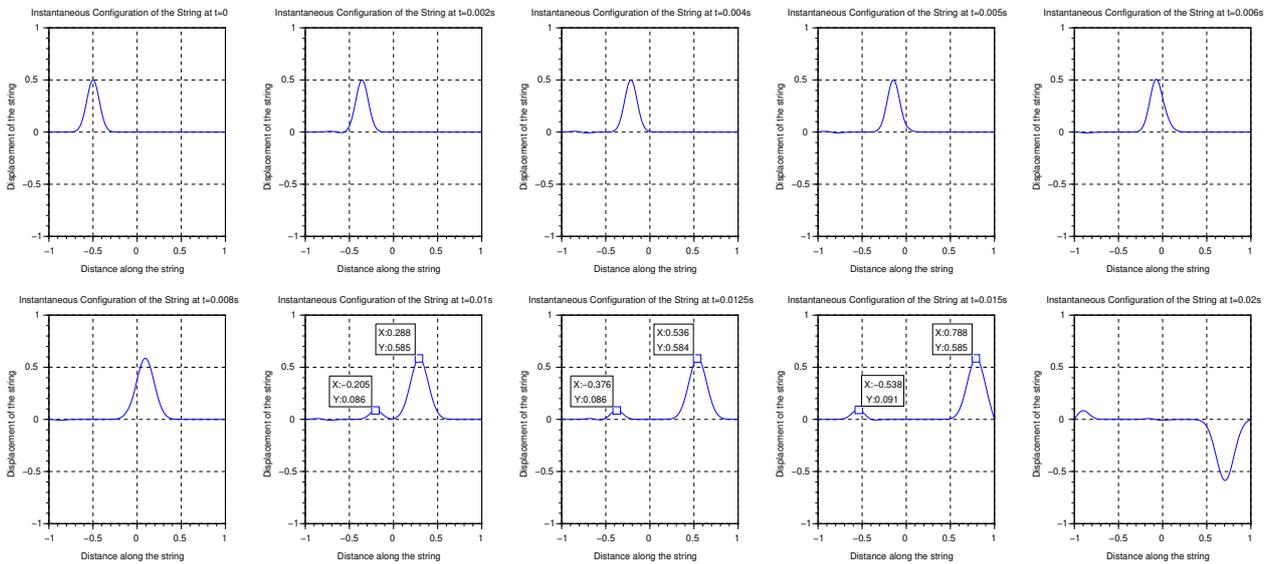
Figure 11. Reflection and Transmission of a Wave at an Impedance Discontinuity at x = 0 : String 1 extends from x = -1 to x = 0, $L_1 = 1m$, $m_1 = 0.002 Kgm^{-1}$, T=10N. String 2 extends from x = 0 to x = 1, $L_2 = 1m$, $m_2 = 0.001 Kgm^{-1}$, T=10N. Reflected pulse has the same polarity as the incident pulse.

in fair agreement with the theoretical value of $v_2 = \sqrt{\frac{T}{m_2}} = \sqrt{\frac{10}{0.004}} = 50ms^{-1}$. Figure 11 depicts the instantaneous configurations of the string corresponding to the propagation of reflected and transmitted pulses in Animation 9 . In this case since $m_1 > m_2$, we have $v_1 < v_2$ and $Z_1 > Z_2$. Thus, the reflected pulse travels slower than the transmitted pulse and there is no change in the polarity of the reflected pulse with respect to the incident pulse. This can be easily verified from the plots corresponding to $t = 0.01s$, $0.0125s$ and $0.015s$ in the figure.

7 STATIONARY WAVES

To illustrate the formation of "Stationary" or "Standing" waves we simulate a situation in which one end of the string, say, x=0 is initially subjected to transverse simple harmonic motion given by

$$u(0,t) = asin(2\pi v t), \qquad (49)$$

where a is the amplitude and v is the frequency of the motion.The other end of the string at x=L is kept rigidly fixed. A transverse harmonic wave would travel along the string in the positive x direction and get reflected at the end x=L, giving rise to a reflected wave which would travel in the negative x direction. The interference between the two oppositely travelling waves results in the formation of the stationary wave on the string.

Equation (49) is one initial condition and the other initial condition is obtained by differentiating it w.r.t. time as

$$\frac{\partial u(0,t)}{\partial t}t = 0 = 2\pi v a. \qquad (50)$$

The discrete formulation of Eqn.(50) using forward difference approximation is given by

$$u(1, j+1) = u(1,j) + l(2\pi v a), \qquad (51)$$

for all values of the index j corresponding to all instants of time. The boundary condition corresponding to the rigid end at x = L can be written as u(L,t) = 0 or as

$$u(i,j) = 0, \qquad (52)$$

for all values of j and the value of index i corresponding to the end x=L of the string.

In Animation 10 it can be seen that the stationary wave does not come into existence immediately. It is only when the wave reflected at x=L has arrived back at the driven end x=0 that we have the formation of standing wave. Thereafter, the string continues to vibrate in a particular mode provided the frequency of the simple harmonic motion of the end x=0 is given by the relation

$$v_n = \frac{n}{2L}\sqrt{\frac{T}{m}}, \qquad (53)$$

where n = 1, 2, 3, ... is the mode number. Equation (53) can be easily deduced in an analytical treatment of stationary waves. Once the required mode has been established this way, the end at x=0 is held stationary. Figure 12 shows the instantaneous configurations of the string corresponding to Animation 10. In this simulation the end x=0 is made to oscillate with a frequency
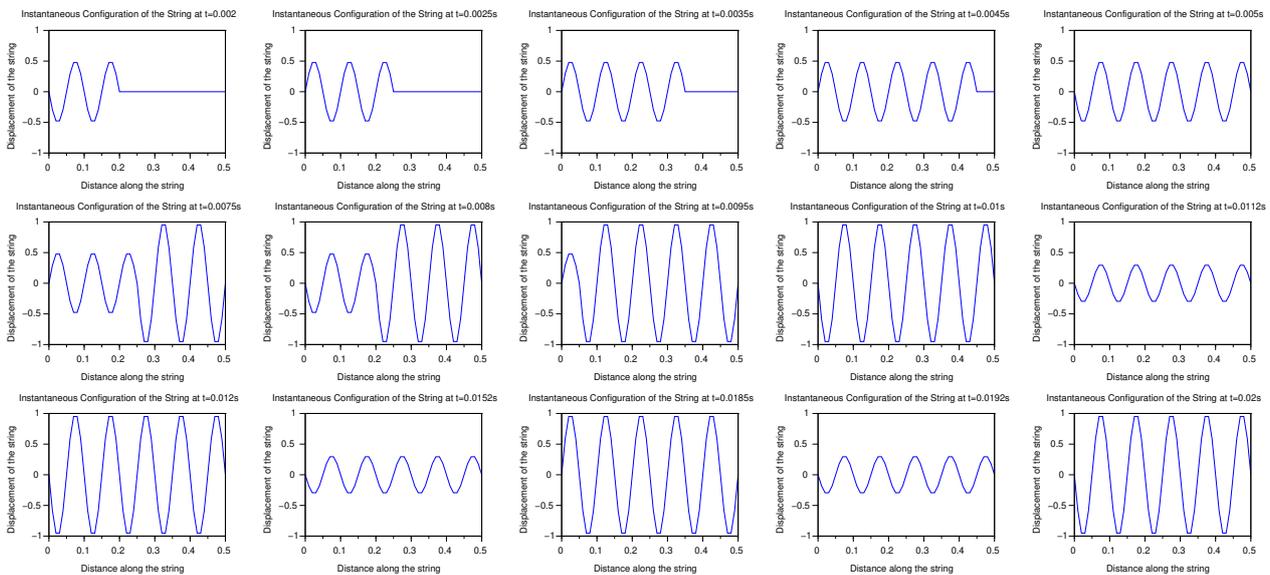
Figure 12. Formation of Stationary Waves in the Tenth Harmonic : L = 0.5m, T = 10N, $m = 0.001 Kgm^{-1}$, v = $1000s^{-1}$, a = 0.5m. String Configuration at t = 0.002s, 0.0025 0.0035s, 0.0045s, 0.005s, 0.0075s 0.008s, 0.0095s, 0.01s, 0.0112s, 0.012s, 0.0152s, 0.0185s, 0.0192s & 0.02s.
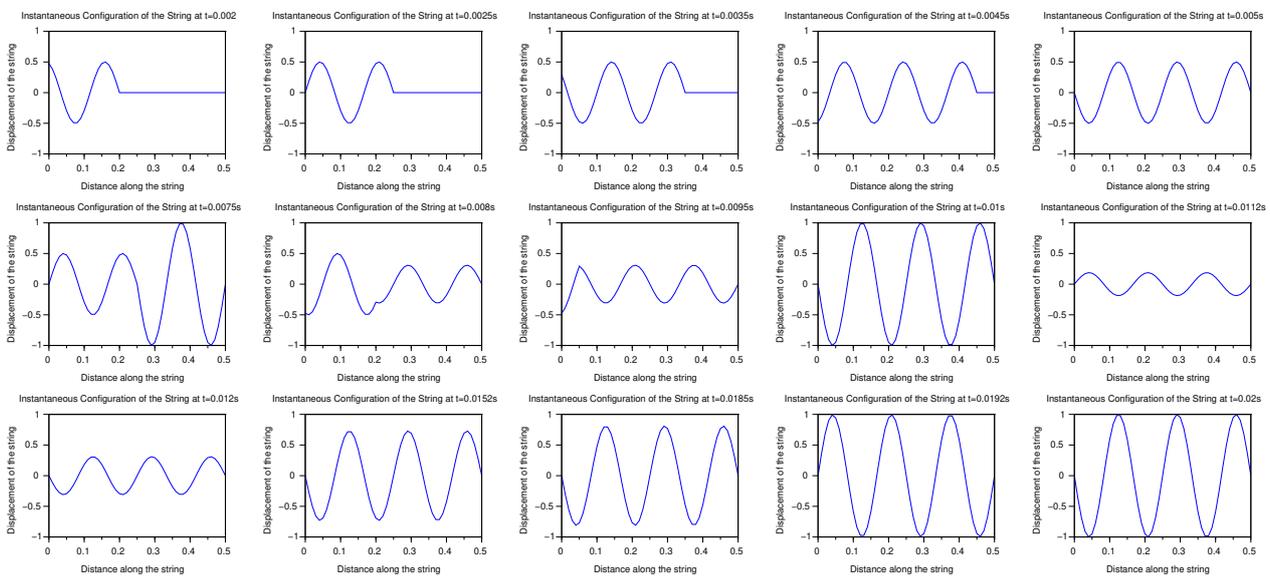
Figure 13. Formation of Stationary Waves in the Sixth Harmonic : L = 0.5m, T = 10N, $m = 0.001 Kgm^{-1}$, v = $600s^{-1}$, a = 0.5m. String Configuration at t = 0.002s, 0.0025 0.0035s, 0.0045s, 0.005s, 0.0075s 0.008s, 0.0095s, 0.01s, 0.0112s, 0.012s, 0.0152s, 0.0185s, 0.0192s & 0.02s.

equal to $1000s^{-1}$, i.e., $v_n = 1000s^{-1}$. With T = 10N, $m = 0.001 Kgm^{-1}$, L=0.5 m, Eq. (53) gives n=10, i.e., the string vibrates in the tenth mode or the tenth harmonic. Similarly, Animation 11 and Figure 13 depict the formation of stationary wave pattern when $v_n = 600s^{-1}$ and the string vibrates in the sixth harmonic (n=6).

Further, it can be seen that in the nth harmonic there are (n-1) positions (between the fixed ends) equally spaced along the string where the displacement is always zero. These points are called "nodal points" or "nodes". The points of maximum displacement are also equally spaced and are called "antinodes". At the antinodes, the displacement can be seen to be equal to the sum of the amplitudes of the constituent superimposing travelling waves.

The values of the physical parameters(L,T,m,v) can be easily changed in the given SCILAB codes to visualize other harmonics.

8   CONCLUSION

The present work is a comprehensive illustration of the boundary behaviour of transverse waves in strings. By using computational techniques to solve a given physical problem the scope of the problem can be increased manifold as compared to its analytical treatment. In some problems pertaining to practical situations an analytical treatment can either be quite challenging or not possible at all. In such cases a computational

approach can be adopted to solve the problem. For example, an analytical treatment of "real boundary" as done in this work would be cumbersome. On the other hand, the scope of the problem can be easily extended to make the support more realistic by modelling it as a damped harmonic oscillator.The discussion on reflection and transmission of waves at an "impedance discontinuity" can also be extended to illustrate the phenomenon of "impedance matching", a concept with immense practical importance in electromagnetic wave propagation and communication. Another extension that can be implemented with ease in the present treatment of stationary waves is to investigate and illustrate the concept of "Standing Wave Ratio". This ratio is of significance when the boundaries encountered by progressive waves are not perfectly rigid, resulting in incomplete or partial reflection of these waves. In such cases we have the formation of stationary wave pattern superposed on travelling waves so that the amplitude at nodes is not zero.

9   REFERENCES

1. G.Andaloro, V. Donzelli and R.M. Sperandeo-Mineo (1991), Modelling in Physics teaching: the role of computer simulation, International Journal of Science Education, 13:3, 243-254.

2. Sawormir Binek, Damian Kimla, Jerzy Jarosz, Using computer simulation to aid the interactive learning of physics

in secondary education, Phys. Edu. 53(2018)055006(7pp).

3. Jack M. Wilson and Edward F. Redish, Using Computers in Teaching Physics, Physics Today, January 1989, 34-41.

4. A.P. French, *Vibrations and Waves*, CBS Publishers and Distributors Pvt. Ltd.

5. Frank S. Crawford Jr. , *Waves, Berkley Physics Course, Volume 3*, Tata McGraw - Hill Publishing Company Limited.

6. H.J. Pain, *The Physics of Vibrations and Waves*, English Language Book Society / John Wiley.

7. N.K. Bajaj, *The Physics of Waves and Oscillations*, Tata McGraw - Hill Publishing Company Limited.

8. Paul L. DeVries and Javier E. Hasbun, *A First Course in Computational Physics*, Jones and Barlett Learning.

9. Joel Fraklin, *Computational Methods for Physics*, Cambridge University Press.

10. Darren Walker, *Computational Physics*, MedTec Introductory Physics Series.

11. SCILAB Version 6.0.0, `<http://www.scilab.org/>`.

Shape of interference fringes in Young's double slit experiment

S. Nath[1] and P. Mandal[2]

[1]Department of Physics, Scottish Church College, Kolkata 700006, India.
santunath7@gmail.com

[2]Department of Physics, St. Paul's C. M. College, Kolkata 700009, India.
pintuphys@gmail.com

Abstract

A detailed study on the shape of interference fringes in Young's double slit experiment (YDSE) is reported in this article. The nonlocalized fringes are confocal hyperboloids with the coherent sources as the foci and the observed shape depends on the position and orientation of the screen. It has been shown that the fringes are hyperbolic in shape on a screen placed at a position parallel to the plane containing the coherent sources while they are concentric circles on a screen placed perpendicular to the line joining the sources. We consider some practical values of path difference between the interfering waves of visible light and explain that fringes actually appear as straight lines. We use Android version of GeoGebra, a freeware, to simulate the shape of interference fringes.

1 Introduction

Interference, a phenomenon of superposition of waves, is a topic of paramount interest in the field of physics and is taught from the school level to the postgraduate level. Young's double slit experiment (YDSE) is a demonstrative experiment on interference of light waves first performed by Thomas Young in 1801. It is the experiment that established the wave theory of light. Similar experiment, later in 1961, was performed by Claus Jönsson but with electron beam [1] demonstrating the prediction of quantum theory.

Figure 1 depicts the schematic of YDSE, which is basically an interference produced by the method of division of wavefront. As the wave theory suggests, the interfering beams produce a constructive pattern for a path difference equal to the integral multiple of the wavelength ($\Delta = m\lambda$), while
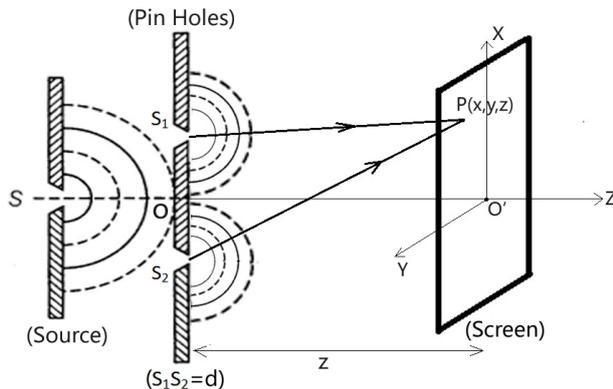
Figure 1: Schematic of Young's double slit experiment. Coherent sources S_1 and S_2 are produced from a monochromatic source S.

a destructive pattern for a path difference equal to the odd integral multiple of the half-wavelength ($\Delta = (2m+1)\lambda/2$)[2, 3, 4]. Fringe of a given order is the locus of the points having same path difference. The fringes in YDSE are nonlocalized meaning that the fringes have infinite extent and can be viewed, in principle, from any position. The shape of the fringes i.e. the equation of loci on the xy plane (Figure 1) is hyperbola which is shown in textbooks [3, 4]. Here, in this article, we study the shape of the fringes in detail and discuss the shape as appears on a screen placed at different positions. We consider two pinholes as coherent sources and show the shape of fringes in three dimension are hyperboloid with the sources as the foci. However, the fringes may appear as circular, hyperbolic or even as straight line depending on the position and orientation of the screen with respect to the sources. If

the screen is placed on xy plane in front of two point sources, the fringes are hyperbolic in nature. However, the fringes on this plane actually appear as straight line as the eccentricity of the hyperbolae is high. If the screen is placed perpendicular to the line joining the two point sources (yz plane), the fringes appear as concentric circles. The study takes into account of some practical values for the visible light. We use GeoGebra 2D and 3D graphing calculator [5], a free online software, for studying the shape of interference fringes. A recent article [6] reported a similar study with a different software Scilab. With the android version of GeoGebra, the present article reports a more detail study including the position of the screen perpendicular to the line joining the sources, and considering practical values for visible light in a more lucid way.

2 Hyperbola and Hyperboloid

In this section we will briefly discuss the geometry of hyperbola and hyperboloid, as relevant to the subject matter of the article. A hyperbola represents the locus of the points P on a plane which have a constant difference of the distances PF_1 and PF_2 with respect to two fixed points F_1 and F_2, called foci i.e. $PF_1 \sim PF_2 = 2a$, $a = $ constant (figure 2(a)). The center of the hyperbola is the midpoint M of the line segment joining the foci ($MF_1 = MF_2 = c$). This line segment, called the major axis, contains the vertices V_1 and V_2 which have equal distance a to

(a)   (b)
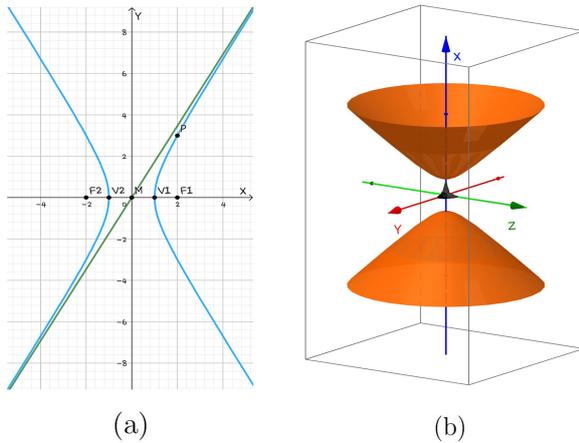
Figure 2: (a) Hyperbola with straight line as the asymptote (b) Hyperboloid.

Figure 3: Projection of a hyperboloid on xy plane-a hyperbola.

the center M. The ratio $e = c/a$ is called the eccentricity of the hyperbola. The equation of a hyperbola is given by

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \qquad (1)$$

where $b^2 = c^2 - a^2$ and the eccentricity, in terms of a and b is

$$e = \frac{c}{a} = \frac{\sqrt{a^2 + b^2}}{a} = \sqrt{1 + \left(\frac{b}{a}\right)^2} \qquad (2)$$

From eq. 1,

$$y = \pm \frac{b}{a}\sqrt{x^2 - a^2} \approx \pm \frac{b}{a}x \qquad (3)$$

for larger values of $|x|$. Thus it follows from eq. 3 that hyperbola approaches two straight lines, called asymptotes for larger values of $|x|$. One asymptote is shown in figure 2(a). It should also be noted that the hyperbola approaches to the asymptotic nature even at smaller values of $|x|$ when the eccentricity is very large i.e. $e >> 1$ or, $b >> a$.

A hyperboloid is the three dimensional geometry of a hyperbola, having symmetry of revolution in a plane, say yz. It is just like a nest but in opposite direction (figure 2(b)). The equation of a hyperboloid is given by

$$\frac{x^2}{a^2} - \frac{y^2 + z^2}{b^2} = 1 \qquad (4)$$

It is clear from figure 3, a hyperboloid if projected on xy plane, appears as hyperbola. On the other hand, it appears as circle when projected on yz plane (figure 4).

3 Shape of Fringes

We consider two pinholes S_1 and S_2 illuminated by a monochromatic point source of light S (figure 1). The pinholes act as two coherent sources which emit spherical wavelets. Superposition of these wavelets produces interference.
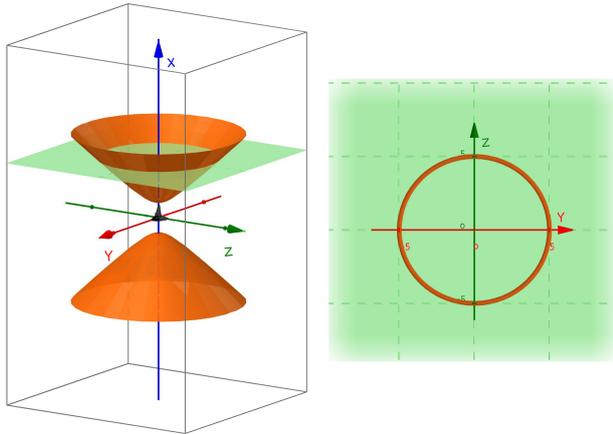
Figure 4: Projection of a hyperboloid on yz plane-a circle.

3.1 Mathematical Derivation

We calculate the path difference between the interfering waves at any point P (x, y, z) on the screen. As seen from the figure 1,

$$S_2P = \left[z^2 + y^2 + \left(x + \frac{d}{2}\right)^2\right]^{1/2}$$

$$S_1P = \left[z^2 + y^2 + \left(x - \frac{d}{2}\right)^2\right]^{1/2}$$

Hence, the path difference at the point P is $\Delta = S_2P - S_1P$ and is given by

$$\Delta = \left[z^2 + y^2 + \left(x + \frac{d}{2}\right)^2\right]^{1/2} - \left[z^2 + y^2 + \left(x - \frac{d}{2}\right)^2\right]^{1/2}$$

Or,

$$\Delta + \left[z^2 + y^2 + \left(x - \frac{d}{2}\right)^2\right]^{1/2} = \left[z^2 + y^2 + \left(x + \frac{d}{2}\right)^2\right]^{1/2} \quad (5)$$

Squaring both sides of eq. 5 and simplifying, we have

$$\frac{x^2}{\left(\frac{\Delta}{2}\right)^2} - \frac{y^2 + z^2}{\left(\frac{d}{2}\right)^2 - \left(\frac{\Delta}{2}\right)^2} = 1 \quad (6)$$

Eq. 6 represents the locus of an interference fringe corresponding to a given path difference Δ. Comparing eq. 6 with eq. 4, $a = \Delta/2$, $b = \sqrt{(d/2)^2 - (\Delta/2)^2}$ and hence $c = \sqrt{a^2 + b^2} = d/2$ i.e. the distance from the center to the pinholes S_1, S_2 (figure 1) and the distance from the center to the foci F_1, F_2 (figure 2) are same and they lie on same axis (x axis). Thus the shape of interference fringes is described by eq. 6 in three dimension which resemblances hyperboloid (eq. 4) with foci at S_1 and S_2.

3.2 Simulation of Fringe Shape

We use GeoGebra, a noncommercial freeware, developed by Markus Hohenwarter et. al. [5] to simulate the shape of interference fringes for different positions of the screen with respect to the point sources. It is an interactive mathematical software that provides geometry, algebra, statistics and calculus applications mainly intended for teaching and learning at all levels. GeoGebra is available on desktop applications for Windows, macOS, Linux, and tablet applications for Android, iPad and Windows. Its web application is based on HTML5 technology. The fringe shape simulation, reported in this article, has been performed with the android application of the software.

While simulating the fringe shape we consider some practical values of the wavelength (λ), the separation between the point sources (d) and the position of the screen with respect to the sources (D or D'). Here we take $\lambda = 6000$ Å and $d = 1$ mm for all calculations.

(1) Screen placed on xy plane: Let the screen is placed on xy plane at $z = D = 1$ m. The path difference for constructive fringes $\Delta = m\lambda$ ($m = 1, 2, 3, ...$) is on the order of λ and hence $\Delta << d$. With respect to the origin taken at O' ($x' = x = 0, y' = y = 0, z' = z - D = 0$) (figure 1), eq. 6 reduces to

$$\frac{x^2}{\left(\frac{\Delta}{2}\right)^2} - \frac{y^2}{\left(\frac{d}{2}\right)^2} = 1 \qquad (7)$$

which represents hyperbola on xy plane with the point sources S$_1$ and S$_2$ as the foci that lie on the x axis at $z = 0$. A set of confocal hyperbolae will be obtained for a set of values of the path difference Δ. In other words, fringes of different orders will form a set of confocal hyperbolae as shown in the figure 5.

The eccentricities of the hyperbolae are given by (eq. 2)

$$e = \sqrt{1 + \left(\frac{d}{\Delta}\right)^2} \approx \frac{d}{\Delta}$$

For $d = 1$ mm and $\Delta = 10\lambda = 60$ nm, the eccentricity $e \approx 167$ which is very high. The observation, in practice, is made over a very small region and hence the fringes appear as straight fringes (figure 6). The fringe width, defined as the separation between two consecutive bright or dark fringes is

$$\beta = \frac{\lambda D}{d} = 0.6\,\text{mm}$$

From the plot we see, same value of the fringe width β when the observation is made near the center.
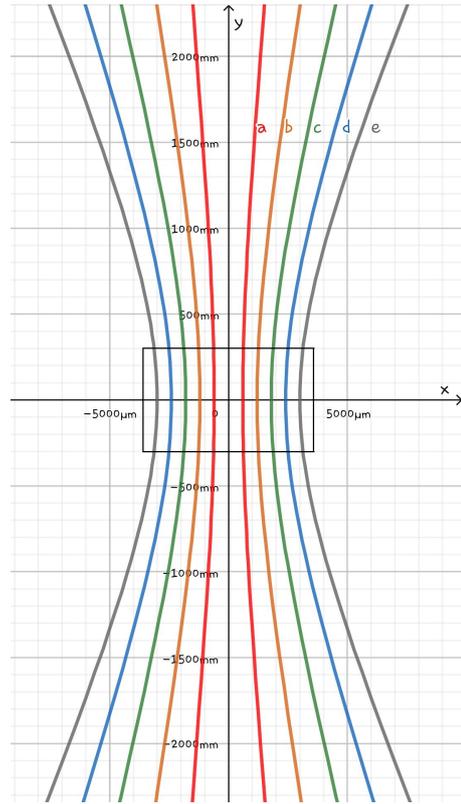
Figure 5: A set of confocal hyperbolae representing fringes of different orders on xy plane at $z = D = 1$ m. Hyperbolae marked as a, b, c, d, e correspond to different values of Δ.

(2) Screen placed on yz plane: Now we consider a different position of the screen. Let the screen is placed on yz plane at $x = D'$, perpendicular to the line joining the point sources S$_1$ and S$_2$. For large values of D'/Δ, eq. 6 reduces to
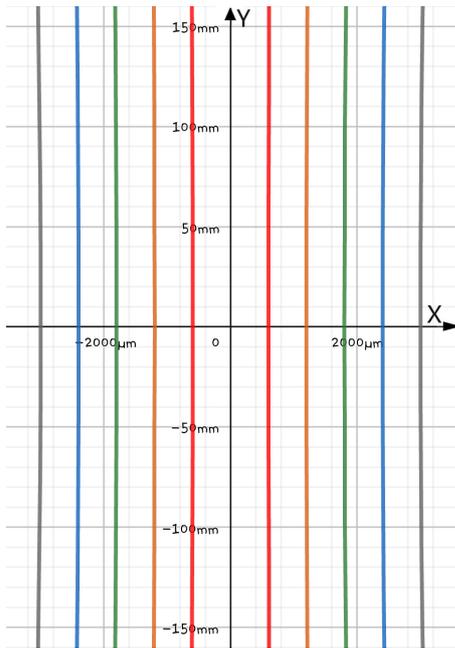
Figure 6: Hyperbolic fringes of high eccentricity as appear on xy plane when observation is made over small region.

$$\frac{y^2 + z^2}{\left(\frac{d}{2}\right)^2 - \left(\frac{\Delta}{2}\right)^2} = \frac{D'^2}{\left(\frac{\Delta}{2}\right)^2}$$

$$\Rightarrow y^2 + z^2 = D'^2 \left[\left(\frac{d}{\Delta}\right)^2 - 1\right] \quad (8)$$

Eq. 8 represents a circle of radius $r = D'\sqrt{(d/\Delta)^2 - 1}$ for a given Δ. For this position of the screen, the path difference between the interfering waves is smaller than the separation between the sources i.e. $\Delta \leq d$. It should be noted that $\Delta = d$ at the center while $\Delta \to 0$ as $r \to \infty$. Thus we may replace Δ by $d - m\lambda$ in eq. 8 where m is the order number of fringes ($m = 0$ for the central fringe). The fringes of different orders will appear as concentric circles (figure 7) on
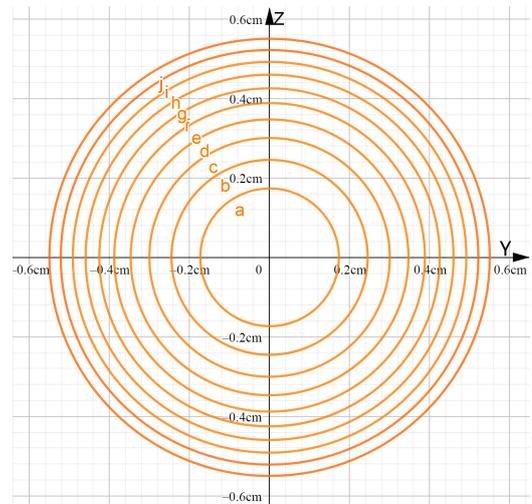
Figure 7: Fringes of concentric circles as expected on yz plane ($x = D' = 5$ cm). Circles marked as $a, b, c, d, ...$ correspond to different values of Δ.

the screen. As seen from figure 7, the circular fringes have large radii and they are widely separated around the center. In practice, observation is made over small region and hence the fringe pattern will remain unobserved on the screen at yz plane around the center.

In order to visualize the fringes, observation may be made somewhere else. Fig-
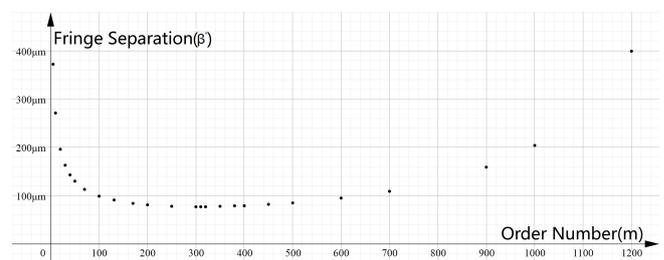
Figure 8: Variation in the separation between consecutive circular fringes (β') with the order number (m).

Figure 9: Circular fringes as appear within small field of view.

ure 7 shows that the separation between consecutive fringes (β') decreases for larger values of r i.e. at higher orders. However, the decrease in β' with the increase in m is not monotonic in nature. The separation β' decreases monotonically upto some value of m, above which it increases monotonically; a feature expected from the geometry of confocal hyperboloids. We calculated the radii of the circular fringes for a wide variation in the order number m with $D' = 5$ cm and plotted in figure 8. As can be seen from figure 8, the separation becomes minimum ($\beta' \approx 80\ \mu$m) somewhere around $m = 300$. If observation is made around these fringes over a small region, they will appear as straight fringes (figure 9).

4   Conclusion

Using the android version of a freeware GeoGebra, we report a detailed study on the shape of fringes produced in Young's dou-

ble slit experiment. The shape of the fringes is hyperboloid in three dimension and when projected on a screen, appears as hyperbola or circle depending on its position and orientation relative to the sources. It is shown that the eccentricity of the hyperbola and the radius of the circle are very large in practice while the observation is made over small region, and hence the fringes actually appear as straight lines. Students at the undergraduate level have observed the shape in Fresnel's biprism experiment and can easily correlate with the observations reported here. The study can be further extended to simulate the fringe-shape for other orientation of the screen at a fixed position.

References

[1] C. Jnsson (1961) Zeitschrift fr Physik 161 (4), 454

[2] F. A. Jenkins, H. E. White Fundamentals of Optics, (The McGraw-Hill Company, 2001)

[3] A. Ghatak Optics, (Tata McGraw-Hill Publishing Company Limited, 2007)

[4] B. Ghosh, K. G. Mazumdar A Text Book on Light, (Sreedhar Publishers, 2014)

[5] https://www.geogebra.org/

[6] P. Ashdhir, N. Batra, N. Goyal, M. Tyagi (2019) Physics Education 35, 2

Equipartition theorem for non-linear oscillators

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

[1] Department of Physics, Govt. Engineering College,
Thrissur - 680 009, India.
`prasanthpnairg@gmail.com`

[2]School of Pure and Applied Physics, Kannur University,
Kannur - 670 327, India.
`mailnishanthp@yahoo.com`

[3]Department of Physics, Nehru Arts and Science College,
Kanhangad - 671 314, India.
`udayanandan@gmail.com`

Abstract

In many text books used for undergraduate courses, equipartition theorem (EPT) is defined only for quadratic Hamiltonian. In this article we apply the EPT for many non linear (non quadratic) Hamiltonian, which will help the students to have a more understanding on EPT. We also give some applications of Tolman's definition of EPT

1   Introduction

Equipartition theorem is a very useful result of thermodynamics which says "If a system described by classical statistical mechanics is in equilibrium at the absolute temperature T, every independent quadratic term in its energy has a mean value equal to $\frac{1}{2}kT$ " [1]. In many text books [1, 2, 3, 4, 5] it is defined only for quadratic Hamiltonian and energies of non quadratic forms are not discussed. A discussion with the students and teachers by the authors, found that many of the students are not aware about the energy partition by non quadratic Hamiltonian. Many of them believe that all types of energies contribute $\frac{1}{2}kT$ internal energy, where k is the Boltzmann constant.

2 Finding internal energy

Different physicists had studied and contributed to the equipartition theorem. Waterstone [6] studied about the motion of molecules in gases and presented the first statement of equipartition theorem. Lord Rayleigh [7] used equipartition theorem in the black body problem to find the distribution of energy. Maxwell [8] worked out the general principles of statistical equilibrium by studying molecular collisions. Method of kinetic energy is valid for short duration of collision of molecules and Maxwell overcame this problem by postulating a mechanical system in generalized Lagrange Hamiltonian coordinates. For computing equilibrium properties of the system, Maxwell used equipartition theorem. Paradox of specific heat and ultraviolet catastrophe were the observations that mismatches with the assumptions of equipartition theorem which were later solved with the emergence of quantization of energy [9, 10]. We have two methods to find average energy, they are the probability method given by Maxwell [8] and the canonical ensemble method given in text books [1, 2].

2.1 Probability method

For a classical system in equilibrium, we write the energy as

$$E = E(q_1, q_2...q_f, p_1, p_2, ...p_f)$$

where q_s and p_s are the generalized coordinates and conjugate momenta. The average internal energy can be found using the equation

$$< E >= \frac{\int E e^{-\beta E} dq_1...dp_f}{\int e^{-\beta E} dq_1...dp_f}$$

By integration and rearrangement we get [1]

$$< E >= -\frac{\partial}{\partial \beta} \ln \left(\int e^{-\beta E} d^3q \right)$$

For 1D ideal gas with $E = \frac{p^2}{2m}$, we get

$$< E >=< \frac{p^2}{2m} >= \frac{3k\,T}{2}$$

For energy $\frac{kq^2}{2}$, we get

$$< E >=< \frac{kq^2}{2} >= \frac{3k\,T}{2}$$

Similar calculations show that for a 3D harmonic oscillator

$$< E >=< \frac{p^2}{2m} + \frac{kq^2}{2} >= 3kT$$

and for a particle in a gravitational field

$$< \frac{p^2}{2m} + mgz >= \frac{5kT}{2}$$

where z is the height of the particle.

2.2 Canonical ensemble method

Now we can find the internal energy from the partition function of a system. Let the energy of the system be

$$E(q) = aq^n$$

where q is either generalized coordinate or conjugate momentum. The energy dependence on n may be of three types $n > 0, n =$

0 and $n < 0$ where n represents a number. Let us evaluate the average energy for $n = 1, 2, 3, 4$. The partition function for a system in 3 dimension (3D) is

$$Q = C \int q^2 dq e^{-\beta E(q)}$$

where $C = \frac{4\pi V}{h^3}$ for energy in terms of momentum only and $C = 4\pi$ for energy in terms of position only. Internal energy is given by [5]

$$U = kT^2 \left(\frac{\partial \ln Q}{\partial T} \right)_{N,V}$$

1. For $n > 0$

$$Q = C \int q^2 dq e^{-\beta E(q)}$$

When $E(q) = q$ we get

$$Q = C \int q^2 dq e^{-\beta q}$$

Using the standard integral

$$\int_0^\infty x^n dx e^{-\mu x} = n!(\mu)^{-n-1}$$

we get

$$U = 3NkT$$

Similarly for $E(q) = q^2, q^3$ and q^4 we get

$$U = \frac{3NkT}{2}$$

$$U = \frac{3NkT}{3}$$

$$U = \frac{3NkT}{4}$$

The above expressions show that in general, the mean internal energy depends on power of the energy function as

$$U = \frac{DNkT}{n} \tag{1}$$

where n is the power of energy function and D is the dimension.

2. For $n < 0$ or negative

$$U = \frac{DkT}{n}$$

We get the same expression for the internal energy, but when we substitute negative values for n we get average energy as negative, which means the particle is confined .

3. For $n = 0$ we have

$$E = a$$

Here energy is a constant. So, the average energy or the internal energy will be same as the given energy. So

$$< E >= a$$

which is an expected result.

3 Tolman's EPT

Tolman [11] proposed a generalized equipartition theorem (called I generalized EPT) given by

$$\left\langle q \frac{\partial E}{\partial q} \right\rangle = k T \tag{2}$$

This is derived from the partition function as below. Partition function

$$Q = \int dq_1 dq_2...dq_n C e^{(-\beta E(q_1,q_2....q_n))}$$

where q_i are generalized coordinates and generalized momenta. Taking a derivative of $q_1 e^{-\beta E}$ and substituting in the partition function and integrating with proper limits we get

$$kT = \frac{\int dq_1 dq_2...dq_n q_1 \frac{\partial E}{\partial q_1} e^{-\beta E}}{Q}$$

Right hand side is the definition for the average of $\left\langle q_1 \frac{\partial E}{\partial q_1} \right\rangle$. So

$$\left\langle q_1 \frac{\partial E}{\partial q_1} \right\rangle = kT$$

This equation is Tolman's definition of EPT given by Eqn (2).

3.1   Some Examples

Using Tolman's equation we get

1. For $E = ax^5$
$$\left\langle ax^5 \right\rangle = \frac{kT}{5}$$

2. For $E = p_x c$
$$\left\langle p_x c \right\rangle = kT$$

3. For $E = \frac{p_x^2}{2m}$
$$\left\langle \frac{p_x^2}{2m} \right\rangle = \frac{kT}{2}$$

These results agree with the general formula obtained for internal energy or average energy given by Eqn (1). Here we had taken 1 D systems.

4   Application of Tolman's EPT

Phillies [12, 13] proposed an application for Tolman's EPT formula and he called the result as second generalized equipartition theorem. Accordingly,

$$\left\langle q_a q_b \frac{\partial E}{\partial q_a} \frac{\partial E}{\partial q_b} \right\rangle = (kT)^2 +$$

$$\left\langle q_a q_b \, kT \frac{\partial^2 E}{\partial q_a \partial q_b} + q_b \frac{\partial q_a}{\partial q_b} kT \frac{\partial E}{\partial q_a} \right\rangle$$

If $q_a = q_b = p_x$ we get

$$\left\langle p_x^2 \frac{p_x}{m} \cdot \frac{p_x}{m} \right\rangle = (kT)^2 + \left\langle p_x^2 kT \frac{1}{m} + p_x kT \frac{p_x}{m} \right\rangle$$

$$\left\langle \left(\frac{p^2}{2m} \right)^2 \right\rangle = \frac{3(kT)^2}{4}$$

So

$$< E^2 > = \frac{3}{4} k^2 T^2$$

Thus this application gives mean square energy of a system. This is very interesting because the mean square fluctuation of energy can be easily found if $< E^2 >$ and $< E >^2$ are known. For example, for 1 D ideal gas

$$< E >^2 = \left(\frac{1}{2} kT \right)^2$$

and here mean square fluctuation is given by

$$< \Delta E^2 > = < E^2 > - < E >^2$$

On substituting $< E^2 >$ and $< E >^2$ we get

$$< \Delta E^2 > = \frac{1}{2} k^2 T^2$$

Thus using first generalized equipartition theorem we can find $< E >^2$ and using second generalized equipartition theorem we can find $< E^2 >$

4.1   Example

For a pure anharmonic oscillator (without kinetic energy)

$$E = aq^n$$

Using the second equipartition theorem for 1D anharmonic oscillator we get

$$< E^2 >= \frac{n+1}{n^2}(kT)^2$$

Using the value for $< E >$ we get

$$< \Delta E^2 >= \frac{k^2 T^2}{n}$$

5 $< E^2 >$ for 3D systems

The above formula to find $< E^2 >$ is suitable only for a 1 D system. To find $< E^2 >$ for 3D system we have to use the probability method. We know

$$< E^2 >= \frac{C \int E^2 p^2 dp e^{-\beta E}}{\int C p^2 dp e^{-\beta E}}$$

where C is a constant. Using the above formula we can show that for

1. $E = p^2$

$$< E^2 >= \frac{5 \times 3}{2 \times 2}(kT)^2$$

2. $E = p^4$

$$< E^2 >= \frac{7 \times 3}{4 \times 4}(kT)^2$$

3. $E = p^6$

$$< E^2 >= \frac{9 \times 3}{2 \times 2}(kT)^2$$

Generalising for $E = p^n$

$$< E^2 >= \frac{3(n+3)}{n^2}(kT)^2$$

5.1 $< \Delta E^2 >$ for 3D systems

From $< E >^2$ and $< E^2 >$ for systems with energy

1. $E = pc$ we get

$$< \Delta E^2 >= 3(kT)^2$$

   and for

2. $E = \frac{P^2}{2m}$

$$< \Delta E^2 >= \frac{3}{2}(kT)^2$$

6   Non linear oscillators

6.1   Quartic oscillator

Quartic oscillator is an anharmonic oscillator studied in detail over decades [15, 16, 17]. This oscillator is non linear in nature and it shows chaotic behaviour in most regions where it is defined. This oscillator is studied both in classical and quantum realms [15, 16]. Biswas et al [14] studied λx^{2m} type general anharmonic oscillators and found the energy spectrum, where $m = 2, 4..$ and λ is a constant. For a quartic oscillator with N degrees of freedom

$$H = \sum_i \frac{p_i^2}{2} + \sum_{i,j} \frac{\alpha_{ij}}{2} q_i^2 q_j^2$$

and for one degree of freedom

$$H = p^2 + q^4$$

We will apply EPT for above hamiltonian. As per EPT

$$< E >= \frac{1}{2}(kT) + \frac{1}{4}(kT)$$

$$< E >= \frac{3}{4}(kT)$$

$$< \Delta E^2 >= \frac{1}{2}(kT)^2 + \frac{1}{4}(kT)^2$$

$$< \Delta E^2 >= \frac{3}{4}(kT)^2$$

These results matches very accurately with V. M. Bannur et. al. [15] found using the method of statistical mechanics. Also we get

$$< E^2 >= \frac{17}{16}(kT)^2$$

for 1D and

$$< E^2 >= \frac{81}{16}(kT)^2$$

for 3D.

6.2   Duffing oscillator

Anharmonic oscillations and different physical processes are studied with Duffing oscillators [18]. Different physical systems are solved with Duffing oscillator as prototype. Hamiltonian for a Duffing oscillator is

$$H = \frac{1}{2}p_y^2 + \frac{1}{2}\alpha x^2 + \frac{1}{4}\beta x^4$$

Using EPT

$$< E >= \frac{1}{2}(kT) + \frac{1}{2}(kT) + \frac{1}{4}(kT)$$

$$< E >= \frac{5}{4}(kT)$$

$$< \Delta E^2 >= \frac{1}{2}(kT)^2 + \frac{1}{2}(kT)^2 + \frac{1}{4}(kT)^2$$

$$< \Delta E^2 >= \frac{5}{4}(kT)^2$$

Also

$$< E^2 >= \frac{29}{16}(kT)^2$$

for 1D and

$$< E^2 >= \frac{141}{16}(kT)^2$$

for 3D.

7   Conclusions

In this article EPT is applied for systems with non quadratic and non linear Hamiltonians. We used canonical ensemble method and also Tolman's definition for EPT for finding internal energies. We hope that undergraduate students will get more insights about EPT by studying about these types of Hamiltonian.

References

[1] F. Reif,   Fundamentals of statistical and thermal physics, Waveland Press, (2009).

[2] K. Huang,   Introduction to Statistical Physics, Chapman and Hall/CRC (2009).

[3] S. Salinas,   Introduction to Statistical Physics. Springer Science and Business Media, (2001).

[4] R. Bowley and M. Sanchez, Introductory Statistical Mechanics, 40, Oxford: Clarendon Press, (1999).

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

[6] J. J. Waterston, On a General Theory of Gases, Brit. Assoc. Rep., (Part 2), 6 (1851).

[7] Lord Rayleigh [Third Baron], Remarks Upon the Law of Complete Radiation Phil. Mag., Series 5, 49, 53940, (1900).

[8] J. C. Maxwell, XXXIV, On Boltzmann's theorem on the average distribution of energy in a system of material points, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 14 (88), 299-312, (1882)

[9] S. G. Brush, Foundations of Statistical Mechanics 18451915, Archive for History of Exact Sciences, 4(3), 145-183, (1967).

[10] A. Einstein, Plancks theory of radiation and the theory of specific heat, Ann. Phys, 22, 180-190, (1907).

[11] R. C. Tolman, The Principles of Statistical Mechanics, Courier Corporation (1979).

[12] G. D. J. Phillies, J. Chem. Phys. 78, 1620 (1983).

[13] G. D. J. Phillies, Elementary Lectures in Statistical Mechanics, Springer Science and Business Media (2012).

[14] S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, V. S. Varma Eigenvalues of λx^{2m} anharmonic oscillators, Journal of Mathematical Physics, 14(9), 1190-1195, (1973).

[15] V. M. Bannur and R. B. Thayyullathil Formulation of statistical mechanics for chaotic systems, Pramana, 72(2), 315-323, (2009).

[16] Balian R., G. Parisi, and A. Voros Quartic oscillator − Feynman path integrals, Springer, Berlin, Heidelberg, 337-360, (1979).

[17] E. Delabaere, and F. Pham Unfolding the quartic oscillator, Annals of Physics, 261 (2), 180-218, (1997).

[18] I. Kovacic and M. J. Brennan The Duffing equation: nonlinear oscillators and their behaviour, John Wiley and Sons, (2011).

An Illustrated Guide to Effect of Phase to Wave Phenomena

Yukio Kobayashi

Department of Information Systems Science,

Faculty of Science and Engineering, Soka University

1-236 Tangi-machi, Hachioji-shi, Tokyo 192-8577, Japan

koba@t.soka.ac.jp

Abstract

Visual clues to realize the effect of phase to wave phenomena such as the acoustic Doppler effect and the interference of two waves are presented on the basis of elementary geometry. Diagrams focused on the phase of waves are effective to understand the mechanisms of these phenomena intuitively. Students can understand those diagrams without preliminary knowledge about addition theorems of circular functions. This subject can be regarded as a cross-curricular study of mathematics education and physics education.

1   Introduction

Visual thinking by pictures or diagrams is effective to raise an image and understand wave phenomena deeply. Explanation by equations is not necessarily help students realize why a particular phenomenon occurs and how it progresses, although students can improve their proficiency level of mathematical description. For example, the essence of the acoustic Doppler effect is the invariance of the wave number held between a listener and the source of sound[1]. Some students raise a question how the invariance of the wave number, or phase, induces the acoustic Doppler effect. Textbooks on elementary physics, however, have not provided instructive pictures or diagrams that we can understand the causal relationship between the invariance of the number of waves and the Doppler effect. Interference is also difficult for students without the preliminary knowledge of the trigonometric identities to understand. Circular functions, sin and cos, are necessary to learn the constructive and deconstructive interference of traveling waves. The present article provides diagrams as visual clues to realize the mechanisms of these wave phenomena on the basis of elementary geometry without requiring preliminary knowledge about ad-

dition theorems of circular functions. The aim is not to develop a simple and intuitive derivation of these phenomena but to illustrate the effect of phase to these phenomena, although the present approach seems a little cumbersome for the acoustic Doppler effect.

2 Diagrams for explaining acoustic Doppler effect

When a sound source, a listener, or both move relative to the medium, the frequency of the sound heard by the listener is not the same as when the source and the listener are at rest in the medium. These phenomena are called as the acoustic Doppler effect, which are separated into three cases:

Case 1: The listener is moving relative to the medium, while the source is at rest there.

Case 2: The source is moving relative to the medium, while the listener is at rest there.

Case 3: The listener and the source are moving relative to the medium.

We set three coordinate reference systems, K_M, K_L, and K_S, which are fixed to the medium, the listener, and the source, respectively. The velocity of sound, the listener, and the source with respect to the medium are denoted by u, v_L, and v_S, respectively. The sound frequency, f, is constant, and the frequency with respect to the listener in each case are denoted by ^1f_L, ^2f_L, and ^3f_L, respectively. We define the phase function $\varphi(x,t)$

of a sinusoidal traveling wave propagating in the x direction to be the argument of the wave function $\sin(\omega t - kx)$ at a fixed time t and fixed x,

$$\varphi(x,t) = \omega t - kx, \qquad (1)$$

where x expresses the position x_M, x_L, and x_S measured at the reference systems, K_M, K_L, and K_S, respectively [2]. The frequency, $f = \omega/2\pi$, is expressed with the angular frequency, ω, and f can be replaced by $^1f_L =^1 \omega_L/2\pi$, where $^1\omega_L$ is the angular frequency with respect to the listener in Case 1. The frequency and angular frequency in Cases 2 and 3 are also denoted in the same way. The rate of increase of phase angle per unit length is $k = 2\pi/\lambda$, where the wave length, λ, expresses $^1\lambda$, $^2\lambda$, and $^3\lambda$ measured in Cases 1, 2, and 3, respectively. The number of waves is $\varphi(x,t)/(2\pi)$, and thus we can consider the phase function instead of the number of waves. In contrast to the number of waves, the phase function is effective to visual thinking, because the value of the phase function is the angle in the diagram of phase such as Figures 1, 3, and 5. At a given time t the phase angle decreases linearly with x in the term $-kx$, while at a given position x the phase angle increases linearly with t in the term ωt [2]. The phase function expressed by (1) indicates that going to greater x decreases the phase angle. In Figures 1, 3, and 5, a counterclockwise direction is a positive direction. For any one of the three cases, the numbers of waves measured at K_L and K_S are the same.

By drawing the invariance of the phase

angle, or the numbers of waves, in the diagrams of Figures 1, 3, and 5, we can consider the mechanism of the acoustic Doppler effect in each case.

Case 1. The Galileo transformation connects x_L with x_S as shown in Figure 2. Assuming the origins $x_L = 0$ and $x_S = 0$ of K_L and K_S coincide at time $t = 0$, we have $x_L = x_S - v_L t$. The phase angle decrease, $-kx$, is different depending on the coordinate reference systems, while the phase angle measured in each system, $\omega t - kx$, is the same. Thus, the frequency, or the angular frequency, with respect to the listener is different in each system. For any one of the three cases, the wavelength measured in K_L and K_S are the same[3, 4]. This fact implies that $^1k = 2\pi/^1\lambda$ is the same in the two systems. From Figures 1 and 2, the relation

$$^1\omega_L = \omega - {}^1kv_L \qquad (2)$$

is clear at a glance, and (2) can be rewritten as

$$2\pi\,{}^1f_L = 2\pi f - \frac{2\pi}{^1\lambda}v_L, \qquad (3)$$

where $f = u/^1\lambda$ is the source frequency. With slight rearrangements, we obtain the desired relation

$$^1f_L = \frac{u - v_L}{u}f. \qquad (4)$$

The Galileo transformation causes the phase difference of $-2\pi v_L f t/u$ between $2\pi\,{}^1f_L t$ and $2\pi f t$ with that of $-^1kv_L t$ between $-^1kx_L$ and $-^1kx_S$ at the same time. In some textbooks [1], (4) is deduced from the phase invariance, $^1\omega_L t - {}^1kx_L = \omega t - {}^1kx_S$, and

the Galileo transformation equation, $x_L = x_S - v_L t$. Figure 1 shows visually the cause of the difference of the frequency between K_L and K_S.

Case 2. In contrast to Case 1, the source frequency is $f = (u - v_S)/^2\lambda$. From Figures 3 and 4, the relation

$$^2\omega = \omega + {}^2kv_S \qquad (5)$$

is clear at a glance, where $^2k = 2\pi/^2\lambda$. In the same way as Case 1, we obtain the desired relation

$$^2f_L = \frac{u}{u - v_S}f. \qquad (6)$$

Case 3. From Figure 6, we have $x_S - x_L = (v_L - v_S)t$, and thus the difference between $-^3kx_S$ and $-^3kx_L$ is $^3k(v_L - v_S)t$, where $^3k = 2\pi/^3\lambda$. The source frequency is $f = (u - v_S)/^3\lambda$. From Figures 5 and 6, the relation

$$\omega = {}^3\omega_L + {}^3k(v_L - v_S) \qquad (7)$$

is clear at a glance. In the same way as Cases 1 and 2, we obtain the desired relation

$$^3f_L = \frac{u - v_L}{u - v_S}f. \qquad (8)$$

REMARK.

We have two approaches to deriving the formulae of the Doppler effect, (4), (6), and (8). One approach is based on the key points that the wavelength of the sound wave is invariant across the coordinate reference systems, K_M, K_L, and K_S, and that the sound velocity relative to the listener and the sound source
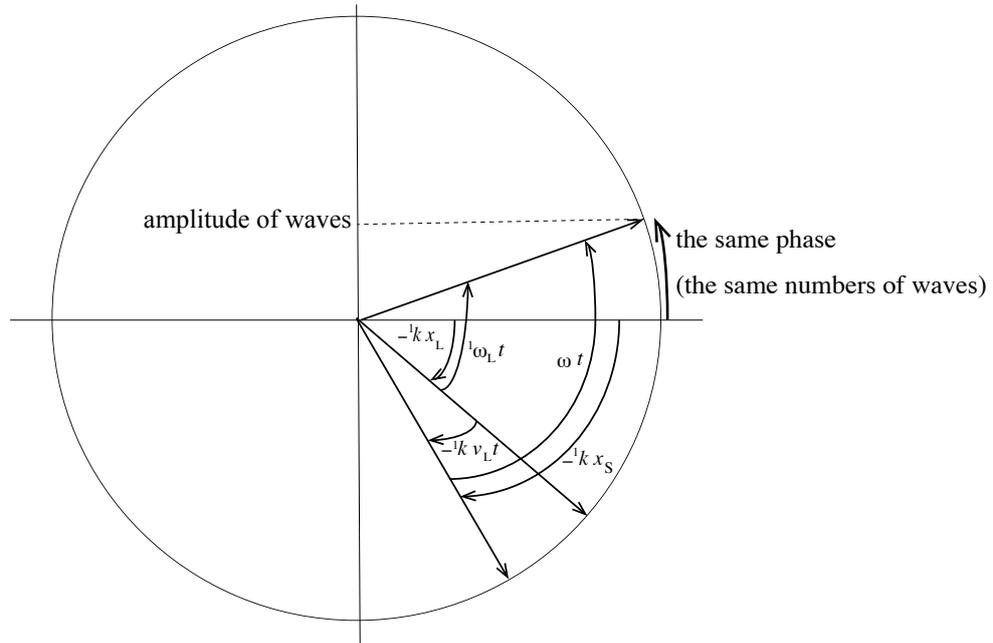
Figure 1. Phase invariance of the sound wave across K_L and K_S in Case 1. The listener moves away from a stationary source. The projection to the vertical axis represents the amplitude of sound wave. From Figure 2, $(-^1kx_S) - (-^1kx_L) = -^1kv_Lt$.
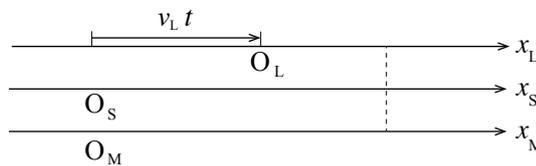
Figure 2. The relation among x_M, x_L, and x_S in Case 1. The Galileo transformation equation is $x_L - x_S = v_Lt$.

is depend on their motion [3, 4]. For example, in Case 3,

$$^3\lambda = \frac{u - v_L}{^3f_L} = \frac{u - v_S}{f}, \qquad (9)$$

which is rewritten as (8). Another approach is based on the invariance of the phase and wavelength of the sound wave across the coordinate reference systems at the same time

as shown in the present study. As suggested by these two approaches, the relative velocity of the sound wave is related to the phase invariance of the sound wave under the invariance of the wave length across the coordinate reference systems. From $\varphi(x_S, t) = \omega t - kx_S$ and $\varphi(x_L, t) = \omega_L t - kx_L$ at the same time t, we obtain
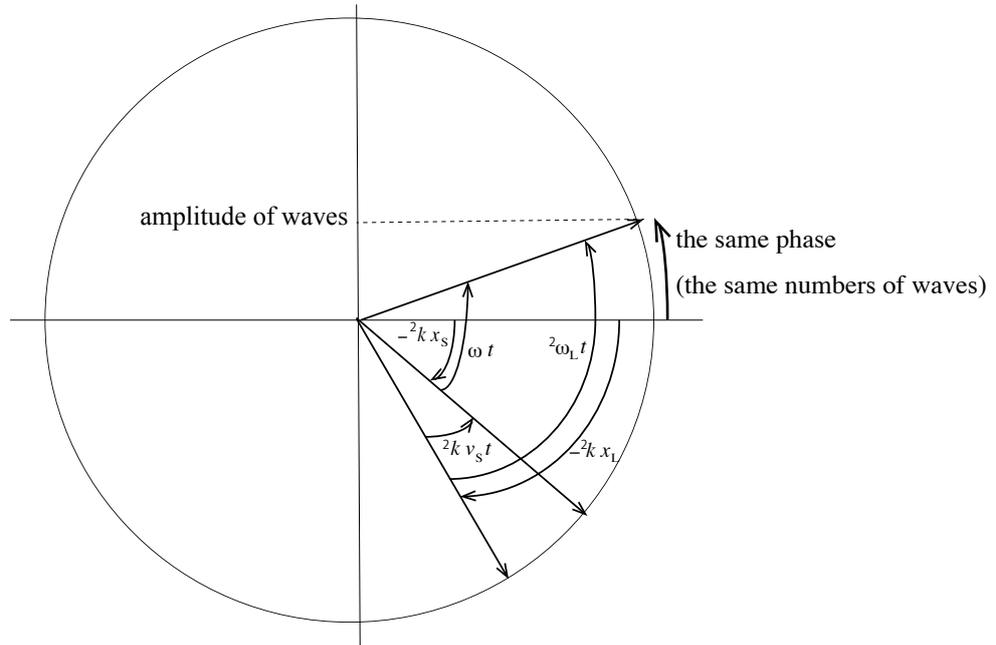
Figure 3. Phase invariance of the sound wave across K_L and K_S in Case 2. The source moves away from a stationary listener. The projection to the vertical axis represents the amplitude of sound wave. From Figure 4, $(-^2kx_S) - (-^2kx_L) =^2 kv_St$.
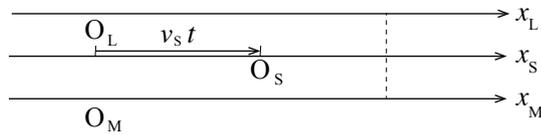
Figure 4. The relation among x_M, x_L, and x_S in Case 2. The Galileo transformation equation is $x_S - x_L = v_St$.

$$\varphi(x_S, t) - \varphi(x_L, t) = 2\pi \frac{u_S t - x_S}{3\lambda} - 2\pi \frac{u_L t - x_L}{3\lambda}$$

$$= \frac{2\pi}{3\lambda}[(u_S - u_L) - (v_L - v_S)]t, \qquad (10)$$

where u_S and u_L are the sound velocity relative to the source and the listener,
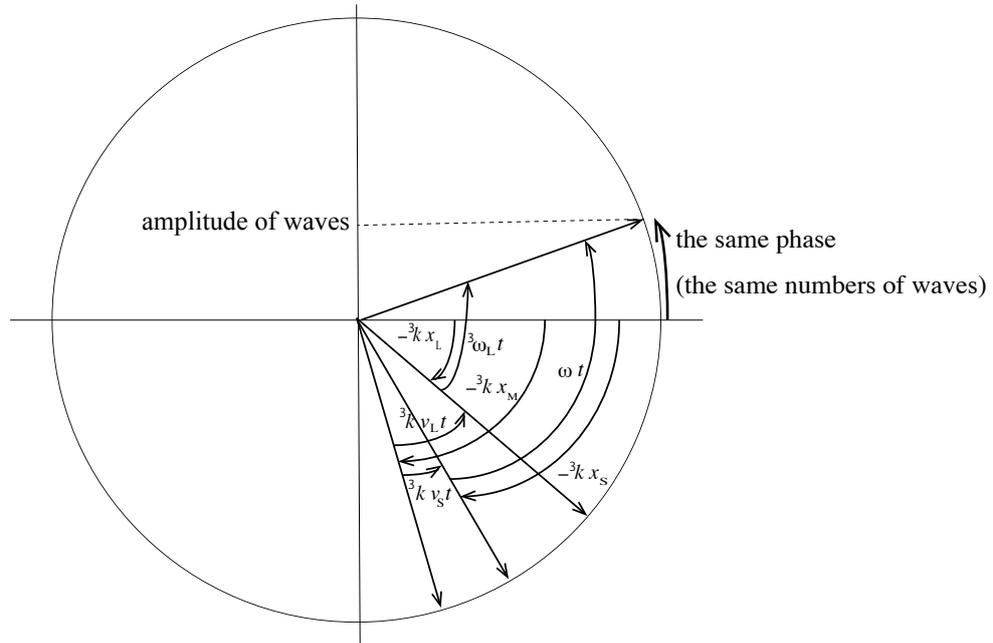
Figure 5. Phase invariance of the sound wave across K_L and K_S in Case 3. Both of the listener and the source move in the medium. The projection to the vertical axis represents the amplitude of sound wave. From Figure 6, $(-^3kx_S) - (-^3kx_L) = -^3k(v_L - v_S)t$.
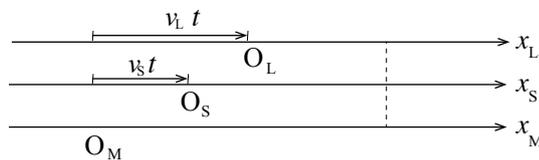
Figure 6. The relation among x_M, x_L, and x_S in Case 3. The Galileo transformation equation is $x_S - x_L = (v_L - v_S)t$.

respectively, and the Galileo transformation equation, $x_S - x_L = (v_L - v_S)t$, holds at the same time t as seen from Figure 6. The invariance of the phase, $\varphi(x_S, t) = \varphi(x_L, t)$, deduces $u_S + v_S = u_L + v_L$, either side of which indicates the sound velocity with respect to the medium, u, and thus $u_S = u - v_S$ and $u_L = u - v_L$ can be confirmed. Conversely, the phase invariance, $\varphi(x_S, t) =$ $\varphi(x_L, t)$, is deduced by assuming $u_S + v_S = u_L + v_L$ in (10). Thus, relative motion with respect to a third entity, the medium, is involved. In contrast to the acoustic Doppler effect, the relativistic Doppler effect of light is caused by the relative motion of the source and the observer. Figures 2, 4, and 6 cannot be applied to explanation for the relativistic Doppler effect. However, the formu-

lae of the acoustic Doppler effect in Cases 1 and 2 are also applicable to the relativistic Doppler effect under the following restrictions. We replace K_L and K_M for K_O and K in Case 1, respectively, where K_O and K are the reference system fixed to the observer and the inertial system, respectively, and t_O, x_O, x represent time and the position measured at K_O, and the position measured at K, respectively. The source of light is stationary at K. First, only the relative velocity of the observer and the source of light, $v_S - v_O$, is key to the relativistic Doppler effect, where v_O and v_S are the velocity of the observer and the source of light measured at the inertial system, respectively. Second, v/c can be neglected in comparison with 1, where v expresses $|v_S - v_O|$ and c is the velocity of light. We get the Galileo transformation equations,

$$x_O = x - v_O t$$

and

$$t_O = t,$$

from the Lorentz transformation equations,

$$x_O = \frac{x - v_O t}{\sqrt{1 - \dfrac{v^2}{c^2}}}$$

and

$$t_O = \frac{t - \dfrac{v}{c^2}x}{\sqrt{1 - \dfrac{v^2}{c^2}}}.$$

3 Diagrams for explaining superposition of traveling waves

Incident wave and reflected wave interfere constructively or destructively. The condition for constructive or deconstructive interference depends on the difference of path length. A question may arise: How is the difference of path length related to the phase difference of incident and reflected waves ? Suppose the amplitude of a sinusoidal wave traveling to the right along the x-axis is

$$\Psi_i(x,t) = A \sin(\omega t - kx), \qquad (11)$$

where A is the peak amplitude, ω is the angular frequency, and k is the wavenumber of the wave. When the wave of the same frequency and amplitude is reflected from a wall at the position x, the reflected wave is traveling to the left and thus can be expressed as

$$\Psi_r(x,t) = A \sin(\omega t + kx + \delta), \qquad (12)$$

where δ is the phase shift determined by a boundary condition. Using the trigonometric identity for the sum of two sines, we can obtain the representation of the superposition of the two waves, in which the frequency is original but the amplitude is proportional to the cosine of $kx + \delta/2$. Therefore constructive or deconstructive interference depends on the phase difference between the two waves regardless of ωt.

Description using equations does not necessarily help students without the preliminary knowledge of the trigonometric

identities derive $\cos(kx + \delta/2)$. Diagrams representing the phase of waves are effective to the derivation of the sum of two sines. According to the idea of proofs without words [5], the sum-to-product identities of the trigonometric function can be realized by the relation between the phase of the two waves on the basis of elementary geometry. The phase functions of the incident and reflected waves are

$$\varphi_i = \omega t - kx \qquad (13)$$

and

$$\varphi_r = \omega t + kx + \delta, \qquad (14)$$

respectively. Figures 7 and 8 can be drawn in the same way as Figures 1, 3, and 5. The

radius of the circle in these figures is A. In Figure 9, the projections of the vectors, $\overrightarrow{OI}$ and $\overrightarrow{OR}$ to the vertical axis represent the amplitude of the incident and reflected waves, respectively. Thus, the projection, ON, of the vector, $\overrightarrow{OM}$, is half the amplitude of the sum of these waves, because M is the middle point of the line segment, IR. From the geometric relations,

$$ON = OM \sin\left(\frac{\varphi_r + \varphi_i}{2}\right)$$

and

$$OM = OI \cos\left(\frac{\varphi_r - \varphi_i}{2}\right),$$

we obtain the desired identity

$$A \sin \varphi_i + A \sin \varphi_r = 2A \cos\left(\frac{\varphi_r - \varphi_i}{2}\right) \sin\left(\frac{\varphi_r + \varphi_i}{2}\right), \qquad (15)$$

where $(\varphi_r - \varphi_i)/2$ is $kx + \delta/2$ and $(\varphi_r + \varphi_i)/2$ is $\omega t + \delta/2$.

4 Concluding remarks

The essence of relative motion is how each observer in a different reference system experiences the same phenomenon. On the subject of relative motion, description using equations is difficult for students to realize intuitively. As shown in the previous study [6], visual aid using diagrams is effective to understanding relative circular motion in a two-body system. Even though subjects are phenomenon caused by

relative motion along a straight line, the mechanism of the acoustic Doppler effect is not necessarily easy for students to understand using intuition. Diagrams representing the phase of the sound wave indicate the physical meaning of the invariance of the phase across the coordinate reference systems at the same time. Figures 1, 3, and 5 show clearly the Galileo transformation, the sound velocity relative to the listener and to the source, and the invariance of wave-
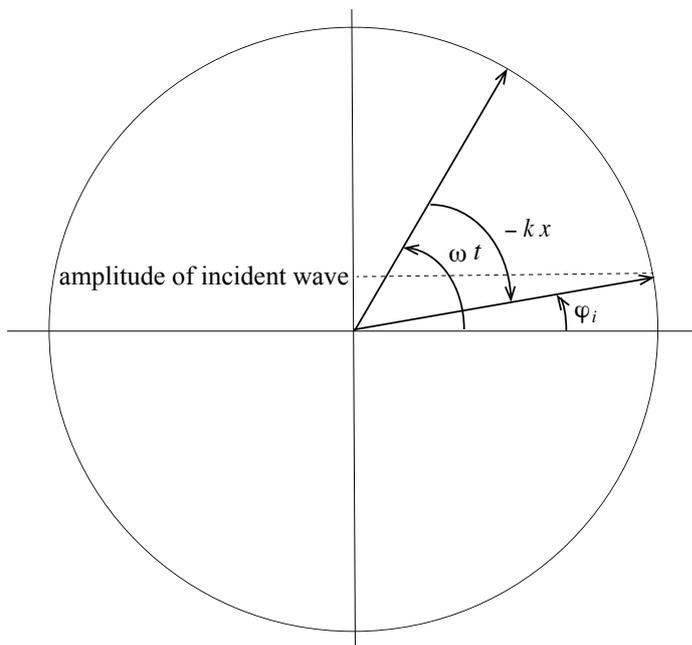
Figure 7. Phase of the incident wave.
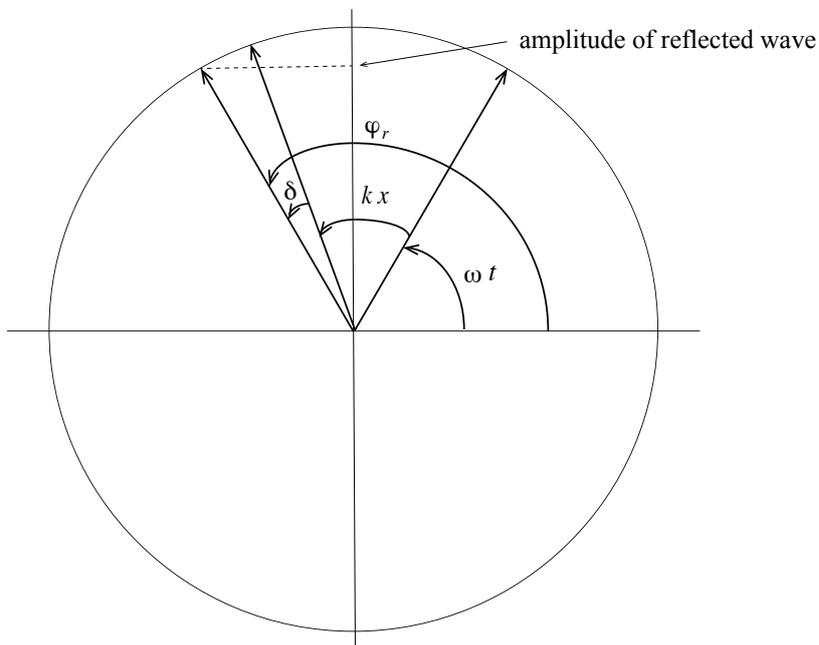
Figure 8. Phase of the reflected wave.

length of the sound wave. The superposition of two traveling waves is determined by the sum and difference between the phase angle of these waves. To understand con-
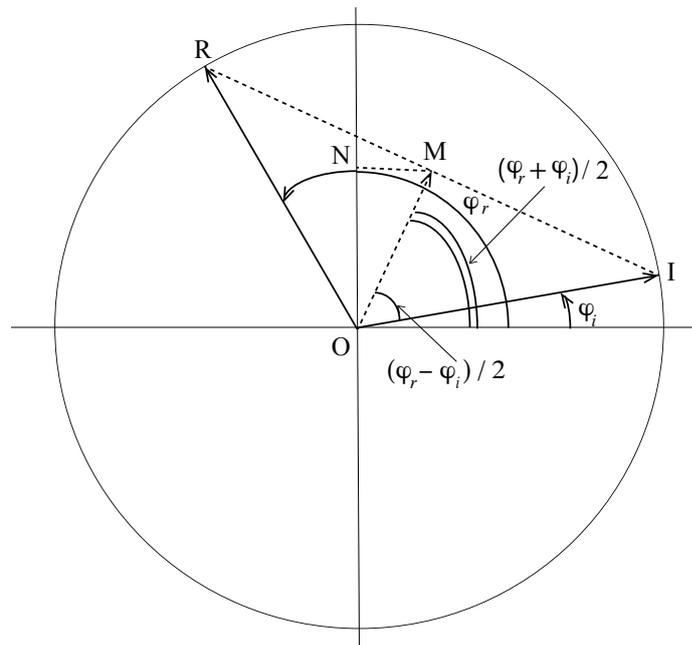
Figure 9. Relation between the phase of the incident and reflected waves.

structive or deconstructive interference, Figures 7, 8, and 9 are instructive diagrams with the aid of elementary geometry. Thus, preliminary knowledge of addition theorems of circular functions is not necessary. The present approach is a pedagogical application of elementary mathematics to fundamental physics from the viewpoint of a cross-curricular study of mathematics education and physics education.

Acknowledgements

References

[1] Born M 1962 *Einstein's Theory of Relativity* (New York, Dover) p 124.

[2] Crawford Jr F S 1965 Berkeley Physics Course Vol. 3 *Waves* (New York: McGraw-Hill) p 160.

[3] Kobayashi Y 2010 A simple derivation of the Doppler effect formulae *Inst. Phys. Singapore Phys. Update* **8** pp 22–4.

[4] Kobayashi Y 2011 Significance of the wavelength in explanation of Doppler effect *J. Phys. Educ. Soc. Jpn.* **59** pp 108–11 (in Japanese).

[5] Nelsen R B 2000 *Proofs without Words More Exercises in Visual Thinking* (Wash-

ington: The Mathematical Association of America) p 50.

[6] Kobayashi Y 2010 Simple demonstration in the classroom: intuitive visu-alization of relative circular motion in a two-body problem *Phys. Educ.* **45** pp 654–7 [Erratum 2011 **46** p 384].

Is interstellar travel to an exoplanet possible?

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

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

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

Submitted on 03-12-2019

Abstract

In this article, we examine the possibility of interstellar travel to reach some exoplanet orbiting around a star, beyond our Solar system. Such travels have been in the realm of science fiction for long. However, in the last 50 years or so, this question has gained further impetus in the mind of a man on the street, after the interplanetary travel has become a reality. Of course, the distances we shall encounter for travel to even some of the nearest stars outside the Solar system could be millions of time larger than those till now covered, to reach our celestial neighbours within the Solar system. Consequently, the time and energy requirements for such a travel could be immensely prohibitive. The questions we want to explore here are: What could be the possible limitations, if any, for such interstellar travels, and could humans ever undertake such a voyage, with hopefully a positive outcome? What could be a possible scenario for such an adventure in a near or even distant future? And what could be the reality of UFOs – Unidentified Flying Objects – that get reported in the media from time to time?

1   Introduction

In the last three decades many thousands of exoplanets, planets that orbit around stars beyond our Solar system, have been discovered. Many of them are in the habitable zone, possibly with some forms of life evolved on a fraction of them, and hopefully, the existence of intelligent life on some of them. Can we ever get in physical contact with the extra-terrestrials, assuming they are there? Radio communication over inter-

stellar distances is one possibility [1]. What about the possibility of humans ever visiting "them"? Or an even more pertinent question first – Is an interstellar space travel to an exoplanet around a star beyond our Solar system possible?

In last 50-60 years, the mankind, first time in its history, has not only ventured into outer space, humans have successfully stepped on the Moon, the first time ever on another celestial body. Rover explorations of the surface of Mars have been made many times, probes have landed on Venus, and many other missions have been sent to other planets. The Galileo spacecraft that entered orbit around Jupiter, made a number of close flybys to study Jupiter's satellite Ganymede. In the Cassini-Huygens mission, while Cassini orbited Saturn and studies its rings before it plunged into Saturns atmosphere, the Huygens probe successfully landed on Saturn's moon Titan.

In recent years, India too has sent two missions, Chandrayan-1 and Chandrayan-2, to the Moon, and Mars Orbiter Mission (MOM), India's first interplanetary mission, has successfully reached Mars. A third mission to the Moon is now being planned, and other interplanetary missions are in the offing. Perhaps in a decade or so, India may also achieve a human landing on the the Moon. After that one could imagine such manned trips to Mars. Other countries are also planning such expeditions in near future. As for the Jovian planets like Jupiter or Saturn, manned missions if any, will have

to have bases on one of their satellites, e.g. Ganymede or Titan, as the planets themselves are all gaseous, lacking a solid surface to make a landing.

This begs a question: Could man possibly ever travel to distant stars to visit some exoplanets, perhaps in a habitable zone, to possibly encounter some extraterrestrial life? After all, a mere century back, a trip to the Moon, culminating in a human landing on it, looked as much impossible and such accounts in science fiction seemed to be just a fig of imagination, as an interstellar travel to an exoplanet may appear now. Such analogies though may have their own justification grounds, but the fact remains that the distances involved in interstellar travel are immensely larger. The nearest star outside the solar system (Proxima Centauri) is as many times ($\sim$ a hundred million times) farther than the Moon, as the latter is compared to distance between adjacent rooms (($\sim$ 4 m) in a building. From a simple logic one could then expect that going to a star will at least be as much more difficult than going to the Moon as the going-to-the-Moon was with respect to a walk just next door within an office building. Of course, the shortness of human lifetime makes things all the more difficult. With the maximum speeds achieved so far by the spaceships within the solar system, it will require about 80,000 years on a one-way journey to this nearest star. Thus it may not look possible to reach other stars within a human lifetime, although on a theoretical basis theory

of relativity could allow one to do so. For instance, a spaceship accelerating continuously with a convenient value of g, that is the acceleration that we are used to on the surface of the Earth, could travel to the most distant parts of the universe within a human lifetime, without violating the speed-limit of c, the speed of light. In principle, interstellar travel may thus appear possible.

However, energies involved in such an endeavour would make it next to impossible. In a spaceship the fuel needed for the later parts of the journey has to be carried aboard and thus also needs to be accelerated till it is utilized. Therefore the initial mass at the start of such a voyage is exponentially larger than the final payload. With conventional chemical fuel such an arduous journey will need a fuel-mass of a whole galaxy. Even within the best possible scenario, where almost 100% of mass is converted into energy (in a typical thermonuclear reaction only about 0.7% of mass is converted into energy), one would require initial mass to be millions of times the mass of the final payload and the energy required may be worth hundreds of years of total energy consumption of the whole world. If we imagine that the energy is beamed from power plants on the Earth to the spaceship, it will again require many hundred million megawatts of power throughout the duration of such a trip, which might last for a very long time. It therefore looks that at most we might travel to other planets within our solar system but the distant stars will

ever remain within the realm of a distant dream only.

In this article, we ignore the technical aspects of the mission as technology is bound to improve rapidly over time. Further, we assume 100% efficiency of the rocket engine in converting fuel energy into kinetic energy of the exhaust, something that might not really be possible. We carry forth the possibility of our endeavour without delving into many other equally important issues such as the long term effects of cosmic radiation on the health of space travellers and their requirements for food, medical and other life-sustaining needs. We consider mainly the minimum basics of the travel, which are distance, time and energy.

2   The story so far

Till date there have been five spacecrafts that have crossed the threshold of escape velocity from the solar system and four of them are already headed towards the interstellar space.

Pioneer 10 was launched in 1972, flew past Jupiter in 1973 and became the first spacecraft to achieve escape velocity from the solar system. The contact was lost in January 2003 and is heading in the direction of Aldebaran in Taurus. Pioneer 11 was launched in 1973, flew past Jupiter in 1974 and Saturn in 1979. The contact was lost in November 1995. The spacecraft is headed toward the constellation of Aquila.

Pioneer 10, as well as Pioneer 11, carry
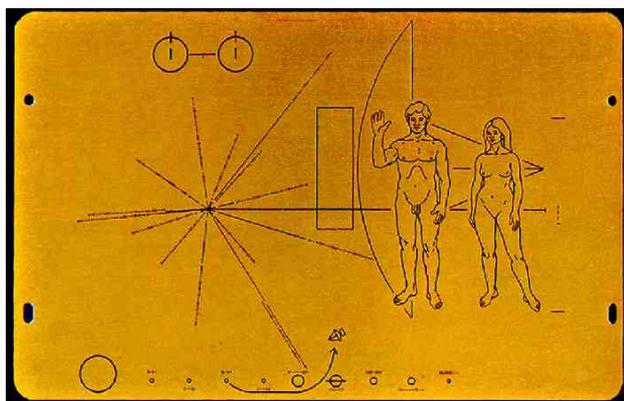
Figure 1: The message, featuring human figures along with several coded-symbols inscribed on the gold-anodized aluminium plaques, carried aboard Pioneer and Voyager spacecrafts.

gold-anodized aluminium plaques in case either spacecraft is ever found by intelligent life-forms from another planetary system. The plaques feature the human figures along with several coded-symbols that are designed to provide information about the origin of the spacecraft, and the message may hopefully survive for hundreds of millions of years during its long travel through the interstellar space. It is, thus, the artefact of mankind with the longest expected lifetime [2].

The content of the message should be clear to an advanced extraterrestrial civilization, which will have, of course, the entire Pioneer 10 spacecraft itself at its disposal to examine as well. But being the product of billions of years of independent biological evolution, they may not at all resemble humans, nor may the perspective and line-drawing conventions be the same there as

here. The human beings will perhaps be the most mysterious part of the whole message for them [2].

Voyager 1 was launched in September 1977, flew past Jupiter and Saturn, made a close approach to Saturn's moon Titan and is now at a distance of about 145 astronomical unit (au), where one au (=1.5×10^8 km) is the average distance of the Earth from the Sun. Voyager 2 was launched in August 1977, flew past Jupiter, Saturn, Uranus, Neptune and is now at a distance of about 125 au. Both probes are already past heliopause, the region where the solar wind interacts with the interstellar medium at distances around 120 au from the Sun. Voyagers are thus presently exploring the boundary between the Sun's influence and interstellar space, where nothing from the Earth has flown before, and are expected to return valuable data, hopefully, for another decade. Since the Pioneers were launched first, they had a head start on the Voyagers, but because they were travelling slower they were eventually overtaken by Voyagers.

New Horizons, launched in 2006, made a flyby of Jupiter in 2007, and then in 2015 it made a flyby of Pluto, where it flew 12,500 km above the surface of Pluto, making it the first spacecraft to explore this dwarf planet. After that, New Horizons made a flyby of Kuiper belt object 486958 Arrokoth, at $\sim$ 43 au from the Sun. New Horizons was launched with the largest-ever launch speed for a man-made object. It will, however, slow down to an escape velocity of only 2.5

au per year as it moves away from the Sun, and it will never overtake the Voyagers.

2.1   The Pale Blue Dot

The pale blue dot is a photograph of planet Earth taken in 1990 by the Voyager 1 spacecraft when the spacecraft reached ~ 6 billion km, or about 40 au (the distance of Pluto), from the Sun. This is an actual photograph (Fig. 2) of the Earth, taken from the farthest distance till now, and it appears as a tiny pale blue dot against the background of an apparent void (the faint brown band is due to the reflection of sunlight from camera optics). This picture is very significant as a perspective on our place in the cosmos as our blue planet literally pales into insignificance within the larger scheme of things. And this is the only actual image of the Earth ever seen by anybody from such a vantage point. It is both a chastening and humbling realization for us humans that our huge planet is such a tiny speck of dust seen from the distance of an outpost (Pluto!) of our planetary system. If it could be photographed from near our nearest star [Proxima Centauri], its diameter will appear about 7000 times smaller and it would be still fainter in brilliance by a factor of 50 million (with the flux-density falling as a square of distance), and that the Earth may not even qualify to be called "a tiny speck of dust" from our just next-door neighbour star.

As Carl Sagan writes [3] "The Earth is a very small stage in a vast cosmic arena. Think of the rivers of blood spilled by all

Figure 2: A panoramic (!) view of our Earth, that appears as a pale bluish dot in the centre of the image. The faint brown band across the image is due to the reflection of sunlight from camera optics.

those generals and emperors so that in glory and triumph they could become the momentary masters of a fraction of a dot. Think of the endless cruelties visited by the inhabitants of one corner of this pixel on the scarcely distinguishable inhabitants of some other corner. How frequent their misunderstandings, how eager they are to kill one another, how fervent their hatreds. Our posturings, our imagined self-importance, the delusion that we have some privileged position in the universe, are challenged by this point of pale light. Our planet is a lonely speck in the great enveloping cosmic dark. In our obscurity - in all this vastness - there is no hint that help will come from elsewhere to save us from ourselves. The Earth is the only world known, so far, to harbour life."

Table 1: An idea of the cosmic distances involved

Cosmic object	Distance
Moon	1.28 light seconds (384000 km)
Sun	500 light seconds (150 million km)
Proxima Centauri	4.24 light years
Orion Nebula	1300 light years
Centre of Milky-way	25,000 light years
Andromeda Galaxy	2 million light years
Size of Universe!	14 billion light years

And to think further that somewhere on a far-off world perhaps some intelligent being looking at this "not-even-a-speck-of-dust" could amusedly imagine that some two-legged creatures, populating that utterly insignificant part of the universe, believe that some of their ancestors (saints, gurus or prophets), confined to a minuscule part of this tiniest of dots, had figured out the grandest design of the whole Universe or even of its so-called creator – and have the audacity to claim that the creator himself or his some messenger had appeared in the form of these very two-legged creatures on their own planet. It should also humble us and put into total insignificance the occurrence of all our daily squabbles, aspirations, the desire to preserve our DNA through our children and grandchildren, political upheavals, love-affairs, wars between nations, and above all it should show us the hollowness of our religious beliefs – perhaps the greatest folly of all – and our chauvinism that we are the best of all, with an utter contempt for others who may not agree with us, and our willingness to condemn those others or even kill and die for some totally unfounded beliefs uttered or penned down by someone perhaps with good intentions but based on the limited knowledge at that moment of time, or much worse, based on a pure whim and fancy, thrust upon other gullible fellow beings.

3 Cosmic distances involved

The main challenge facing interstellar travel is the vast distances that have to be covered, requiring very high speeds as well as long travel times. The latter make it particularly difficult to design manned missions.

3.1 How far can a manned mission travel from Earth?

As one cannot travel faster than light, one might conclude that a human can never make a round-trip farther than 20 light years (1 light year $\approx 9.5 \times 10^{17}$, the distance travelled by light in one year), assuming the

traveller is active between the ages of 20 and 60. Thus one would never be able to go beyond a few star systems which exist within the limit of ~ 20 light years from the Earth. Even if we design a spaceship that can travel at $0.99c$, where $c \approx 3 \times 10^{10}$ cm/sec is the speed of light, which, from the theory of relativity is the maximum possible speed an object could ever attain, interstellar travel beyond some nearest stars seems next to impossible.

To survive for long years on a spaceship, it would be ideal to maintain a constant acceleration, $g \approx 9.8 \times 10^2$ cm sec^{-2}, the acceleration due to gravity that the humans have evolved in and are accustomed to on the Earth, with the rocket continuously accelerating the spaceship by this amount. Since we may want soft landings on the surface of the exoplanet as well as on our return to the Earth, we divide our journey into four separate stages. In the onward Journey while the spaceship is moving towards the destination, it will be accelerated in the first half of the journey, while in the second half it will have to be decelerated to attain an almost zero speed. In the same way during the return journey, it will have to be accelerated in first half of the return journey and then decelerated in the second half, for an ultimate soft landing.

3.2 The relativity comes to the rescue – time dilation

A constant acceleration of $1g$ for a year would bring the speed of spaceship ap-

proximately close to c. Therefore relativistic effects of time dilation would have to be taken into consideration. We know that time passes relatively slower by a relativistic factor $\gamma = 1/\sqrt{1-(v/c)^2}$ for an observer moving with a relative speed v. Detailed calculations show that by the time the spaceship lands back on the Earth, the time t, that would have passed on Earth, would be related to the time T, that passed on the spaceship, as (Appendix A; also see [4])

$$t = \frac{4c}{g} \sinh \frac{gT}{4c},\qquad(1)$$

a factor of 4 in the formula appears because of the four stages of the journey. During this time, the maximum relative speed the spaceship would achieve, midway of the journey, is

$$v = c \tanh(gT/c)\qquad(2)$$

The maximum distance d, of the destination that the spaceship would have arrived at and returned from, will be given by

$$d = \frac{2c^2}{g}\left[\cosh \frac{gT}{4c} - 1\right].\qquad(3)$$

The destination distance D can be expressed in terms of time t of the Earth, as

$$D = \frac{2c^2}{g}\left[\sqrt{1+\left(\frac{gt}{4c}\right)^2} - 1\right].\qquad(4)$$

It helps to remember that for $g \approx 9.8 \times 10^2$ cm sec^{-2}, time $c/g = 0.97 \approx 1$ year and the distance $c^2/g \approx 1$ light year. Table 2 gives us an idea of the time dilation involved from

Table 2: Effects of time dilation

Time on spaceship T (years)	Time on Earth t (years)	Distance reached d (light years)
1	1.01	0.065
2	2.1	0.26
5	6.5	1.85
7	11.5	4.1
10	25	11
15	95	45
20	335	165
25	1,225	610
30	4,450	2,225
40	58,800	29,400
50	0.8 million	0.4 million
60	10.2 million	5.1 million
75	0.5 billion	0.25 billion
90	23.5 billion	11.75 billion

the total duration and distance reached in a round trip, involving a constant acceleration of $1g$ for the crew. A future spacecraft, using technologies that we haven't even dreamed of, may use an engine that could sustain a constant acceleration of $1g$. Travelling even at the speed of light, visiting the stellar nursery in Orion nebula would require at least 2600 years on the earth time, while a cruise to the centre of our Milky-way galaxy will take more than 50,000 years, and a round trip to Andromeda, the nearest spiral galaxy, will need at least 4 million years. But due to the relativistic time dilation, for the traveller the time spent could be much smaller. With a $1g$ engine, a vacation trip to Andromeda may be possible within a human lifetime!

For astronauts, while those who left in their twenties might be still in their seventies at the end of the voyage, to hope for a family reunion on return, however, is out of question. Back on Earth, millions of years would have passed and entire civilizations would have come and gone. Table 2 gives time spent by the astronaut, travelling in a rocket with a constant acceleration g, for the time and distance as observed from Earth.

4   The rocket equation

If the fuel needed for the journey has to be carried aboard it also needs to be acceler-

ated till it is utilized. Therefore the initial mass, M_i, at the start of the journey is much more than M_f, the final payload mass. This is given by the rocket equation, which gives the final reachable speed v as a function of the exhaust speed u of gas/ion/light emission and $\mathcal{R} = M_i/M_f$, the ratio of the initial mass (payload + fuel) to the final mass (only payload). From the momentum conservation we have

$$M\frac{dv}{dt} = -\frac{dM}{dt}u,$$

We can integrate it

$$\int_0^v dv = -u \int_{M_i}^{M_f} \frac{dM}{M},$$

which gives

$$\frac{v}{u} = \ln \mathcal{R}.$$

or

$$\mathcal{R} = \exp(v/u). \qquad (5)$$

The exponential makes the required mass ratio increase very fast with v/u.
For example,
$\mathcal{R} = 1$, for $v = 2.3u$,
but
$\mathcal{R} = 10^{10}$, for $v = 23u$. Thus, to obtain a final speed, v close to c, it is necessary for u to be of the order of c as well, otherwise the required mass ratio will be prohibitively large.

If the motion is with a constant acceleration g, then $v = gt$ and we get

$$\mathcal{R} = \exp(gt/u). \qquad (6)$$

In a relativistic case, the rocket equation becomes (see Appendix B)

$$\frac{v}{c} = \frac{1 - \mathcal{R}^{-2u/c}}{1 + \mathcal{R}^{-2u/c}}.$$

For the mass ratio $\mathcal{R}$, we then get

$$\mathcal{R} = \left[\frac{(1 + \frac{v}{c})}{(1 - \frac{v}{c})}\right]^{c/2u},$$

or

$$\mathcal{R} = \left[\gamma\left(1 + \frac{v}{c}\right)\right]^{c/u}. \qquad (7)$$

For $v \ll u \le c$, Equation (7) reduces to the familiar non-relativistic, Equation (5).

In the case of a constant proper acceleration g, a trip planned for a travel time T of the crew, using a rocket with an exhaust speed u, would require

$$\mathcal{R} = \exp(gT/u). \qquad (8)$$

In a non-relativistic case, $T = t$, and the Equation (8) becomes Equation (6).

The power of the rocket engine needed can be calculated from the required thrust of the rocket, which is nothing but the total mass, M_i, of the spaceship (payload+fuel) multiplied by its acceleration, g. The thrust of the rocket is obtained from the exhaust mass-flow rate times the exhaust velocity [5]. For a non-relativistic case, the needed power, P, of the engine thus equals the mass-flow rate times one-half the square of the exhaust velocity. From that we get, $P = M_i gu/2$. For a relativistic exhaust speed ($u \sim c$) it becomes $P = M_i gc$.

If at the maximum speed so far achieved, which is 16 km s^{-1} for the New Horizons probe to Pluto, we could make a return trip to the Moon in a little more than half a day (ignoring the slowing down due to the Earth's gravity), a similar return trip at this speed to Proxima Centauri, the star

nearest to our solar system, will take about 160,000 years which is over 6,000 human generations, and this is of the order of time that has passed since the homo sapiens (humans) first appeared on the scene. One can thus conclude that in order to reach these interstellar destinations, one would have to travel much faster, in fact with speeds close to that of light, c, which is the maximum attainable speed for any object. Otherwise such a trip would be unimaginable. And to get close to c, we need alternative fuels.

5 Various rocket concepts

5.1 Chemical fuel rocket

Till now the chemical energy being used comes from a mixture of liquid oxygen and hydrogen, which yields 100 MJ (Mega Joules) per kg of fuel. The highest efficiency is achieved if the end products of the chemical reactions themselves can be expelled for propulsion with the energy produced. Then one will get an exhaust speed of $u = 14$ km/s. Attaining a modest maximum final value of one thousandth of the speed of light, would mean $\sim 17,000$ years of travel time for a return trip to Proxima Centauri, at a distance of 4.24 light years. This would itself require, due to the four stages of the journey, an extremely high mass ratio ($\mathcal{R} \sim (1.001)^{4c/u} \sim 1.6 \times 10^{37}$). This implies that a ten ton payload (a minimum from any standards) will need a fuel $\sim 1.6 \times 10^{44}$) gm, the mass equivalent of ~ 100 billion suns or a whole galaxy. Not at all a viable possibility,

considered from any angle. Perhaps nuclear fuel might be a better option.

5.2 Nuclear fuel - fission or fusion?

Uranium yields about 6.5×10^7 MJ/kg of energy through fission, or about a million times better than the chemical reactions. In this case, we could get an exhaust velocity, 12,000 km/s or $u = c/25$, and we could possibly attain a maximum travel speed, $v = 0.1c$, which implies $\mathcal{R} \sim 12$. Considering, however, four stages of the journey, $\mathcal{R} > 20,000$ will be needed. A round trip to the nearest star would, however, require a minimum of 170 years of travel time. Relativistic effects of time dilation would be insignificant at such speeds.

Fusion could provide ten times more energy per unit fuel mass. Despite the fact that controlled reactions of fusion of lighter nuclei have not yet been very successful, we can imagine that the technology required for it could be developed in the years to come. Banking on this assumption, one could propose the energy required for interstellar travel to come from nuclear fusion.

Using fusion of lighter nuclei, an exhaust speed of c/8.4 may become possible (see Appendix C), and that we could attain a top speed of 0.3c, requiring for a return journey a mass ratio more than 32,000. At these speeds a trip to the nearest star would require for the return journey a minimum of 60 years of total travel time, slightly more than the average working life span of a single generation. Of course a ten ton payload

will mean more than 320,000 tons of hydrogen to be carried aboard and to be converted into helium and propelled behind during the journey. This will be $\sim 2 \times 10^{17}$ MJ of energy, which is around 400 years worth of total energy consumption (5×10^{14} MJ for the year 2018) of the whole world!

The examples discussed so far were for accelerations much lower than g, the acceleration due to gravity on the Earth, an ideal value for journeys made by humans for long durations. In fact, an acceleration of $1g$ could make it possible to attain much higher speeds for the spaceship and thus substantially cut down the travel time. However, as we will show later, the mass ratio, $\mathcal{R}$, then snowballs to extremely high values, making even the nuclear fusion energy as a mode of locomotion for journey to other stars, not very promising. Thus a vision of interstellar space travel will be highly unrealistic, if we were to depend only on these energy sources.

5.3 Antimatter rockets

An antimatter rocket would have a far higher energy density and specific impulse, i.e. total impulse (or change in momentum) delivered per unit of propellant mass, than any other proposed class of rocket. When matter and anti-matter is made to fuse, the entire mass gets converted to radiation, but the technology supporting such a mode of energy production, would require matter and anti-matter to be stored at a safe distance from each other and to be able to com-bine them, a proper amount, at a proper time in order to be able to use the energy which is produced due to annihilation.

The problem, however, is that all of the current methods of manufacturing antimatter require enormous particle accelerators and produce antimatter in very small quantities, and to store antimatter, if we need a ton of magnets for one gram of antimatter, the entire idea of a lightweight way to store and carry immense amounts of energy remains no longer meaningful. Antimatter could nevertheless perhaps find use in interstellar spaceships as a way to help trigger nuclear reactions.

6 Non-rocket concepts

6.1 A scoop on the way

In a fusion rocket a huge scoop could collect diffuse hydrogen from the interstellar space and burn it on flight, using proton-proton fusion reaction and expel the fusion product to get the thrust. The idea is attractive as the fuel would be collected en route, but all attempts to design some kind of a scoop has the unfortunate effect of producing more drag than you get back thrust.

6.2 Sailing away

Solar sails are a form of spacecraft propulsion using the solar pressure, of a combination of photons and solar wind from the Sun, to push large ultra-thin mirrors to high speeds. Comets tails are pushed away from

the Sun by the same mechanism.

The momentum of a photon or an entire flux is given by $p = E/c$, where E is the photon or flux energy, p is the momentum. At 1 au the flux density of solar radiation is 1.36 kW/m^2, resulting in a pressure of $\sim 4.5\mu$Pa. A perfectly reflecting sail with 1-sq. km area could thus yield a force ~ 9 N, while the Sun's gravitational force on one ton mass there is about 6 N. As both the radiation pressure and the gravity fall with the square of distance from the Sun, a 1-ton load attached to a sail of 1-sq. km area could get pushed outward by the radiation pressure and thus escape the solar system.

Solar wind on the other hand exerts only a nominal dynamic pressure of about 3 to 4 nPa, three orders of magnitude less than solar radiation pressure on a reflective sail, and would not relatively have much effect.

A physically realistic approach would be to use the light from the Sun to accelerate. The ship would begin its trip away from the system using the light from the Sun to keep accelerating. Beyond some distance, the ship would no longer receive enough light to accelerate it significantly, but would maintain its course due to inertia. When nearing the target star, the ship could turn its sails toward it and begin to decelerate. Additional forward and reverse thrust could be achieved with more conventional means of propulsion such as rockets.

6.3 Laser sails or particle beams

Laser sails might be another way to go. Instead of relying just on the enormous amount of light given off by the Sun, laser sails to Proxima Centauri could also ride laser beams that the earthlings would fire carefully at those ships to give an extra boost, especially when sails were too far away to catch much light from our Sun. The problem with laser sails is that a lot of light needs to be used for a long time to get fast enough to get to Proxima Centauri within a human lifetime. This means very powerful and extraordinarily large lasers are needed in order to focus on sails that get farther and farther away.

An idea similar to light sails could be firing a particle beam at a spaceship that would ride that energy. The problem with laser beams is that they disperse over distance, so we could use particle beams. The beam would have to have a neutral electrical charge so as not to disperse itself over time.

6.4 Bombs!

Another idea for space travel would involve riding explosions through space. Such "pulsed propulsion" would hurl bombs behind a ship, which is shielded with a giant plate. The explosions would push against the plate, propelling the ship. Nuclear pulsed propulsion works best for really big systems. If we want to send a colony of 1,000 people to space, this might be the way to do it.

7   Some other fanciful ideas

7.1   Interstellar travel by transmission

If physical entities could be decomposed as "information", then transmitted and then reconstructed at a destination, travel at nearly the speed of light would be possible, which for the "travellers" would be instantaneous. However, sending an atom-by-atom description of (say) a human body would be a daunting task. Extracting and sending only a computer brain simulation is a significant part of that problem. "Journey" time would be the light-travel time plus the time needed to encode, send and reconstruct the whole transmission.

7.2   Generation-ships

A generation-ship is a kind of interstellar ark in which crew that arrive at the destination are descendants of those who started the journey. Generation ships are not currently feasible, because of the difficulty of constructing a ship of the enormous required scale, and the great biological and sociological problems that life aboard such a ship raises.

7.3   Suspended animation

Scientists and writers have postulated various techniques for suspended animation. These include human hibernation and cryonic preservation. While neither is currently practical, they offer the possibility of sleeper ships in which the passengers lie in-

ert for the long years of the voyage, hopefully without many after-effects.

8   Other difficulties of interstellar travel

8.1   Ex-communication!

The round-trip delay time is the minimum time taken for to-and-fro communication between the probe and the Earth. For Proxima Centauri this time would be 8.5 years. Of course, in the case of a manned flight the crew can respond immediately to their emergencies. However, the round-trip delay time makes them not only extremely distant from but, in terms of communication, also extremely isolated from the Earth. In fact the communication issue could become the biggest problem. How will the people born in an interstellar colony identify themselves with no attachment to the Earth? Will they not feel literally excommunicated from the Earth?

8.2   Hard-hitting interstellar medium

A major issue with traveling at extremely high speeds is that interstellar dust and gas may cause considerable damage to the craft, due to the high relative speeds and large kinetic energies involved. A robust shielding method to mitigate this problem would be needed. Larger objects (such as macroscopic dust grains) are far less common, but would be much more destructive. The risks of impacting such objects, and methods of

mitigating these risks, will have to be adequately addressed.

8.3 Manned missions

The mass of any craft capable of carrying humans would inevitably be substantially larger than that necessary for an unmanned interstellar probe. The requirements for food, water, medical and other life-sustaining needs of the crew will literally put huge burden on the mission. In the case of interstellar missions, given the vastly greater travel times involved, there will thus be the necessity of a closed-cycle life support system, which would last over decades. In generation ships, will there be a large enough gene pool for healthy future generations? There will be the ethical questions – Should a new-born be condemned to a life-time of journey in which he or she may have no choice whatsoever. Then there is the possibility that the new generations aboard might change their mind and abandon the mission or go elsewhere, keeping no contact with the Earth.

9 A hypothetical journey!

Let us make a hypothetical journey to Proxima Centauri, the star closest to the Solar system, at a distance of 4.24 light years. For this we expand on a scenario created by Purcell [1], with the crew always under an acceleration of $1g$, the acceleration due to gravity, so that the they "feel at home". From Equation (4) we find that the return trip will take

a total of 12 years of the earth time, with the top speed (Equation (2)) reaching $0.95c$ midway point of the journey. However, from Equation (3), the traveller would age only by about 7 years. We already saw that a chemical fuel cannot provide enough thrust as it does not give rise to large enough exhaust speed. So let us try nuclear fusion of hydrogen into helium, for which the best possible exhaust speed is $u = c/8.4$ (see Appendix C). Then assuming a 100% efficiency, the relativistic rocket equation (5) yields a mass ratio $\mathcal{R} \sim 4.7 \times 10^6$ to reach a maximum speed $0.95c$. However, if we consider the deceleration and the return journey as well, the scenario becomes impossible as the mass ratio for the nuclear fusion case swells to $\mathcal{R} \sim (4.7 \times 10^6)^4 \approx 5 \times 10^{26}$. So for a 10 ton payload, we will need a fuel mass of $\sim 5 \times 10^{33}$ gm, that is, equivalent to more than two suns. Thus one will have to tug along fuel mass equivalent to two suns or more, in order to accomplish a return trip to the nearest star beyond the Solar system. The fuel requirement could be reduced substantially if we are able to somehow achieve nuclear fusion of hydrogen into iron, the ultimate stage in the nuclear fusion, where the maimum exhaust speed becomes $u = c/7.4$ (Appendix C). In that case the fuel needed for a return journey to Proxima Centauri, with a 10 ton payload, reduces to $\sim 3 \times 10^{30}$, equivalent to the mass of ~ 500 earths. Still an impossible amount of fuel.

Though recently an earth-size planet has been found orbiting around α-Centauri

B, but it seems too close to the parent star and would be very hot and perhaps not habitable. It is estimated that to visit a habitable planet and hopefully encounter some extraterrestrial life, we may have to probe stars up to about 12 light years. For instance, Ross 128 b, a confirmed Earth-sized exoplanet, orbiting within the inner habitable zone of the red dwarf Ross 128, lies at a distance of about 11 light years from the Earth. Another exoplanet, Luyten b, orbiting within the habitable zone of the red dwarf Luyten's Star, is at a distance of 12.2 light years from our Solar system. With this in mind, let us make a hypothetical return trip to an exoplanet, say, at a distance of 12 light years. From Equation (4) we find that the return trip will take a total of 28 years of the earth time, with the top speed (Equation (2)) reaching $0.99c$ midway point of the journey. However, from Equation (3), the traveller would age only by about 10 years. For the best possible exhaust speed is $u = c/7.4$, to reach $0.99c$, the mass ratio for the nuclear fusion case swells to $\mathcal{R} \sim 2 \times 10^{34}$. So for a 10 ton payload we will need a fuel mass of $\sim 2 \times 10^{41}$ gm, that is, equivalent to ~ 100 million suns. This would imply consuming, throughout the journey of 10 years on board, on the average, fuel mass about one third of the sun every second. This means the energy that the Sun produces during its life time of $\sim 10^{10}$ years, would be consumed every three seconds to accelerate the spaceship. In fact the fuel consumption will be orders of magnitude higher in the initial

stages, being at a rate $M_i\, g/u$, that is ~ 25 suns per second. A scenario not imaginable even in the wildest of our fantasies.

Thus forgetting the chemical fuel, even the nuclear fusion could not be the source of energy for interstellar travel. And that too when we restricted travels to only a few light years within the reach of the Solar System. It is quite clear that, one would need an exhaust speed, $u \approx c$ and matter-antimatter annihilation only may provide it. To reach $0.99c$, the mass ratio in such a case may appear to be manageable, $\mathcal{R} = 14$, at least for one leg of the journey, which however, snowballs to $\mathcal{R} = (14)^4 = 40,000$ for the complete journey in four stages, implying 200,000 tons of matter and antimatter each. For the early part of the journey we will need $\sim 1.2 \times 10^{12}$ MW, about seven times more than the radiation that the Earth receives from the Sun. But with all that in gamma-rays, our problem will be not only to shield the payload but also to shield the Earth. Again, not a very promising scenario!

10   Could we? Or should we?

So far no one has created technology that is widely agreed upon as capable of caring for or preserving humans across the lifetimes it might take to get to even Proxima Centauri; it might easily take more than one lifetime to reach any star system! If that is so, mission designers might have to take procreation and family into account so that offspring of the original crew would get properly edu-

cated and trained to manage the ship in due course.

Thus a trip to our nearest star requires not only ingenious methods of propulsion and a minimum of decades en route, but also a sophisticated system of life support for the human crew to survive the journey. Not only the costs and difficulties are almost insurmountable, but they would also require almost unparalleled public and governmental support. The ultimate question then might change from – Could we to should we?

Even if the constraints imposed by the technology are ignored, the requirement of energy plays a huge constraint by itself. A huge amount of fuel would have to be put to use for such an endeavour and many generations of earthlings would have to work on such a project, putting all the available resources and the work force dedicated on something that may or may not result in a successful outcome.

There is a very strong likelihood that the mission would fail due to many other factors. We have ignored the requisites of food and water and other medicinal requirements for the crew. There is also the effect of the harmful radiation such as cosmic rays and impacts with other larger bodies. What if some deadly disease strikes? It is unlikely that living beings will be able to survive such ordeals for time periods of the order of decades.

Further we have not even considered the time and resources needed for possible research and conduction of experiments at the place of the destination, without which such a trip would not be of much advantage to us, anyway.

11   Conclusions

Taking these severe limitations into account, we can conclude that space travel, even in the most distant future, will remain confined to our own planetary system, and a similar conclusion will hold forth for any other civilization, no matter how advanced it might be, unless those extraterrestrial species have life spans order of magnitude longer than ours. Even in such a case it is unlikely that they will travel much farther than their immediate stellar neighbourhood, as each such excursion will exhaust the resources of their home planet so much that those will dwindle rather fast and there might not be much left for the further scientific and technological advancements. So the science-fiction fancy of a Galactic Empire may ever remain in our fantasies only. And as for the mythical UFOs, whose quiet appearances do get reported in the press once in a while, recent explorations have shown no evidence that any such thing could have an origination within our own solar system itself, a "quiet" return trip from a distant star is almost impossible as it could not be so quiet as the exhaust in any such trip will dazzle the sky like many suns or perhaps more like a gamma ray burst occurring, but not in a distant part of the universe, instead going off

right in our own solar backyard.

Appendix A: The distance-time relation for an accelerated motion with relativistic speeds

We can compute time T of a spaceship traveller, undergoing a proper acceleration g to achieve relativistic speeds, in terms of the time t and distance x, as measured by a set of observers stationary with respect to the launching station. We assume it to be a 1-dimensional motion, say, along the x-axis, taking $x = 0$ and $t = 0$ at the start of the journey at $T = 0$. From relativistic transformations, we have the time dilation formula, $dt = \gamma \, dT$, while for the longitudinal acceleration we have, $dv/dt = g\,\gamma^{-3}$ [4]. The equation of motion then is

$$\gamma^3 dv = g \, dt = g\gamma \, dT.$$

We can integrate it

$$\int_0^v \frac{dv}{1 - (v/c)^2} = \int_0^T g \, dT.$$

For a constant g, we then get

$$\frac{v}{c} = \tanh(gT/c),$$

which gives $\gamma = \cosh(gT/c)$.

From this we can get a relation between t and T as

$$\int_0^t dt = \int_0^T \gamma \, dT = \int_0^T \cosh \frac{gT}{c} \, dT,$$

or

$$t = \frac{c}{g} \sinh \frac{gT}{c}.$$

The distance covered is

$$\int_0^x dx = \int_0^t v \, dt = c \int_0^T \sinh \frac{gT}{c} \, dT,$$

or

$$x = \frac{c^2}{g} \left[\cosh \frac{gT}{c} - 1 \right].$$

Distance x can be expressed in terms of t as

$$x = \frac{c^2}{g} \left[\sqrt{1 + \left(\frac{gt}{c} \right)^2} - 1 \right].$$

Appendix B: The relativistic rocket equation

If in the instantaneous rest frame of the rocket, a fuel mass Δm is consumed during a proper time ΔT, to generate energy that causes the expulsion of the propellent with an exhaust speed u, with a corresponding Lorentz factor $\gamma_u = 1/\sqrt{1 - (u/c)^2}$, from the energy conservation we have, $\gamma_u \Delta m' c^2 = \Delta m c^2$, where $\Delta m'$ is the mass in the expelled fuel's rest frame. The expelled mass carries a momentum, $\gamma_u \Delta m' u = \Delta m u$ and from momentum conservation, we get

$$Mg = -\frac{dM}{dT} u,$$

Using $dT = dt/\gamma$ and $g = \gamma^3 \, dv/dt$ from Appendix A, we get

$$M\gamma^2 \frac{dv}{dt} = -\frac{dM}{dt} u,$$

or

$$\frac{dv}{1 - (v/c)^2} = -\frac{dM}{M} u.$$

We can integrate it

$$\int_0^v \frac{\mathrm{d}v}{1-(v/c)^2} = -u \int_{M_i}^{M_f} \frac{\mathrm{d}M}{M},$$

to get

$$\tanh^{-1}\frac{v}{c} = \frac{u}{c}\ln\mathcal{R},$$

which can be written as

$$\frac{v}{c} = \tanh[\ln\mathcal{R}^{u/c}] = \frac{\mathcal{R}^{u/c} - \mathcal{R}^{-u/c}}{\mathcal{R}^{u/c} + \mathcal{R}^{-u/c}}.$$

The relativistic rocket equation then is

$$\frac{v}{c} = \frac{1-\mathcal{R}^{-2u/c}}{1+\mathcal{R}^{-2u/c}},$$

or

$$\mathcal{R} = \left[\frac{(1+\frac{v}{c})}{(1-\frac{v}{c})}\right]^{c/2u} = \left[\gamma\left(1+\frac{v}{c}\right)\right]^{c/u}.$$

For a constant proper acceleration g, we substitute for v and γ from Appendix A, to get

$$\begin{aligned}\mathcal{R} &= [\cosh(gT/c) + \sinh(gT/c)]^{c/u}\\ &= \exp(gT/u).\end{aligned}$$

Appendix C: The exhaust velocity limit for a nuclear fusion rocket

In a nuclear fusion reaction of hydrogen into helium, an amount $\epsilon = 0.71\%$ of the fuel mass gets converted into energy, while for a conversion from hydrogen to iron, the ultimate stage in the nuclear fusion, the amount is $\epsilon = 0.92\%$.

The energy released by this amount could be converted into the kinetic energy

$[(\gamma_u - 1)\Delta m' c^2]$ of the expelled fuel mass, giving

$$(\gamma_u - 1)\Delta m' c^2 = \epsilon \Delta m c^2.$$

Using $\gamma_u \Delta m' = \Delta m$ (Appendix B), we get

$$(1 - \sqrt{1-(u/c)^2}) = \epsilon.$$

This yields for a nuclear fusion rocket, the best possible values for the exhaust speed, $u = c/8.4$, for $\epsilon = 0.71\%$
and
$u = c/7.4$, for $\epsilon = 0.92\%$.

References

[1] Purcell, E., Interstellar Communication, ed. Cameron, A. G. W., Benjamin Inc. (1963), p. 121-143.

[2] Sagan, C., Carl Sagan's Cosmic Connection - An Extraterrestrial Perspective, Cambridge University Press (2000).

[3] Sagan, C., Pale Blue Dot: A Vision of the Human Future in Space, Ballantine Books (1997).

[4] Datta, S., Introduction to special theory of relativity, Allied Publishers, Mumbai (1998), p. 95-100.

[5] von Hoerner, S., Interstellar Communication, ed. Cameron, A. G. W., Benjamin Inc. (1963), p. 144-159.