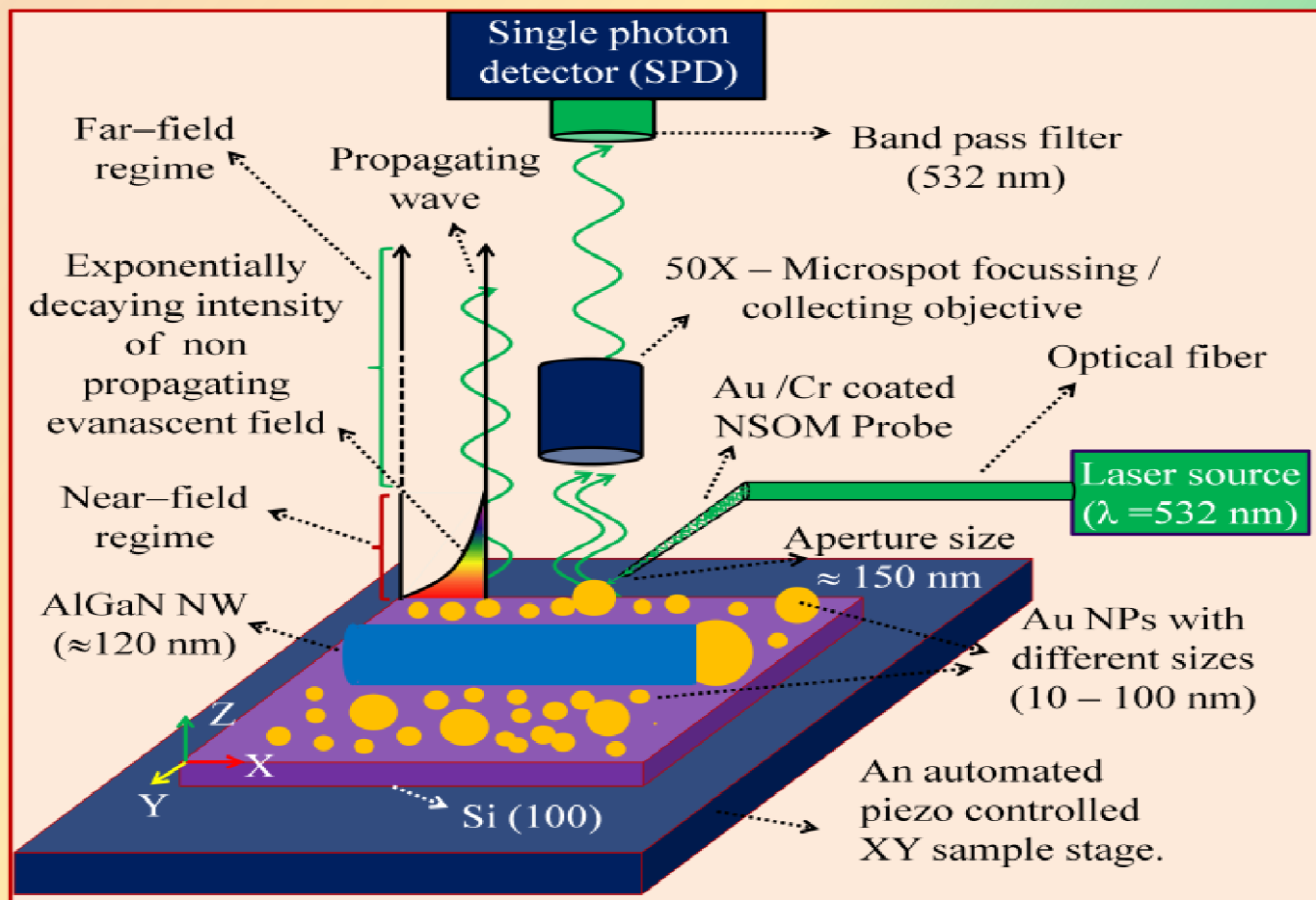
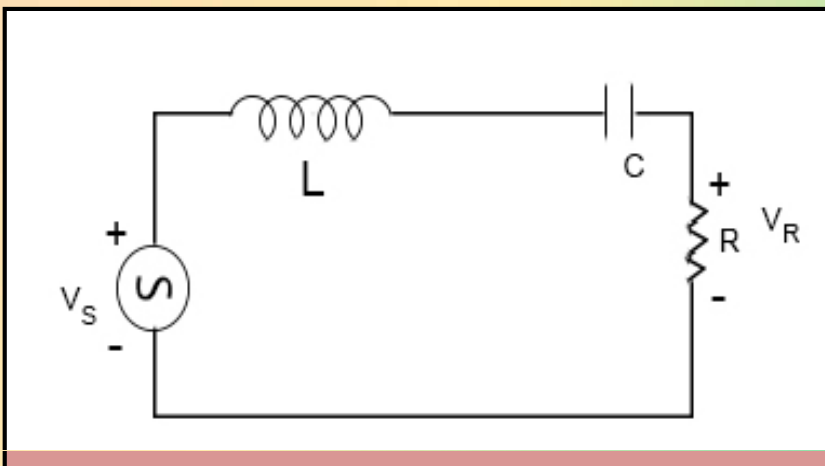
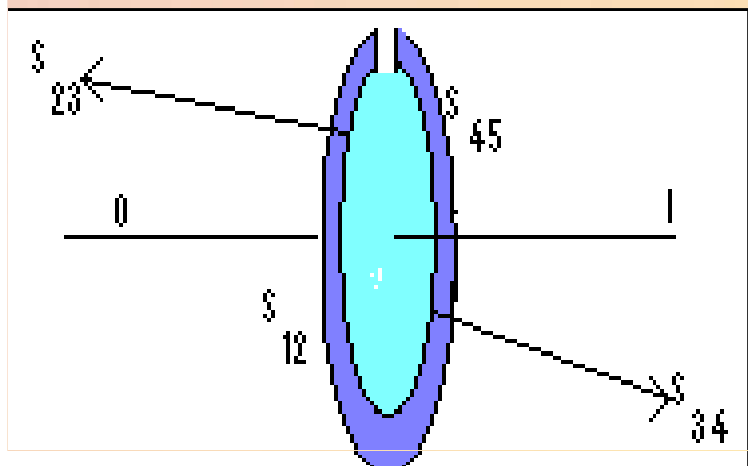


PHYSICS EDUCATION



Volume 33, Number 1**In this Issue**

- **Editorial** 01 Page
M. S. Santhanam
- **Determination of Refractive Index of a liquid using Lensmakers' formula** 06 Pages
Rabindranath Chattopadhyay
- **More than Three Hundred Years on - The Infinite Series keeps Providing Finite Results** 12 Pages
Seema Vats and Chinmoy Kumar Ghosh
- **Revisiting the physics of turning water into wine, a pedagogical approach** 04 Pages
A. C. F. Santos
- **The paradox of power loss in a lossless infinite transmission line** 15 Pages
Ashok K. Singal
- **Mermin-Wagner Theorem and its implications** 15 Pages
Debnarayan Jana
- **The genesis of the internal resistance of a battery -- a physical perspective** 07 Pages
Ashok K. Singal
- **Determination of the dimensions of the two co-axial cylindrical cavities hidden inside a mechanical black box** 10 Pages
Bhupati Chakravarti, Shirish Pathare, Saurabhee Huli
- **Optical imaging of metallic and semiconductor nanostructures at sub-wavelength regime** 06 Pages
A. K. Sivadasan, Kishore K. Madapu and Prajit Dhara
- **Electrical Circuits as the Nerve Cells** 07 Pages
M.R.Khoshbin-e-Khoshnazar

EDITORIAL

We will soon be entering a new phase in the life of *Physics Education*. From the next issue onwards, major changes to journal website, its content and format will be implemented. The website content will be more dynamic and will engage in a continuous conversation and debate about many aspects of physics and its interface with all other aspects of life in general. All these changes have become possible, thanks to the continued support of the Board of Research in Nuclear Sciences, Department of Atomic Energy, Government of India.

At the same time, we will maintain the standards of the journal. *Physics Education* will continue to

focus on pedagogical aspects of physics with special reference to India, while keeping in mind that physics, and science in general, is an international enterprise without reference to national borders.

We hope to have your continued support for the journal.

M. S. Santhanam
Chief Editor
Physics Education

Determination of Refractive Index of A Liquid Using Lensmakers' formula

Rabindranath Chattopadhyay

Haripal G.D.Institution,
W.B. India & Indian Centre for Space Physics ,
Hooghly , Kolkata, 712405.

(Submitted: 21 -11 - 2015)

Abstract

A High School Project is undertaken to demonstrate the Lensmakers' formula and use it to determine the refractive index of liquids like water and glycerine. Such a formula is derived particularly for thin lens. A compound lens, made of a pair of thin convex glass cover enveloping a liquid (water and glycerine separately in our experiment) and are glued to each other is prepared and used as a compound liquid lens. The interface equation-based thin lens formula is then applied to the cases of two such lenses and then manipulating the experimentally found out value of focal length in respective cases the refractive indices are found out.

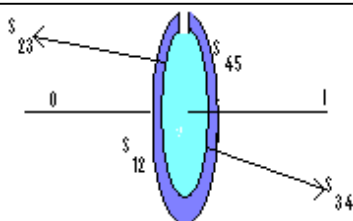


Fig.1

1. Introduction

Two thin glass-shells of same thickness and of nearly equal radii of curvature and of same aperture of rear edge are glued together keeping a small opening at one place on the edge-boundary and thus a hollow lens is formed (Fig.1). The thickness of the shells and the radii of curvature of those two spherical shells are measured prior to gluing them. Liquid under consideration is then poured into the hollow lens until it is completely filled up. It is thus made ready for use in the experiment. The compound lens thus produced is then clamped on a stand attached to an optical bench keeping the small opening upward so that

liquid poured into it does not come out of it. Keeping that lens in a fixed position on the optical bench the object-pin and the image-pin are then moved as is usually done in U-V experiment for finding the focal length of a lens. The data are put down in tabular form and are then plotted with the help of Microsoft Words Worksheet. From that graph the focal lengths in the two different cases of two compound lenses are found out. On the other hand the focal length of a lens is also found out from theoretical consideration (discussed in the following section). The two values are equated to each other and from that equation the ' μ (refractive index)' of the liquid concerned is found out.

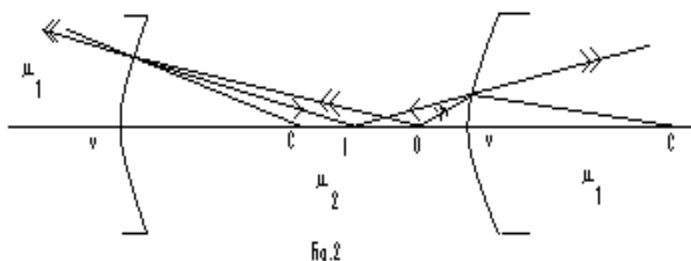
Theory : The very physical parameter which most directly determines the optical behavior of an optical medium is the refractive index (μ) of the medium. When two different optical media touch each other at a boundary the boundary is called an 'interface' between the two media. An optical system produced by encompassing an optical

medium with two such curved interfaces is called a 'lens'. Ray-tracing diagram in Geometrical Optics helps one establishing interface equation corresponding to an interface(Ref.1). Interface-equation corresponding to interface, either of concave or of convex nature(Fig.2) is given by ,

$$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{r} \dots\dots\dots(\text{for concave interface})$$

$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{-r}$ (for convex interface) where 'r' is the radius of curvature of the interface ,interfaces being considered to be spherical in nature. The interface-equations given above are derived on the basis of a set of assumptions as follow ; (i)Rays incident on interface are paraxial (ii)Rays considered are only those which pass through a small aperture while incident on interface. (iii) Interfaces are perfectly spherical

Brief interpretation of Fig.2: O → The point source situated on the common principal axis of the interfaces. I → The point image situated on the common principal axis of the



interfaces. V → The vertices ,the points where the principal axis cut the interfaces. C → The centre of curvature of the spherical interface that lie on the principal axis. P → A point on the boundary of the small aperture of the interface through which the paraxial beam passes.

For the assumptions mentioned earlier to be followed the following approximations are taken granted: OP ≈ OV = u = the object distance , IP ≈ IV = v = the image distance CP ≈ CV = r = the radius of curvature. In our experiment, in case of the liquid-filled biconvex compound lens there are actually four interfaces namely, S₁₂ ,S₂₃ ,S₃₄ ,S₄₅ and the five separate media-entity labeled 1,2,3,4 &5 respectively. Here μ₁ = μ₅ = 1 (for air μ =1) & μ₂ = μ₄ = μ_g = 1.5 (for glass μ =1.5) and μ₃ the refractive index of liquid considered the value of which is to be determined.

The interfaces S₃₄ and S₄₅ on the side of the image are both concave with respect to the object while S₁₂ and S₂₃ on the side of object are convex with respect to the same object. Therefore the set of interface-equations required to be considered for establishing the relation between the object-distance and image-distance are as follows;

$$\frac{1.5}{v} - \frac{1}{(-u)} = \frac{1.5-1}{-r_{12}} , \quad \frac{\mu_l}{v'} - \frac{1.5}{-(\tau-v)} = \frac{\mu_l-1.5}{-r_{23}} , \quad \frac{1.5}{v''} - \frac{\mu_l}{-(\tau'-v')} = \frac{1.5-\mu_l}{r_{34}} , \quad \frac{1}{v'''} - \frac{1.5}{-(\tau''-v'')} = \frac{1-1.5}{r_{45}}$$

τ being the width between S₁₂ & S₂₃ , τ' being the width between S₂₃ & S₃₄ and τ'' being the width between S₃₄ & S₄₅ . Adding all these four equations one gets,

$$\left(\frac{1}{v'''} + \frac{1}{u}\right) + 1.5\left(\frac{1}{v} + \frac{1}{(\tau-v)}\right) + \mu_l\left(\frac{1}{v'} + \frac{1}{(\tau'-v')}\right) + 1.5\left(\frac{1}{v''} + \frac{1}{(\tau''-v'')}\right) = -\left[\mu_l\left(\frac{1}{r_{23}} + \frac{1}{r_{34}}\right) + 1.5\left(\frac{1}{r_{12}} - \frac{1}{r_{23}} - \frac{1}{r_{34}} + \frac{1}{r_{45}}\right) - \left(\frac{1}{r_{12}} + \frac{1}{r_{45}}\right)\right] \dots\dots\dots(1)$$

Determination of radius of curvature of lens-interfaces and the final relation between the object-distance and image-distance: The radii of curvature of both outer sides of the lens were found out with the help of spherometer and are given below;

$r_{12}= 6.61$ cms. , $r_{45}= 5.80$ cms. .As the glass thickness was found to be 0.15cm. the inner interfaces' radii are given by $r_{23}= 6.46$ cms. and $r_{34}= 5.65$ cms. Here $\tau = 0.15\text{cm} = \tau''$ and $\tau' = 0.50\text{cm}$. (as measured). Hence $\tau, \tau', \tau'' \ll v, v', v''$ and are therefore neglected in Eqn.(1) and putting these values in the same equation one gets,

$$\left(\frac{1}{v'''} + \frac{1}{u}\right) = -\left[0.5\left(\frac{1}{6.61} + \frac{1}{5.80}\right) + 1.5\left(\frac{1}{6.46} + \frac{1}{5.65}\right) + \mu_l\left(\frac{1}{6.46} + \frac{1}{5.65}\right)\right] = -[0.3318\mu_l - 0.3358] \dots\dots\dots(2)$$

Table I(U-V values for water lens)

Serial No.	Fixed position of lens	Position of object	Position of image	Object(u) distance	Image(v) distance	Focal(f) length	Mean(\bar{f}) of f(cms)
1	70 cms.	85 cms.	50 cms.	15 cms.	20 cms.	- 08.600	-09.500
2	70 cms.	90 cms.	52 cms.	20 cms.	18 cms.	- 09.474	
3	70 cms.	95 cms.	54 cms.	25 cms.	16 cms.	- 09.756	
4	70 cms.	100 cms.	54.5 cms.	30 cms.	15.5 cms.	- 10.219	
5	70 cms.	105 cms.	57 cms.	35 cms.	13 cms.	- 09.479	
6	70 cms.	110 cms.	57.5 cms.	40 cms.	12.5 cms.	- 09.524	

TableII(U-V values for glycerine lens)

Serial No.	Fxd.posn. of lens	Position of object	Position of image	Object(u) distance	Image(v) distance	Focal(f) length	Mean(\bar{f}) of f(cms)
1	70 cms.	80 cms.	52.0 cms.	10 cms.	18.0 cms.	-6.43	-06.777
2	70 cms.	82 cms.	56.0 cms.	12 cms.	14.0 cms.	-6.46	
3	70 cms.	88 cms.	58.8 cms.	18 cms.	11.2 cms.	-6.90	
4	70 cms.	92 cms.	59.8 cms.	22 cms.	10.2 cms.	-6.97	
5	70 cms.	94 cms.	60.2 cms.	24 cms.	09.8 cms.	-6.90	
6	70 cms.	97 cms.	60.8 cms.	27 cms.	09.2 cms.	-6.86	
7	70 cms.	100 cms.	61.0 cms.	30 cms.	09.0 cms.	-6.92	

Using the well-known formula for finding the focal length of a thin biconvex lens in case of real image $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$ (2) (as obtained from the definition of principal focus)

the focal length of two compound liquid lenses are found out and are tabulated in the same table above(Table I & II).Moreover the focal lengths of the two liquid lenses are also found out from graphs

plotted for u versus v (Fig.3 & 4). Now combining Eqns.(1) & (2) one gets

$$[0.3318\mu_l - 0.3358] = \frac{1}{f} \dots\dots\dots(3)$$

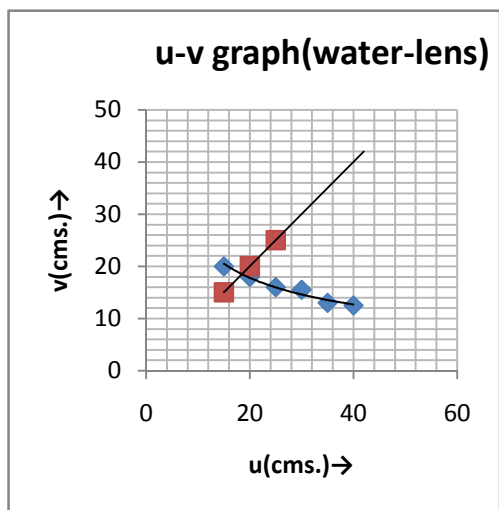


Fig.3

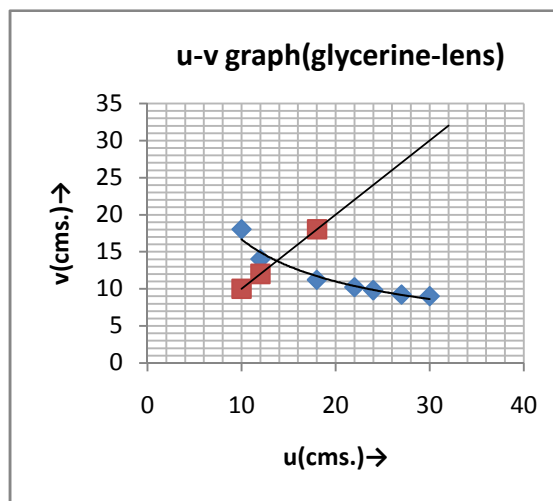


Fig.4

The values of 'f', as found out from graph are '9.15' and '9.25' respectively. Putting these values in Eqn.(3) one gets the corresponding values of 'μ_w' (Table III). Similarly the values of 'f' for glycerine-lens from graph are found to be equal to '6.9' and '6.95' respectively. The corresponding values of 'μ_{gl}', as obtained putting these values in Eqn.(3) are given in the same table.

Table-III

Liquid in lens	Value of μ From expt.	Value of μ(1) From graph	Value of μ(2) From graph	Mean Value of μ	Standard Value of μ	% of deviation
Water	1.329	1.341	1.338	1.336	1.330	+0.45%
Glycerine	1.457	1.450	1.446	1.451	1.466	-1.00%

The value of 'μ' of water is thus experimentally found to be equal to '1.336' which is very close to the standard reference value of it at normal temperature. The value of 'μ' of glycerine, as found

out here in this experiment is '1.451'. Glycerine, taken in this experiment was 95% solution by weight for which at room-temperature, as was recorded during our experiment i.e. 20°C the

reference standard value (Table IV) is '1.46597' which roughly may be taken as '1.466'(Ref.3).From the result thus obtained it is observed that our experimental result deviates less than or equal to 1% from reference standard value of the refractive indices of the liquids concerned and therefore this experimental method may be regarded as quite a reliable method of finding refractive index of a liquid.

Conclusion: The result of the experiment described above implies that on one hand the object-distance-image-distance relationship based on interface equations for thin lens is verified and on the other hand this method of experiment may be adopted for finding out the refractive index of a liquid.The advantage of this method of determination of focal length through interface-equation over other usual method in Geometrical Optics is that this interface equation-based formula may quite satisfactorily be applied thick lens as well.Moreover if instead of glass semi-rigid transparent plastic is used as enclosure of

References:

1.'Physics the Pass Word",Rabindranath Chattopadhyay,South Asian Publishers Pvt.Ltd.,New Delhi,2004,pp.28-31,ISBN 81-7003-271-7

liquid to produce compound liquid lens it will resemble certainly to a large extent to the lens-system within eyes of a living being or more specifically to human eye.Using such a lens in an experiment as described in this article in a more sophisticated way and performing the experiment altering varieties of other variables such as pressure etc. many more possibility regarding the focal length depending on other factors may be explored too.

Acknowledgement: The author would like to gratefully acknowledge the support that Mr.Swadesh Kumar Das ,TIC (Headmaster) of Haripal G.D.Institution by providing the author with the opportunity to use the physics-laboratory for performing the experiment.The author likes also to acknowledge the support Mr.Mriganka Patra and Mr. Tanmoy Ghosh (both H.S.Students) extended to the author by supplying with the lens used in this experiment.

2. G.W.C. Kaye& T.H.Laby,*Tables of Physical & Chemical Constants and some Mathematical Functions*,Longman,London(1960)
3. [PDF]Refractive Index of Glycerine-Water-Solutionsat 20°C (69°F),www.dow.com/webpps/lit/litorder.asp?file path= glycerine/pdf/...

Appendix-I**Table-IV(Ref.3)**

Refractive Index of Glycerine-Water Solutions at 20°C (69°F)					
Glycerine % by Weight	Refractive Index <i>n</i>D20	Difference for1%	Glycerine % by Weight	Refractive Index <i>n</i>D20	Difference for1%
100	1.47399	0.00165	50	1.39809	0.00149
99	1.47234	0.00163	49	1.39660	0.00147
98	1.47071	0.00161	48	1.39513	0.00145
97	1.46909	0.00157	47	1.39368	0.00141
96	1.46752	0.00156	46	1.39227	0.00138
95	1.46597	0.00154	45	1.39089	0.00136
94	1.46443	0.00153	44	1.38953	0.00135
93	1.46290	0.00151	43	1.38818	0.00135
92	1.46139	0.00150	42	1.38683	0.00135
91	1.45989	0.00150	41	1.38548	0.00135
90	1.45839	0.00150	40	1.38413	0.00135
89	1.45689	0.00150	39	1.38278	0.00135
88	1.45539	0.00150	38	1.38143	0.00135
87	1.45389	0.00152	37	1.38008	0.00134
86	1.45237	0.00152	36	1.37874	0.00134
85	1.45085	0.00155	35	1.37740	0.00134
84	1.44930	0.00156	34	1.37606	0.00134
83	1.44770	0.00160	33	1.37472	0.00134
82	1.44612	0.00162	32	1.37338	0.00134
81	1.44450	0.00160	31	1.37204	0.00134
80	1.44290	0.00155	30	1.37070	0.00134
79	1.44135	0.00153	29	1.36936	0.00134
78	1.43982	0.00150	28	1.36802	0.00133
77	1.43832	0.00149	27	1.36669	0.00133
76	1.43683	0.00149	26	1.36536	0.00132
75	1.43534	0.00149	25	1.36404	0.00132
74	1.43385	0.00149	24	1.36272	0.00131
73	1.43236	0.00149	23	1.36141	0.00131
72	1.43087	0.00149	22	1.36010	0.00131
71	1.42938	0.00149	21	1.35879	0.00130
70	1.42789	0.00149	20	1.35749	0.00130
69	1.42640	0.00149	19	1.35619	0.00129
68	1.42491	0.00149	18	1.35490	0.00129
67	1.42342	0.00149	17	1.35361	0.00128
66	1.42193	0.00149	16	1.35233	0.00127
65	1.42044	0.00149	15	1.35106	0.00126
64	1.41895	0.00149	14	1.34980	0.00126
63	1.41746	0.00149	13	1.34854	0.00125
62	1.41597	0.00149	12	1.34729	0.00125
61	1.41448	0.00149	11	1.34604	0.00123
60	1.41299	0.00149	10	1.34481	0.00122
59	1.41150	0.00149	9	1.34359	0.00121
58	1.41001	0.00149	8	1.34238	0.00120
57	1.40852	0.00149	7	1.34118	0.00119
56	1.40703	0.00149	6	1.33999	0.00119
55	1.40554	0.00149	5	1.33880	0.00118
54	1.40405	0.00149	4	1.33762	0.00117

More Than Three Hundred Years on - the Infinite Series Keeps Providing Finite Results

Dr. Seema Vats¹ and Dr. Chinmoy Kumar Ghosh²

1. Assistant Professor , Department of Physics,
Motilal Nehru College (Day),
University Of Delhi,
New Delhi -110021

Email: (seema_sharmas@yahoo.co.in)

2. Regional Director, IGNOU Regional Centre,
Dwarka, New Delhi – 110077

Email: (ckghosh@ignou.ac.in)

(Submitted: 09-05 - 2016)

Abstract :

The world of mathematics got enriched by way of introduction of Taylor Series Expansion(TSE) in 1715 which is just above three centuries ago. Ever since its advent it had been making inroads into the world of physics and facilitating explanation of several phenomenon.

The article is an attempt to sensitize the readers, in particular the undergraduate students of physics about the wide range of application of TSE. Here we have chosen examples from the areas of Oscillation and Waves, Motion under Resistive Force and Alternative Current Circuit. Using TSE as a vehicle we have made an effort to draw linkages between mathematics and physics.

Introduction

In 1715, Brook Taylor had developed in *Methodus incrementorum directa et inversa* the calculus of

finite differences, which *inter alia* consisted the famous infinite series named after him. It was the culmination of the works of James Gregory, Leonhard Euler, Brook Taylor and Colin Maclaurian (Figs.1-4).



James Gregory (Fig. 1)



Leonhard Euler (Fig. 2)



Brook Taylor (Fig. 3)



Colin Maclaurian (Fig. 4)

Before going into the mathematical descriptions, we would like to express a concern which is about the about various mathematical derivations which they come across in the study of physics. Mathematics is not only to be used for the explanation of several phenomenon in physics but the learners are to be sensitized to ponder over examples, where in a way physics emerges from mathematics. The Taylor series expansion presents many such situations .

We have picked up examples from diverse areas of physics which are very much the integral parts of the undergraduate curriculum, and the vital link between them is TSE. The idea is to revisit the prolific mathematical tool as a facilitation to arrive at finite results using an infinite series.

learners of physics at undergraduate level suffering from a fear psychosis particularly

The Taylor Series can be used to represent any function $f(x)$ that is infinitely differentiable about a point x_0 called the expansion point and the series is

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2!}(x - x_0)^2 f''(x_0) + \dots \tag{1}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} (x - x_0)^n f^n(x_0)$$

Where f' - first derivative of $f(x)$
 $f''(x)$ - Second derivative of $f(x)$
 $f^n(x)$ - n th derivative of $f(x)$

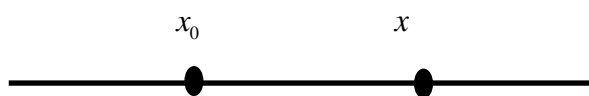


Fig. 5

x_0 - the point about which Taylor series expansion takes place;
 x - location where we evaluate the series

The location x , where the series is evaluated can be taken anywhere within the radius of convergence of the series in order to yield the correct values of $f(x)$
 The Maclaurin series is a special case of Taylor series expanded about the particular expansion point $x_0 = 0$ i.e.

$$f(x) = f(0) + xf'(0) + \frac{1}{2!}x^2 f''(0) + \dots$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} x^n f^n(0)$$

(2)

Euler discovered the Maclaurin series for e^x

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots,$$

$$-\infty < x < \infty \quad (3)$$

Often prolific examples of Maclaurin series are

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

$$-\infty < x < \infty \quad (4)$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots$$

$$-\infty < x < \infty \quad (5)$$

Taylor series finds many applications in several areas of physics which we shall present here –

1. Simple Pendulum

A simple pendulum acts like a harmonic oscillator with period dependent on l (length of pendulum) and g (acceleration due to gravity) for sufficiently small angular amplitudes, i.e. θ . With no friction mechanical energy is conserved. Total mechanical energy is a combination of kinetic energy and gravitational potential energy. As the pendulum swings back and forth, there is constant exchange of energy between kinetic and gravitational potential force.

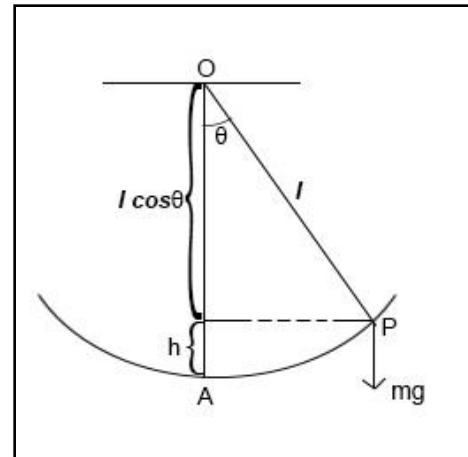


Fig 6 : A simple pendulum with a bob of mass ' m ' at a position having angular displacement, θ .

Refer to Fig 6.

O is the point of suspension of a simple pendulum, OA is the mean position. The effective length of the pendulum is l . The value of the time period of oscillation of the pendulum can be derived in many ways. Here, we shall apply Taylor series to arrive at the desired result.

The potential energy $V = mgh$

Expression h in terms of θ, l we get

$$V = mgl(1 - \cos \theta).$$

Using Eq. 5,
$$\cos \theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} \quad (6)$$

For small θ , we consider only upto the second term, to get

$$V = mgl \left[1 - \left(1 - \frac{\theta^2}{2} \right) \right]$$

or

$$V = mgl \frac{\theta^2}{2} \quad (7)$$

Kinetic energy, $T = \frac{1}{2}l\left(\frac{d\theta}{dt}\right)^2$

or $T = \frac{1}{2}ml^2 \left(\frac{d\theta}{dt}\right)^2 (\because I = ml^2)$ (8)

According to the principle of conservation of energy

$$\frac{1}{2}ml^2\left(\frac{d\theta}{dt}\right)^2 + mgl(1 - \cos\theta) = a, \text{ constant}$$

or $\frac{ml^2}{2}\left(\frac{d\theta}{dt}\right)^2 + mgl(1 - \cos\theta) = a \text{ constant}$ (9)

Differentiating w.r.t. time, t we get

$$\frac{ml^2}{2} \cdot 2 \frac{d\theta}{dt} \frac{d^2\theta}{dt^2} + mgl \cdot \frac{1}{2} \cdot 2\theta \frac{d\theta}{dt} = 0$$

or $ml \frac{d\theta}{dt} \left(\frac{ld^2\theta}{dt^2} + g\theta \right) = 0$

$$\frac{d\theta}{dt} \neq 0,$$

in general

$$\therefore \frac{ld^2\theta}{dt^2} + g\theta = 0$$

or

$$\frac{d^2\theta}{dt^2} + \omega^2\theta = 0$$

where $\omega^2 = \frac{g}{l}$ where ω is the angular frequency

The above is the differential equation of S.H.M

Time period, $T = \frac{2\pi}{\omega}$

$$V(x) = V(x_0) + (x - x_0)V'(x_0) + \frac{1}{2!}(x - x_0)^2V''(x_0) + \frac{1}{3!}(x - x_0)^3V'''(x_0) + \dots$$
 (12)

On the right hand side, the first term is irrelevant because shifting a potential by a constant amount

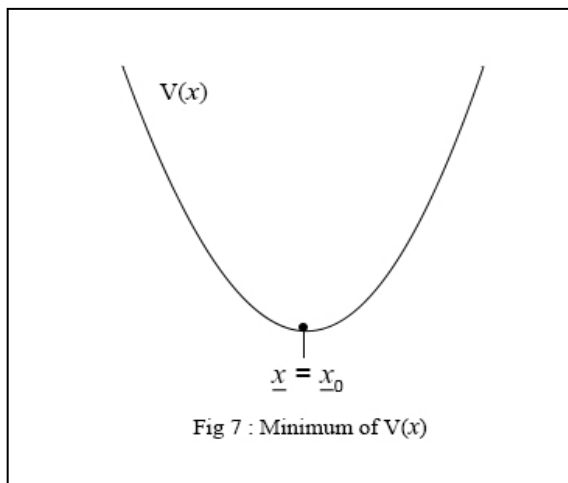
$$\therefore T = 2\pi \sqrt{\frac{l}{g}}$$

Comments : We have discussed an example of simple harmonic motion, that is, motion governed by a Hooke's law force, where the restoring force is proportional to the (negative of the) displacement. The above example was the case of a rotational analogue of Hooke's law force. Hence the restoring torque is $mgl \sin\theta$, and again in the Maclaurin series (Eq. 4) approximation of θ being very small, is $\sin\theta = \theta$. Thus the restoring torque is proportional to the negative of angular displacement, θ .

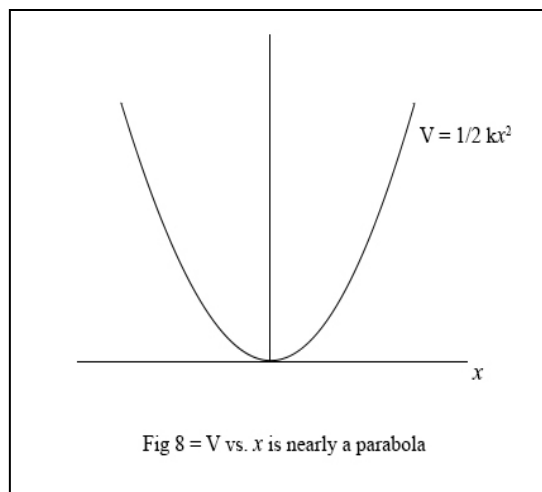
Two points emerge here. First, that a simple harmonic motion is the outcome of an approximation and second, that the basis of the approximation is Taylor series expansion. While handling the example of simple pendulum, we had started with the expression for potential energy and then had applied the Taylor series expansion to arrive at the approximation. Let us now generalize by considering an arbitrary potential, and let us see what it looks like near a local minimum. This is a reasonable place to look, because particles generally hang out near a minimum of whatever potential they are in. An exam (10) potential $V(x)$ is shown in Fig. 7. The best way to see what a function looks like in the vicinity of a given point $x = x_0$ is the Taylor series, so let us expand $V(x)$ in a Taylor series around x_0 (the location of the minimum). We have

does not change the physics. (Equivalently, the force is the derivative of the potential, and the

derivative of a constant is zero.) And the second term is zero due to the fact that we are looking at a minimum of the potential, so the slope $V'(x_0)$ is zero at x_0 . Furthermore, the $(x-x_0)^3$ term (and all higher order term) is negligible compared with the $(x-x_0)^2$ term if x is sufficiently close to x_0 , which we will assume to be the case. So we are left with $V(x) = \frac{1}{2}(x-x_0)^2V''(x_0)$. In other words, we have



to be a potential of the form, $\frac{1}{2}kx^2$ where $k = V''(x_0)$ and where we have shifted the origin of x so that it is located at x_0 . Equivalently, we are just measuring x relative to x_0 . We see that any potential looks basically like a Hooke’s law spring, as long as we are close enough to a local minimum. In other words, the curve can be approximated by a parabola, as shown in Fig 8.



This again establishes that S.H.M. is an approximation. The issue of approximation gets generalised by way of Eq. (12) which is a Taylor series expansion. Retaining the expansion upto the second order derivative term and neglecting the rest is called *Harmonic Approximation*, which finds application in several areas of physics. A very common problem is that of the interatomic potential, $V(r)$, given as under.

$$V(r) = -\frac{a}{r^6} + \frac{b}{r^{12}} \tag{13}$$

The first term on the right hand side is an attractive potential, arising out of van der waals force (dipole – dipole interaction) and the second term is a

repulsive potential having its origin is Pauli’s exclusion principle. ‘ a ’ and ‘ b ’ are constants, ‘ r ’ is the inter-atomic distance. Application of harmonic approximation provides a convenient method of determining the frequency of atomic vibration about their equilibrium position ($r = r_0$).

We know that for $r = r_0$, $V'(r) = 0$

$$V'(r) = -(-6ar^{-7}) + (-12br^{-13})$$

$$\therefore 0 = 6r_0^{-7} a - 12br_0^{-13}$$

$$\therefore r_0^6 = \frac{2b}{a}$$

$$r_0 = \left(\frac{2b}{a}\right)^{\frac{1}{6}} \tag{14}$$

$$V''(r) = -42ar^{-8} + 156br^{-14}$$

$$\begin{aligned} \therefore V''(r_0) &= -6r_0^{-8} 7a + 26br_0^{-6} \\ &= -6 \left(\frac{2b}{a} \right)^{-4} \left(7a - 26b \cdot \frac{a}{2b} \right) \\ \therefore V''(r_0) &= -6 \left(\frac{a}{2b} \right)^4 (-6a) = \frac{36}{2^4} \cdot \frac{a^7}{b^4} \quad (15) \end{aligned}$$

$$\text{And the required frequency} = \frac{1}{2\pi} \sqrt{\frac{V''(r_0)}{\mu}}$$

where μ = the reduced mass of the system.

We have discussed another application of Taylor Series Expansion wherein we have determined the frequency of interatomic oscillation, based on the knowledge of interatomic potential and the fact that the displacement of atoms from their mean position while executing the oscillations is infinitesimal.

Through these examples we observe that the Taylor series is tailor-made for the introduction of harmonic approximation (Eq.12 and as a sequel to that subsequent expression of the potential function as a second order term). It is indeed a testimony about mathematics playing a guiding role for a principle of physics.

We have made an approximation in the expansion of the potential, where we have ignored all the terms beyond the second order. A parallel can be drawn with the derivation of Eq.7, where the expansion of $\cos \theta$ has been done up to second power of θ . And then what about $\cos \theta$ being an even function! But for it, the potential energy function would not have been symmetric about $\theta =$

0 (that is the mean position of the simple pendulum in example 1). Again, the first order term in the expansion of the potential function is identically zero and this is corroborated by the absence of terms having odd powers of θ in TSE.

2. Wave motion in a stretched string

Let us consider a stretched string in which a wave is generated by shaking one end of the taut string; $f(z, t)$ represents the displacement of the string at the point z , at time t . Given the initial shape of the string, $g(z) = f(z, 0)$, the displacement at point z , at the later time t is the same as the displacement at a distance vt to the left (i.e. at $z - vt$), back at time $t = 0$ (Fig. 9) [3].

Mathematically,

$$f(z, t) = f(z - vt, 0) = g(z - vt) \quad (16)$$

The function $f(z, t)$ represents a wave of fixed shape travelling in the z direction at speed v .

We know that in order to sustain a wave motion through a medium, we need inertia to absorb and transmit the push and elasticity for restoration.

A stretched string supports a wave motion because the mass of the string provides the inertia and the tension in the string provides the elasticity. We shall now apply Newton's second law to the given situation. Let us imagine a very long string under tension T . If it is displaced from equilibrium, the net transverse force on the segment between z and $z + \Delta z$ (from Fig 9) is

$$\Delta F = T \sin \theta' - T \sin \theta$$

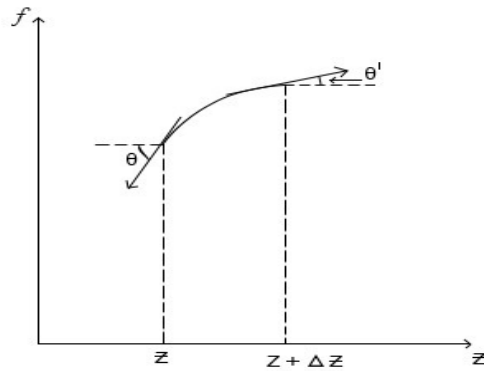


Fig 9: Variation of f vs z

where θ' is the angle the string makes with the z -direction at point $z + \Delta z$, and θ is the corresponding angle at point z . Provided that the

distortion of the string is not too great, these angles are small and we can replace the sine by the tangent.

$$\Delta F = T \tan \theta' - T \tan \theta = T \left[\left(\frac{\partial f}{\partial z} \right)_{z+\Delta z} - \left(\frac{\partial f}{\partial z} \right)_z \right]$$

Using Taylor's theorem as mentioned earlier, this can be expanded about $z + \Delta z$ as the increment,

we get

$$\left(\frac{\partial f}{\partial z} \right)_{z+\Delta z} = \left(\frac{\partial f}{\partial z} \right)_z + \left(\frac{\partial^2 f}{\partial z^2} \right)_z \Delta z$$

Considering only first two term of Taylor's series so, we get

$$\begin{aligned} \Delta F &= T \left(\frac{\partial f}{\partial z} \right)_z + T \left(\frac{\partial^2 f}{\partial z^2} \right)_z \Delta z - T \left(\frac{\partial f}{\partial z} \right)_z \\ &= T \left(\frac{\partial^2 f}{\partial z^2} \right)_z \Delta z \end{aligned}$$

If the mass per unit length is μ , Newton's second law says

$$\Delta F = \mu(\Delta z) \frac{\partial^2 f}{\partial t^2} \tag{17}$$

And therefore

$$\frac{\partial^2 f}{\partial z^2} = \frac{\mu}{T} \left(\frac{\partial^2 f}{\partial t^2} \right) \tag{18}$$

Evidently, small disturbances on the string satisfy wave equation

$$\frac{\partial^2 f}{\partial z^2} = \frac{1}{v^2} \left(\frac{\partial^2 f}{\partial t^2} \right) \dots\dots (19)$$

Where $v = \sqrt{\frac{T}{\mu}}$ is the speed of propagation.

So, we have judiciously used Taylor's series expansion to arrive at the differential equation of a wave through a stretched string.

3. Motion in presence of resistive forces

Let a ball of mass m be moving vertically up along positive y direction and let its initial velocity (pointing up) be v_0 . We assume that the only two forces acting on the ball are gravity and the air resistance, with the latter being directed oppositely to the direction of the ball's motion.

According to Newton's second law

$$\frac{mdv}{dt} = -mg - F_{air} \dots\dots\dots (20)$$

F_{air} is proportional to the first power of velocity,

i.e. $F_{air} = \mu' v$ (21)

This approximation is applicable for small bodies moving with not very high velocity.

$$\therefore m \frac{dv}{dt} = -mg - \mu' v$$

$$\frac{dv}{dt} = -g - \frac{\mu' v}{m}$$

$$\frac{dv}{dt} = -g(1 + kv),$$

Where $k = \frac{\mu'}{mg}$

$$\int \frac{dv}{(1 + kv)} = \int -g dt$$

$$\frac{1}{k} \ln(1 + kv) = -gt + C ,$$

$C =$ constant of integration

Putting initial conditions $v(0) = v_0, y(0) = y_0$ we get

$$\frac{1}{k} \ln(1 + kv) = -gt + \frac{1}{k} \ln(1 + kv_0)$$

$$\frac{1}{k} \ln \left(\frac{1 + kv}{1 + kv_0} \right) = -gt$$

$$\frac{1 + kv}{1 + kv_0} = e^{-kgt}$$

$$v = \frac{1}{k} \left[(1 + kv_0) e^{-kgt} - 1 \right] \dots\dots\dots (22)$$

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

On integrating we get

$$y = y_0 + \frac{1}{gk} \left[(1 + kv_0) \frac{1 - e^{-kgt}}{k} - gt \right] \dots\dots\dots (23)$$

For $k \ll 1$ (very small)

we can use second degree Taylor polynomial for e^{-kgt}

$$e^{-kgt} = 1 - kgt + \frac{k^2 g^2 t^2}{2} + \dots\dots\dots$$

Taylor series expansion upto second degree (24)

Using this Polynomials, we get

$$v = \frac{1}{k} \left[(1 + kv_0) \left(1 - kgt + \frac{k^2 g^2 t^2}{2} \right) - 1 \right]$$

$$= \frac{1}{k} \left[k(-gt + v_0) + k^2 \left[\frac{g^2 t^2}{2} - gtv_0 \right] + k^3 \frac{g^2 t^2}{2} v_0 \right]$$

$$= v_0 - gt + \frac{kg^2t^2}{2} - kgtv_0 + \frac{k^2g^2t^2v_0}{2} \dots\dots\dots (25)$$

If air resistance is not present, i.e. $k \rightarrow 0$ when equation (22) and (23) reduces to

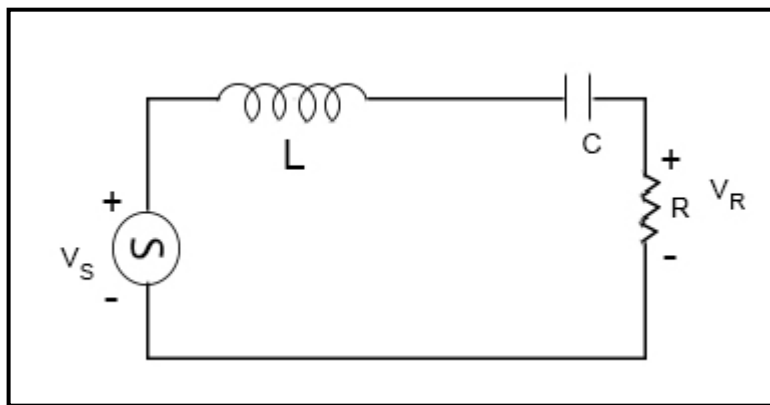
$$v = v_0 - gt \dots\dots\dots (26)$$

$$y = y_0 + \frac{1}{g}(v_0gt - \frac{g^2t^2}{2})$$

i.e. $y = y_0 + v_0t - \frac{1}{2}gt^2$ \dots\dots\dots (27)

At max height $v = 0$, \therefore time needed to reach maximum elevation is given by

4. Sharpness of Resonance - Series LCR Circuit



The impedance experienced by the source V_s is

$$Z = R + j\omega L + \frac{1}{j\omega C}$$

$$= R + j\left(\omega L - \frac{1}{\omega C}\right) \dots\dots\dots (29)$$

Which at $\omega = \omega_0 = \frac{1}{\sqrt{LC}}$ becomes equal to R .

$$t = \frac{1}{kg} \ln(1 + kv_0)$$

Using Taylor's expansion for above function $\ln(1 + kv_0)$, we get

$$t = \frac{v_0}{g} - \frac{k}{2}\left(\frac{v_0}{g}\right)^2 + \frac{k^2}{3}\left(\frac{v_0}{g}\right)^3 + \dots\dots\dots (28)$$

For $k \rightarrow 0$, $t \rightarrow \frac{v_0}{g}$, which is the well-known result for free fall under gravity. Depending on the value of 'k', one can decide about the number of terms up to which the series has to be extended on the right hand side.

This is the condition of resonance, at which the power dissipated in the LCR circuit is equal to the power dissipated by the resistor.

With symbols having usual meanings and since voltage across a resistor ($V_R \cos \omega t$) and the current through it ($I_R \cos \omega t$) are in phase, the Power at time t' is given by,

$$P(t) = V_R I_R \cos^2 \omega t$$

The Average Power becomes

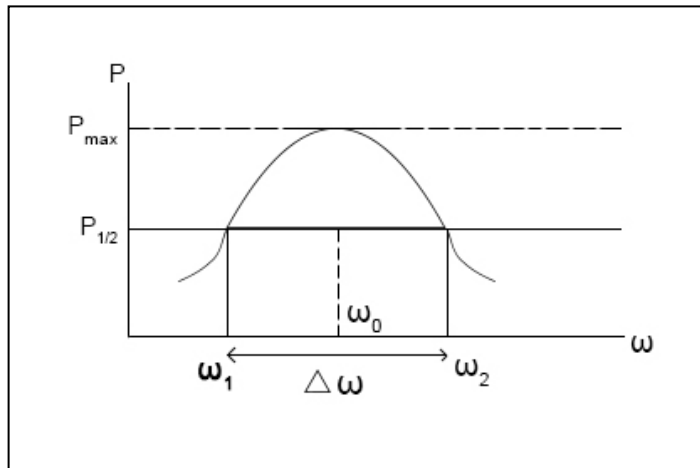
$$P(\omega) = \frac{1}{2} V_R I_R$$

$$(\because \cos^2 \omega t = \frac{1}{2})$$

$$= \frac{1}{2} \frac{V_R^2}{R} \quad \dots (30)$$

The maximum power is dissipated at the resonance frequency.

We shall now use the power vs frequency graph(Fig.11) to determine the sharpness of resonance. It measures the peakedness of the P vs



ω graph around the resonance frequency $\omega = \omega_0$. For this we need to find the band width $\Delta\omega$ which is given by

$$\Delta\omega = \omega_2 - \omega_1 \quad \dots (31)$$

Where ω_1 and ω_2 are the frequency corresponding to half of the maximum value of power ($\omega_1 > \omega_2$)

Now,

$$P_{1/2} = \frac{1}{4} V_{max}^2 / R$$

To determine ω_1 and ω_2 , we write

$$\frac{1}{4} \frac{V^2_{max}}{R} = \frac{V^2_{max} R}{2 \left[R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right]}$$

$$2R^2 = R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2$$

$$R^2 = \left(\omega L - \frac{1}{\omega C} \right)^2$$

$$\omega L - \frac{1}{\omega C} = \pm R$$

$$R = \omega L - \frac{1}{\omega C} \rightarrow a ; -R = \omega L - \frac{1}{\omega C} b$$

From (a)

$$LC\omega^2 - 1 = RC\omega$$

$$LC\omega^2 - RC\omega - 1 = 0$$

$$\omega = \frac{RC \pm \sqrt{R^2 C^2 + 4LC}}{2LC} ;$$

$$\omega = \frac{RC + \sqrt{R^2 C^2 + 4LC}}{2LC}$$

From (b) i.e. $-R = \omega L - \frac{1}{\omega C}$

$$\omega = \frac{-RC \pm \sqrt{R^2 C^2 + 4LC}}{2LC}$$

$$= \frac{-RC \pm \sqrt{R^2 C^2 + 4LC}}{2LC} ;$$

$$\omega = \frac{-RC \pm \sqrt{R^2 C^2 + 4LC}}{2LC}$$

The ω values have been chosen as negative results are inadmissible. Out of these two values the greater one corresponds to ω_2 and the lesser to ω_1 .

Band width

$$\Delta\omega = \omega_2 - \omega_1 = \frac{2RC}{2LC} = \frac{R}{L}$$

..... (33)

The geometric mean of ω_1 and ω_2

$$\sqrt{\omega_1 \omega_2} = \sqrt{\frac{R^2 C^2 + 4LC - R^2 C^2}{4L^2 C^2}} = \frac{1}{\sqrt{LC}} = \omega_0 =$$

The resonant frequency. (34)

The sharpness of resonance is measured by the quality factor , Q

$\Delta\omega$ is in the denominator, because less the band width, higher is the peak

Conclusion

Here we have provided four examples of obtaining extremely crucial and meaningful results from different areas of physics, using Taylor's

$$\therefore Q = \frac{\omega_0 L}{R} = \frac{1}{R} \sqrt{\frac{L}{C}}$$

Hence the sharpness of resonance increases with decreasing R. This result can be obtained much more elegantly using Taylor's series.

According to Taylor's theorem

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \dots$$

We consider

$$f(\omega) = \left(\omega L - \frac{1}{\omega C} \right)$$

Taking the increment of ' ω ' by half the band width about $\omega = \omega_0$, we get

$$f\left(\omega_0 + \frac{\Delta\omega}{2}\right) = f(\omega_0) + \frac{\Delta\omega}{2} \left(L + \frac{1}{C\omega_0^2} \right) + \dots$$

$$= 0 + \frac{\Delta\omega}{2} \cdot 2L$$

$$= L\Delta\omega$$

Now, going back to Eq (32), we get

$$\frac{1}{4} \frac{V^2_{max}}{R} = \frac{V^2_{max} R}{2[R^2 + L^2 \Delta\omega^2]}$$

$$2R^2 = R^2 + L\Delta\omega^2$$

$$R^2 = (L\Delta\omega)^2$$

$$R = \pm L\Delta\omega$$

$$\therefore |\Delta\omega| = \frac{R}{L} \text{ which is same as Eq. (33)}$$

The rest can now follow.

series expansion (TSE). As a matter of fact, such examples can be many. It finds application in derivation of Vector Integral Theorems, Adiabatic Approximation, Principle of Least Action,

Relativistic variation of Mass with Velocity and so on.

In all the examples considered here, the functions taken were single-valued. Application of TSE can be extended to multiple variables, which finds application in many vital problems of physics.

Thus an infinite series which has seen the light of the day more than three centuries ago has been contributing significantly towards yielding finite results; one of the rider in its application being the judicious choice of the order where the series is to be truncated.

We conclude by observing that learners of physics generally have a fear psychosis about mathematical derivation, the main cause behind it being their examination phobia. They are to be encouraged to appreciate the vital link between

mathematics and physics. For example, they can be told that the Taylor series is tailor-made for the introduction of Harmonic Approximation. It is indeed a testimony about mathematics playing a guiding role for a principle of physics.

After going through this article we shall expect the undergraduate students to have a paradigm shift in their approach towards mathematical derivations which would be oriented towards viewing mathematics as a foundation for explaining applications in physics.

Acknowledgement

The authors wish to

acknowledge the support received from Shri Anil Saroj in preparing the manuscript.

References

1. Dennis G. Zill, Advanced Engineering Mathematics, Third edition, Jones and Bartlett Publishers.
2. Daniel Kleppner and Robert Kolenkow, An Introduction to Mechanics, McGraw Hill Education.
3. David J. Griffiths, Introduction to Electrodynamics, PHI Learning Pvt. Ltd.

Revisiting the Physics of Turning Water Into Wine, A Pedagogical Approach

A.C. F. Santos

Instituto de Física
Universidade Federal do Rio de Janeiro
21941-972. Rio de Janeiro, RJ, Brazil (toni@if.ufrj.br)

(Submitted : 12-12-2016)

Abstract

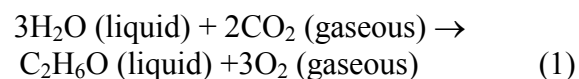
The changing of water into wine at the Marriage at Cana is the earliest miracle credited to Jesus in the Gospel of John. In this paper, I calculate the change in the chemical potential across the reaction first suggested by Irwin, Hicks, and Lerman. Using the simple concepts from statistical physics is not only easier to get the message, but even High-school students can develop a deeper understanding of the process.

1. Introduction

In the Gospel of John (2 : 1-11) [1], Jesus, Maria and his disciples were welcomed to a wedding at Cana. When the wine was over, Maria asked Jesus to endorse his glory by turning water into wine. The main constituents of wine are: water, carbohydrates, acids, alcohols, phenolics, nitrogenous compounds, and inorganic substances. Table I shows the typical concentration ranges of the major chemical components of dry table wine. To provide somewhat larger view of the thermodynamic considerations in the making of wine, the reader is referred to reference [2].

Irwin, Hicks, and Lerman[3] investigated the change in enthalpy across their proposed reaction and determined it to be 1255 kJmol^{-1} , indicating an endothermic reaction. They supposed that pure water was changed into a basic form of water mixed with ethanol, with the strength of 12% ABV (Alcohol by volume). They found the change in entropy to be $4.21 \text{ kJmol}^{-1}\text{K}^{-1}$ and estimated the energy required for the entire reaction occur as 0.25-

0.37GJ. The turning from water (H_2O) to ethanol ($\text{C}_2\text{H}_6\text{O}$) requires a source of carbon. Irwin, Hicks, and Lerman assumed that the most probable source would be carbon dioxide in the air. Then, they set up the reaction which conserves the number of each type of atom.



This reaction was supposed to occur under standard pressure (1 atm) and temperature (298 K). To inspect whether this reaction would occur spontaneously, the authors calculated the total change in bond enthalpy. If it were negative, more energy would be liberated from forming the new bonds than it was required to initially break the original bonds. Consequently, the reaction will occur spontaneously. On the other hand, if it were positive, the reaction would be endothermic and would require a heat input. The authors also assumed that Jesus was able to provide the perfect

catalyst (the agent that modifies the transition state to lower the activation energy) to the reaction by making the additional activation energy negligible.

In this paper, I suggest an alternative and more pedagogical way inspect this reaction. The tendency of water to react with CO_2 via Eq.1 can be expressed quantitatively by means of some simple ideas that are the building blocks of statistical physics. Supplementing the orthodox thermodynamics lecture with these ideas can improve understanding and enthusiasm to introductory physics.

Constituent	Wine (g/L)
Water	800 – 900
Carbohydrates	1 – 10
Glucose	0.5 – 5
Fructose	0.5 – 5
Pectins	Trace
Acids	4.5 – 11
Tartaric	1 – 6
Malic	0 – 8
Lactic	1 – 5
Acetic	0.2 - 1.5
Alcohols	
Ethanol	80 – 150
Glycerol	3 – 14
Phenolics	From trace – 5
Simple phenolics	Trace - 0.2
Anthocyanins	0 - 0.5
Tannins	T – 5
Nitrogenous compounds	0.1 – 1
Inorganic constituents	1.5 – 4
Potassium	0.5 – 2

Table 1 - typical concentration ranges of the major chemical components of grape juice and dry table wine.

2. Introducing the Gibbs free energy

This section introduces the Gibbs free energy to justify the use of the chemical potential,

μ . For a less formal approach, the teacher may skip this part without lack of content.

Processes in biology and chemistry often proceed under conditions of constant temperature T and constant (external) pressure P . For a given energy input energy by heating Q , and volume change ΔV , What does the basic entropy inequality,

$$\Delta S \geq \frac{Q}{T} = \frac{\Delta E + P\Delta V}{T}, \quad (2)$$

imply under those circumstances ?

The temperature T denotes the temperature of the environment. If that temperature remains constant during whatever process is being described, then we may multiply on both sides by T and may include T within the ΔS term:

$$\Delta(TS) \geq \Delta E + P\Delta V. \quad (3)$$

Similarly, the work done in expansion depends on the force put that opposes the expansion; so the pressure P may be taken to be an external pressure. If that pressure remains constant, then P may included within the ΔV term:

$$\Delta(TS) \geq \Delta E + \Delta(PV). \quad (4)$$

Finally, we collect all terms onto the right-hand side:

$$0 \geq \Delta(E - TS + PV). \quad (5)$$

Reading from right to left, we find that the combination $E - TS + PV$ must decrease or remain constant. Given the inequality in equation (5), the combination $E - TS + PV$ warrants a name:

$$\text{Gibbs free energy} = G \equiv E - TS + PV. \quad (6)$$

3. Water into Wine and chemical potential

The chemical potential, μ , of a substance characterizes the tendency of the decomposition

reaction of this substance into the elements in their final states [4]. The calculation of chemical reactions begins with the presentation of the chemical potentials of the reactants and products. The tendency for the Gibbs free energy to decrease under conditions of constant temperature and constant (external) pressure is the key concept. The chemical potential provides a way to use the Gibbs free energy efficiently. This approach is elementary, as it does not demand previous knowledge and leads straightforward to results. The line of attack is through the potential differences between the reactants and products of reaction represented in Eq. 1.

The entropy change, ΔS equals the amount of energy transferred by heating to a system divided by the temperature at which the process takes place, provided the process proceeds slowly. According to the second law of thermodynamics, nature has a tendency for entropy to increase in an isolated system, and the system changes in answer to this propensity. Therefore, alike to the chemical potential, the increase in entropy is the driving force for a spontaneous process in an isolated system.

To each substance, a chemical potential and a standard entropy can be associated. Table II presents the chemical potentials of the substances in Eq. 1.

Substance	Phase	μ (kG)	S(kJ.mol ⁻¹ .K ⁻¹)
H ₂ O	Liquid	-237,2	69.9
CO ₂	Gas	-394,4	213.6
C ₂ H ₆ O	Liquid	- 174,7	283.59
O ₂	Gas	0	205

Table II – chemical potentials (in kiloGibbs, 1 kG = 1 kJ.mol⁻¹) and standard entropies at T = 298.15 K and P = 101325 Pa.

First, let me define the chemical potential of the reactants as $\mu_R = \mu_{H_2O} + 2\mu_{CO_2}$, and the corresponding chemical potential of the products a $\mu_P = \mu_{C_2H_6O} + 3\mu_{O_2}$. So, as to inspect whether the reaction takes places freely or not, we weight μ_R

against μ_P . If $\mu_R > \mu_P$, then the transformation of water into wine takes place spontaneously. If $\mu_R = \mu_P$, then there is a chemical equilibrium. Finally, if $\mu_R < \mu_P$, then the transformation of water into wine does not occur spontaneously. From Table I, we obtain

$$\mu_R = 3\mu_{H_2O} + 2\mu_{CO_2} = -1500.4 \text{ kG}$$

and

$$\mu_P = \mu_{C_2H_6O} + 3\mu_{O_2} = -174.7 \text{ kG}$$

Thus, we have $\mu_P > \mu_R$. This result was to be expected since it is well known that water does not transform spontaneously into wine.

Analogically, water could be converted into wine spontaneously (under conditions of constant temperature and pressure) only if the Gibbs free energy would decrease. The change of the Gibbs free energy when a molecular constituent is changed by one mole (or, in some circumstances, by one molecule) is called chemical potential for that constituent and is denoted by the Greek letter μ . Thus, the change in the Gibbs free energy when the reaction (1) is imagined to proceed by one step may be written as

$$\Delta G = (\mu_{C_2H_6O} + 3\mu_{O_2}) - (3\mu_{H_2O} + 2\mu_{CO_2})$$

Tabulated data (see Table II) show that the right-hand side has the numerical value

$$\Delta G = (-174.7) - (-1500.4) = +1325.7 \text{ kJ.mol}^{-1}$$

The positive value for the imagined reaction indicates that the reaction will not proceed spontaneously.

The entropy change can be calculated in the same way. From Table II:

$$\Delta S = (S_{C_2H_6O} + 3S_{O_2}) - (3S_{H_2O} + 2S_{CO_2}) = 261.89 \text{ kJ.mol}^{-1}\text{K}^{-1}$$

Since $\Delta S > 0$, the reaction is physically allowed and irreversible. In fact, the second law of thermodynamics states that the overall entropy of an isolated system always increases, or stays constant in reversible process. $\Delta S > 0$ accounts for the irreversibility of natural processes. The second law of thermodynamics also states that for any spontaneous process, the overall ΔS must be greater than or equal to zero, which it is not the case for Eq. 1, since, as we saw, $\Delta G > 0$. It is worth to note that spontaneous reactions can result in a negative change in entropy. This fact does not refute the second law of thermodynamics, though. The increase in temperature of the surroundings gives rise to an appropriately large increase in entropy, such that the total change in entropy is still positive. In other words, the ΔS of the surroundings increases

References :

- [1] Gospel of John (2 : 1-11).
- [2] E.Covaci, *Thermodynamic Parameters Of Potassium Bitartrate During The Young Wines Cold Stabilization*, Chemistry Journal of Moldova. General, Industrial and Ecological Chemistry **10**, 42-45(2015)
- [3] B. Irwin, P. Hicks, H. Lerman, *Water into Wine*, Journal of Physics Special Topics P5_11 (2013).
- [4] G. Job and F. Herrmann, *Chemical potential - a quantity in search of recognition*, Eur. J. Phys. **27**, 353 (2006).

sufficiently due to the exothermicity of the reaction so that it overcompensates for the negative ΔS of the system. Thus, the overall change in entropy is still positive.

In order to get a quick estimate for the probability of reaction (1) takes place, using $k_B = 1.38 \times 10^{-23}$ J/K (for room temperature), where k_B is Boltzmann constant, one obtains $\frac{P_{wine}}{P_{water}} = e^{-\frac{\Delta S}{k_B}}$ [5-8], which is so small (my calculator even refused to calculate it). Then, the extract from John persists a miracle.

Conclusions

The ideas discussed in this paper can be introduced smoothly into high-school or introductory college physics with least work. I believe that the contextualization presented here is interesting, and exemplifies the power of the statistical physics.

- [5] R. Baierlein, *The elusive chemical potential*, Am. J. Phys. **69**, 423 (2001).
- [6] M. D'Anna and P. Lubini, *Chemical potential*, The Physics Teacher **48**, 358 (2010).
- [7] G. Cook and R. H. Dickerson, *Understanding the chemical potential*, Am. J. of Phys. **63**, 737 (1995).
- [8] J. J. Prentis, *Thank you, Boltzmann, your constant is so small*, The Physics Teacher **34**, 392 (1996).

The paradox of power loss in a lossless infinite transmission line

Ashok K. Singal

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

(Submitted 14-11-2015)

Abstract

We discuss here the famous paradox of a continuous power drainage from the source at the input of an otherwise lossless infinite transmission line. The solution of the paradox lies in the realization that in an open-circuit finite transmission line/ladder network, there is an incident as well as a reflected wave and the input impedance is determined by the superposition of both waves. It is explicitly shown that the reactive input impedance of even a single block, comprising say a simple LC circuit, is determined at all driving frequencies from the superposition of incident and reflected waves, and that the input impedance remains reactive in nature (i.e., an imaginary value) even when additional blocks are added indefinitely. However in a ladder network or transmission line, taken to be *infinite right from the beginning*, there is no reflected wave (assuming the circuit to be ideal with no discontinuities en route). Thus the source while continuously supplying power in the forward direction, does not retrieve it from a reflected wave and unlike in the case of a finite line, there is a net power loss. This apparently lost energy ultimately appears in the electromagnetic fields in the reactive elements (capacitances and inductances which to begin with had no such stored energy), further down the line as the incident wave advances forward. It is also shown that radiation plays absolutely no role in resolving this intriguing paradox.

Introduction

A transmission line is a channel for transmitting electric signals or power from one point to another along a guided path [1, 2, 3, 4]. A line could be of a finite length or be of infinite length (at least in principle). A circuit comprising lumped parameters is generally called a ladder network, on the other hand if it consists of a continuous distribution of parameters, then it is usually called a transmission line. The two are almost identical in their behavior [1]. The elements of a transmission line could be either reactances (with no power dissipation within them) like capacitances or inductances, or could comprise resistances or shunt leakage conductances, which dissipate power into heat. Most lines will have a mixture of reactances and dissipative elements. An ideal transmission line may be thought of as the one which delivers signal or power across its length without any dissipation on the way. Intuitively one would think a line devoid of elements like resistances should behave as a lossless line without a continuous power drainage from the source at the input, and this does seem to hold true for a line of finite length. However, for an infinite line, even if there were no resistive elements along its length that could dissipate power, the line presents a *real value* of input impedance, implying that power will be drained from the source at a constant rate [5].

Where does this energy go as it is not dissipated in the inductors and capacitors of the circuit? For this Feynman [4] writes “But how can the circuit continuously absorb en-

ergy, as a resistance does, if it is made only of inductances and capacitances? *Answer:* Because there are an infinite number of inductances and capacitances, so that when a source is connected to the circuit, it supplies energy to the first inductance and capacitance, then to the second, to the third, and so on. In a circuit of this kind, energy is continually absorbed from the generator at a constant rate and flows constantly out into the network, supplying energy which is stored in the inductances and capacitances down the line.”

In an alternative approach [6, 7] it has been shown that the input impedance of an open-circuit ladder network, initially consisting of a finite number of blocks comprising inductors and capacitors, does not converge to a unique fixed value when additional identical blocks are added, and always yields pure imaginary (reactive) input impedance value irrespective of the number of the blocks added. The input impedance does not have a real (dissipative) part for any driving frequency, even when the number of blocks is increased indefinitely. This contradicts Feynman’s observation [4] that the infinite ladder network has an input impedance which has a real part at frequencies below certain value. It was argued afterwards [8] that a non-zero real part of impedance appears only if there is a termination in an impedance that has a real part and that a circuit consisting solely of components with purely imaginary impedances has a purely imaginary input impedance. Later the behavior of infinite ladder network, its convergence and solutions have been analyzed in a greater detail [9, 10].

In this paper we examine this intriguing paradox from a fresh view point trying to understand why two alternate approaches lead to conflicting results. We will first review the relevant characteristics of a transmission line/ladder network; the detailed description of various terms and the derivation of the formulas used can be found in standard textbooks [1, 2, 3, 4]. Then we shall show how one arrives at a paradoxical result of an uninterrupted power drain in an otherwise lossless infinite transmission line. This will be followed by a brief account of the alternative approach of extending a finite ladder network by the addition of further blocks, with the circuit always comprising only reactive elements. Subsequently we shall present the resolution of the paradoxical results both for a ladder network as well as the transmission line; the resolution basically ensues the realization that there is an absence of a reflected wave in an infinite ladder network or a transmission line. That reflection could play a role in the paradox had a brief mention without further elaboration [10], which we do here in detail by explicitly calculating the input impedance of a finite ladder network by a superposition of incident and reflected waves. We shall demonstrate that unlike in a finite case, where a termination in a load matched to the characteristic impedance of the line could dissipate all power, or at an open-ended termination could reflect it all back towards the source, in the case of a line taken to be infinite right from the start, there is no termination point to start a reflected wave (provided of course no discontinuities along the line to trigger any reflection) and that results

in the current being in phase with the voltage and net power being drained from the source.

A non-ideal behavior of an ideal circuit

A transmission line

A transmission line is described by its line parameters R, L, C, G , where R is the series resistance per unit length of line (including both wires), L is the series inductance per unit length of line, C is the capacitance between the two conducting wires per unit length of line and G is the shunt leakage conductance between the two conducting wires per unit length of line. For an incremental length Δz of the line, the equivalent circuit is shown in Fig. 1. The increments in voltage and current along the line are [1, 2, 3, 5],

$$\Delta V(z) = -I(z)(R + j\omega L)\Delta z \quad (1)$$

$$\Delta I(z) = -V(z)(G + j\omega C)\Delta z. \quad (2)$$

These could be written in limit $\Delta z \rightarrow 0$ as,

$$dV(z)/dz = -I(z)(R + j\omega L) \quad (3)$$

$$dI(z)/dz = -V(z)(G + j\omega C). \quad (4)$$

From Eqs. (3) and (4) one gets a general solution for voltage along the line,

$$V(z) = V_0' e^{-\gamma z} + V_0'' e^{\gamma z} \quad (5)$$

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta, \quad (6)$$

where γ is the propagation constant. The phasor part is written with an assumed $e^{j\omega t}$

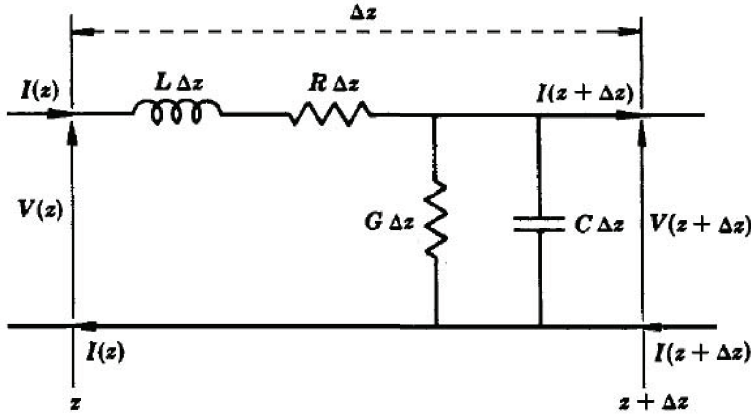


Figure 1: Increments of voltage and current over an incremental length Δz of the transmission line.

time dependence throughout. Now of the two terms in Eq. (5), the first one represents a wave traveling along increasing z commencing at $z = 0$, while the second represents a wave traveling towards decreasing z which in case of an infinite line would have to start from $z = \infty$ an infinite time back and thus must be dropped. Therefore the voltage along an infinite transmission line can be written as,

$$V(z) = V_0 e^{-\gamma z} = V_0 e^{-\alpha z} e^{-j\beta z}. \quad (7)$$

From this one gets for the electric current,

$$I(z) = (V_0/Z_0) e^{-\gamma z} = (V_0/Z_0) e^{-\alpha z} e^{-j\beta z}. \quad (8)$$

Here Z_0 , the characteristic impedance of the line given by,

$$Z_0 = \sqrt{(R + j\omega L) / (G + j\omega C)}. \quad (9)$$

Equations (7) and (8) represent an attenuated sinusoidal wave along z , with α as the

attenuation constant and $\beta = 2\pi/\lambda$ as the wave number.

For an infinite line, the input impedance (at $z = 0$) is calculated from Eqs. (7) and (8) as,

$$Z_i = V(0)/I(0) = Z_0. \quad (10)$$

In a lossless line, $R = 0$ and $G = 0$, and from Eq. (6) we have, $\alpha = 0$ and $\beta = \omega\sqrt{LC}$, i.e., a sinusoidal wave without any attenuation along the line. But we also have $Z_i = Z_0 = \sqrt{L/C}$, i.e., its impedance has a *real* value. This is a paradox because though the transmission line contains no resistive element so there could be no Ohmic losses in the line, yet its input impedance is a pure resistance. That means for an input voltage V_0 , power will be drained from the source at the rate of $V_0^2/(2\sqrt{L/C})$ [5]. The questions therefore arise as to why does a pure resistance show up in a circuit comprising

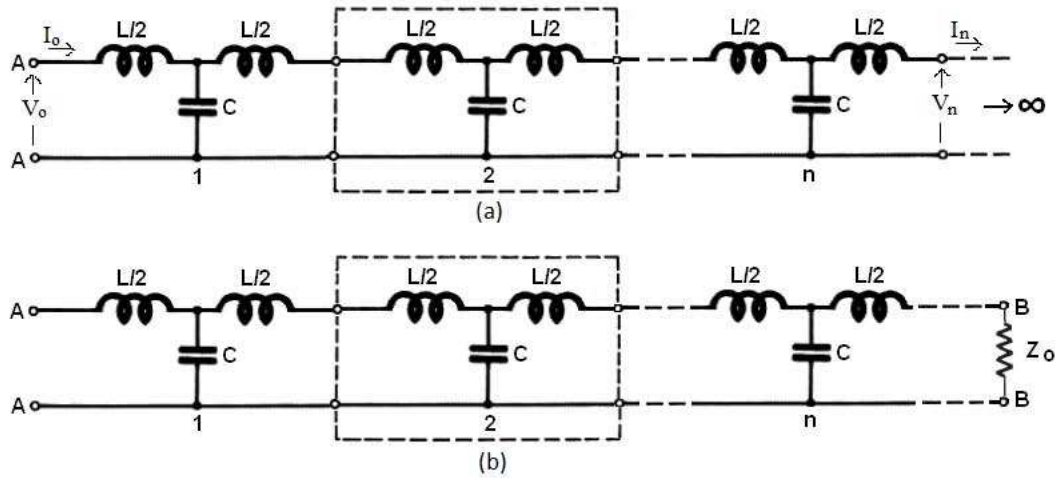


Figure 2: (a) An infinite ladder network comprising lumped parameters. (b) A finite ladder network terminated in its characteristic impedance Z_0 . A typical block in the network is shown by a rectangular box of dashed lines.

only reactances, thereby implying a continuous power drainage and where does this energy ultimately go?

The paradox can be also seen from the Smith chart where the input impedance of a lossless open-circuit line, goes through cycles when its length is varied. Not only does the input impedance not converge to a single unique value when the length of the line is increased indefinitely but also in general it is an *imaginary* value, i.e., a pure reactance [1, 2, 3] for any length of the line, which contradicts the conclusion that the infinite line presents a *real* input impedance.

A ladder network

A transmission line with distributed parameters is almost identical in behavior to a ladder network comprising lumped parameters

[1], and the above paradox appears in the infinite ladder network too. A Ladder network of n blocks, with each block a symmetrical T section consisting of two $L/2$ inductances and a capacitance C , has a characteristic impedance $Z_0 = \sqrt{L/C - \omega^2 L^2/4}$ [1, 2, 3]. The number n of blocks could be finite, or it could even be infinite ($n \rightarrow \infty$). Figure 2(a) shows an infinite ladder network while Fig. 2(b) shows a finite ladder network, but terminated in its characteristic impedance $Z_0 = \sqrt{L/C - \omega^2 L^2/4}$.

A solution for the input impedance Z_i of the infinite network is obtained in the following manner [4, 6, 7, 11, 12]. Since adding another block to the beginning of an infinite ladder network does not change the input impedance (it still remains the same infinite network), Z_i must equal the impedance of a

circuit having a single block terminated in a load impedance equal to Z_i . Therefore we have,

$$Z_i = \frac{j\omega L}{2} + \frac{(Z_i + j\omega L/2)(1/j\omega C)}{Z_i + j\omega L/2 + 1/j\omega C}, \quad (11)$$

which has a solution,

$$Z_i = \sqrt{L/C - \omega^2 L^2/4}. \quad (12)$$

The input impedance of the infinite network equals its characteristic impedance, i.e., $Z_i = Z_0$, and the circuit behaves as if it were terminated in Z_0 somewhere along the line as in Fig. 2(b). Now for $\omega < \omega_0 = 2/\sqrt{LC}$, Z_i is a *real* value. This leads to the same paradox as for the infinite transmission line of distributed parameters – how come a circuit containing only purely imaginary impedances has for its input impedance a *real* value which could absorb energy continuously?

Where does the energy disappear? – Could radiation losses be the answer?

Could the energy be lost into the surrounding medium by the process of radiation, with $Z_0 = \sqrt{L/C}$ as the radiation resistance? In transmission line or ladder network containing resistive elements, power loss by the source is fully accounted for by the energy dissipation in the circuit, for any value of R and G .

Consider the lossy infinite line (i.e., with R and G non-zero), where input power from the source is [5],

$$\begin{aligned} P_i &= [V_0^2/(2|Z_0|)] \cos(\angle Z_0) \\ &= [V_0^2/(2|Z_0|^2)] \text{Re}(Z_0). \end{aligned} \quad (13)$$

On the other hand the power dissipated in an infinitesimal line element (Fig. 1) is,

$$\begin{aligned} dP_d &= (1/2) (|I(z)|^2 R + |V(z)|^2 G) dz \\ &= (V_0^2/2) [(R/|Z_0|^2) + G] e^{-2\alpha z} dz. \end{aligned} \quad (14)$$

Hence the total power dissipated in the infinite line is

$$\begin{aligned} P_d &= (V_0^2/2) [(R/|Z_0|^2) + G] \int_0^\infty e^{-2\alpha z} dz \\ &= (V_0^2/2) [(R/|Z_0|^2) + G] / (2\alpha). \end{aligned} \quad (15)$$

Substitution for $|Z_0|$ and α shows that $P_d = P_i$ [5], and all power losses are accounted for without anything going into radiation. This is true for all R and G , in particular even when in limit $R \rightarrow 0$ and $G \rightarrow 0$. Now it cannot happen that when $R = 0$ and $G = 0$ radiation suddenly shows up into picture from somewhere. Further, even in a lossless line, all the power (assumed to be lost by the source) can at any stage be either reflected back by making the circuit open just after that point, or it could be consumed by terminating the line in its characteristic impedance, irrespective of the length of the line up to that stage. This implies that up to any *arbitrarily selected length* of the line, the radiation losses had not yet taken place. Therefore for resolving this paradox there does not seem any scope for radiation hypothesis at all and a satisfactory resolution of the paradox lies elsewhere.

The paradox reappears!

Actually while writing Eq. (11) for Z_i [4], one implicitly assumed that the infinite series *converges* to a unique value and it is only under this existence supposition that a unique solution Eq. (12) could be obtained. If the series does not converge, then of course this basic assumption itself breaks down and the solution obtained thereby may not represent a true value.

On the other hand if one started with an open-circuit ladder network of a finite number of identical blocks comprising inductors and capacitors, and then added more similar blocks, the input impedance does not converge to a unique fixed value even when the number of blocks is increased indefinitely [6, 7]. Moreover, the input impedance always turns out to be a pure imaginary value with no real (dissipative) part for any driving frequency, even when the number of blocks approaches infinity.

It seems that the infinite ladder networks of type in Fig 2(a) may have different answers for the input impedance, and thereby implying different power consumptions depending upon the method of solution. Hence a paradox still exists as one arrives at different answers using different arguments, and a question still remains whether or not does an infinite ladder network converge to a pure resistance drawing continuous power from an input source, and if so where does this energy go. What could be the missing factor, if any, in these arguments?

Resolution of the paradox – incident versus reflected waves

Here we demonstrate with a detailed analysis that the resolution of the paradox lies in the realization that there is an absence of a reflected wave in an infinite ladder network or an infinite transmission line. That should also help us comprehend why the two alternate approaches led to two conflicting conclusions.

The case of an infinite ladder network

Let us examine the propagation factor $e^{-\gamma}$ between successive blocks in a ladder network. To do this we terminate the ladder network in its characteristic impedance Z_0 , which is basically to ensure that there is no reflected wave and thus we are dealing only with the incident wave. The propagation factor in this way is found to be [1, 4],

$$e^{-\gamma} = \frac{V'_n}{V'_{n-1}} = \frac{I'_n}{I'_{n-1}} = \frac{\sqrt{L/C - \omega^2 L^2/4} - j\omega L/2}{\sqrt{L/C - \omega^2 L^2/4} + j\omega L/2}, \quad (16)$$

which can be simplified to give,

$$e^{-\gamma} = 1 - \omega^2 LC/2 - j\sqrt{\omega^2 LC} \sqrt{1 - \omega^2 LC/4}. \quad (17)$$

A prime (') over voltages and currents merely indicates that these represent an incident wave.

The ladder network at low frequencies

For frequencies below a critical value $\omega_0 = 2/\sqrt{LC}$, the characteristic impedance $Z_0 = \sqrt{L/C - \omega^2 L^2/4}$ can be written as $Z_0 = \sqrt{L/C} \sqrt{1 - (\omega/\omega_0)^2}$, which is a *real* quantity, meaning a pure resistance. In the low frequency ($\omega < \omega_0$) case the propagation factor from Eq. (17) can be written as,

$$e^{-\gamma} = 1 - 2(\omega/\omega_0)^2 - j2(\omega/\omega_0)\sqrt{1 - (\omega/\omega_0)^2}. \quad (18)$$

From the real and imaginary parts in Eq. (18), it can be readily seen that the propagation factor has a unit magnitude and represents a simple phase change $e^{-j\beta} = \cos \beta - j \sin \beta$, between successive blocks in the network.

Although for calculating the propagation factor of the circuit we needed to isolate the *incident wave* by terminating this network with its characteristic impedance Z_0 , yet the propagation properties of the incident wave (that is, the propagation constant calculated from Eq. (18) of *incident wave* between two neighboring blocks, say, $n - 1$ and n) does not depend upon this termination. The incident wave has an input impedance everywhere equal to the characteristic impedance Z_0 of the network. Of course the voltages and currents at any point are decided by the superposition of the incident and reflected waves at that point and the input impedance of a network calculated using the standard procedure [6, 9, 10, 11] is actually what results from the superposition of the incident and the reflected wave with their phases duly taken into account.

To prove our assertion that this indeed is the case in general, we want to calculate input impedance of an open-circuit line, made of any finite number of blocks (say, n), by evaluating voltage and current at $z = 0$ due to the sum of the incident and reflected waves, the latter arising from the termination just after the n th block. For a cascaded network of n identical blocks, the propagation factor is simply $e^{-jn\beta}$. The angle β here is half of θ defined in Eq. (21) of that in Ref. [9]. If the network has a total of n blocks, then voltage V_0 at $z = 0$ includes a reflected wave with a phase change of angle $2n\beta$ from the incident wave, while the current I_0 has a phase change of angle $2n\beta + \pi$ (an extra phase of angle π in the current wave at the reflection point). Therefore the input impedance is given by,

$$\begin{aligned} Z_i = V_0/I_0 &= Z_0 \frac{1 + e^{-j2n\beta}}{1 - e^{-j2n\beta}} \\ &= -jZ_0 \cot(n\beta). \end{aligned} \quad (19)$$

We see that the calculated input impedance is the same what was calculated in an alternative method for a finite open-circuit ladder network [9, 10], which thus proves our assertion that the propagation factor of the incident wave is unaffected by the termination impedance. As n increases, Z_i ever remains an *imaginary* value that goes through cycles, even becoming a 0 or an ∞ , and in general not converging to a unique value even when $n \rightarrow \infty$.

Energy transport - a physical perspective

In a finite open-circuit network there is a reflected wave from its terminated end as it has to match the conditions for a zero net current (implying the electric currents out of phase by angle π for the incident and the reflected waves), although the voltages will be in phase for the incident and reflected waves at the termination point. It is important to note that when we analyze a *finite* network, barring transients, the voltages and currents being considered are the superposition of incident and reflected waves. Therefore the calculated Z_i may depend upon the length of the line or equivalently the number of blocks in the network as that would determine the relative phases of the incident and reflected waves at the input point.

Suppose a generator is connected to the circuit at input terminals AA (Fig. 2(a)). The generator drives the circuit at a frequency ω (say) and will give rise to a voltage as well as a current in the 1st electric block, which (a pure reactance) does not consume the electric power itself, and in turn gives rise to voltage and current in the 2nd block and so on. As we showed above, for $\omega < \omega_0$, there will be no decrease in the amplitude of voltage or current from one block to the next and there will only be a progressive phase change between successive blocks. This “incident wave” will move along the network until a discontinuity, say an open-circuit termination after n th block, is encountered which will cause a reflected wave towards the block $n - 1$, then $n - 2$ and so on continuing onto the genera-

tor.

The generator meanwhile still keeps on supplying further power as incident wave to the 1st block which gets passed further on, until it is finally reflected back towards the generator. This is true even when the line terminates just after the very first block (just a simple LC circuit, see Appendix). And when we have two or more but a finite number of blocks, then the discussion still entails reflected waves implicitly. However when we consider an infinite ladder network or an infinite transmission line, all by itself (and not by an indefinite extension of finite network by adding more successive blocks or increasing the length of the line), then we do not consider the reflected wave since the incident wave does not ever reach a termination point to start a reflected wave.

In that case we have only the incident wave and the source at the input keeps on continuously supplying power to the network or the transmission line but does not get it back as a reflected wave. Therefore in an infinite network or transmission line, it results in a net power drain from the source and this energy of course appears from one block to the next down the line where it has not yet reached due to the long extent of the line. Of course as it will never reach a termination point (at infinity!), so the energy transfer further down the line continues for ever. Initially none of the circuit elements had electric energy (say just before time $t = 0$ when the generator was just connected), but afterwards up to a certain stage the circuit elements have stored electromagnetic energy (shared between the capacitances and induc-

tances or equivalently between the electric and magnetic fields and getting continuously exchanged between them). Ultimately this energy has come from the generator. The energy is not lost as it can be still consumed by terminating the circuit in a matched load somewhere still further down the line or recovered by terminating the line as an open circuit and getting the energy returned as a reflected wave.

The ladder network at high frequencies

The propagation factor $e^{-\gamma}$ between adjacent blocks, for a high frequency ($\omega > \omega_0$) case can be written from Eq. (17) as,

$$e^{-\gamma} = 1 - 2(\omega/\omega_0)^2 + 2(\omega/\omega_0)^2 \sqrt{1 - (\omega_0/\omega)^2}. \quad (20)$$

From Eq. (20) it can be seen that for $\omega > \omega_0$ the propagation factor, written as $e^{-(\alpha+j\pi)} = -(\cosh \alpha - \sinh \alpha)$, is of magnitude less than unity and is always of a negative value, implying a phase change of angle π between successive blocks accompanied by an exponential decrease in amplitude. The voltages and currents do not penetrate too far in the circuit, and there is no continuous transport of energy along z . The input impedance at frequencies $\omega > \omega_0$ for a cascade network of n blocks is,

$$Z_i = V_0/I_0 = Z_0 \frac{1 + e^{-2n\alpha}}{1 - e^{-2n\alpha}} = Z_0 \coth(n\alpha), \quad (21)$$

which is imaginary, in spite of $\coth(n\alpha)$ being always a real value. This is be-

cause the characteristic impedance $Z_0 = (j\omega L/2)\sqrt{1 - (\omega_0/\omega)^2}$ is imaginary for $\omega > \omega_0$. For $n \rightarrow \infty$, $Z_i \rightarrow Z_0$, a pure reactance, thus there is no paradox for the $\omega > \omega_0$ case.

While the incident wave, when considered alone, presents a *real* input impedance for frequencies $\omega < \omega_0 = 2/\sqrt{LC}$, whenever a superposition of incident and reflected waves is considered then we get an *imaginary* input impedance for all driving frequencies, implying in turn the current at the input being $\pi/2$ out of phase with the voltage and as a result no continuous power absorption from the source.

Infinite transmission line

The characteristic impedance of a ladder network in 2.2 could be rewritten as $Z_0 = \sqrt{(L\Delta z)/(C\Delta z) - \omega^2 L^2(\Delta z)^2/4}$ in case of a lossless ideal transmission line with distributed parameters (Fig. 1) and which reduces to $Z_0 = \sqrt{L/C}$ as in limit $\Delta z \rightarrow 0$. Therefore unlike the ladder network case, in the transmission line case there is no cut-off frequency and for *all* driving frequencies a wave travels along the line without any amplitude attenuation since propagation constant has an *imaginary* value implying only a phase change.

In general the input impedance of a line of length l is given by [1],

$$Z_i = Z_0 \left(\frac{e^{\gamma l} + K e^{-\gamma l}}{e^{\gamma l} - K e^{-\gamma l}} \right), \quad (22)$$

where K (=reflected voltage at load/incident

voltage at load) is the reflection coefficient, impedance becomes,

$$K = \left(\frac{Z_r - Z_0}{Z_r + Z_0} \right). \quad (23)$$

$$Z_i = Z_0 \left(\frac{e^{j\beta l} + e^{-j\beta l}}{e^{j\beta l} - e^{-j\beta l}} \right) = -jZ_0 \cot(\beta l)$$

$$= -jZ_0 \cot(2\pi l/\lambda), \quad (25)$$

with Z_r as the impedance at the receiving (load) end. The input impedance reduces to Z_0 when there is no reflected wave, i.e., when $K = 0$.

Now the absence of a reflected wave in a transmission line can be due to three reasons. First, the line is finite but terminates in a load matched to the characteristic impedance of the line, i.e., when $Z_r = Z_0$. Second, the line has small resistance which causes the incident voltage to attenuate over its long length l , i.e., if $\gamma l \rightarrow \infty$, so that the amplitude of the incident and thence of the reflected wave in limit is zero, and then the series does converge to a unique solution [6, 11] which is consistent with $Z_i = Z_0$. Thirdly the line is lossless but truly of infinite extent so that it could be assumed that the incident wave, which presumably started a finite time back, never reaches a termination point to start a reflected wave. In all three cases, the input impedance, which is the ratio of the voltage and current at the input point, is the same as that is not affected by what happens at its termination point, and we obtain the same result for the input impedance, viz. $Z_i = Z_0$.

On the other hand, for an open-circuit line of finite length l ($Z_r = \infty$, $K = 1$), the input impedance is given by,

$$Z_i = Z_0 \left(\frac{e^{\gamma l} + e^{-\gamma l}}{e^{\gamma l} - e^{-\gamma l}} \right). \quad (24)$$

In a lossless line, $\gamma = j\beta = j\omega\sqrt{LC}$, the input

impedance becomes, which is a pure reactance, and thereby no net power consumed, and which is similar to the result derived for the ladder network (Eq. (19)). It should be noted that in case of a ladder network, the quantities γ, α, β , or even L, C etc. are specified as per block of the circuit while in the case of a transmission line with distributed parameters all such quantities are defined per unit length of the line. Therefore in Eq. (19) it is the phase angle change $n\beta$ over n blocks while in Eq. (25) it is the phase angle change βl over length l of the line. In fact with increasing l , Z_i/Z_0 from Eq. (25) is cyclic and is indeed the value read from the Smith chart. One thing that we notice from Eq. (25) is that the input impedance Z_i depends on the length l of the line in terms of wavelength λ . Thus depending upon $2\pi l/\lambda$, Z_i could be zero, a finite value or even infinity, but always a pure *imaginary* value, with a zero *real* part similar to what was seen for the ladder network in 5.1.1. Here as much amount of power is reflected back to the generator as much it supplies in the incident wave.

In the case where there is only an incident wave, i.e., there is no reflected wave, the current is in fact in phase with the voltage, implying power is being drawn from the source. However, if there is a reflected wave as well, then the voltage and current are not in phase everywhere. Thus it is the absence

of reflected wave in infinite transmission line that results in a continuous positive energy flux along the line. The relative phases of V and I depend upon the reflected wave, which in turn depends upon at how far away along the line reflection took place. Of course no reflection will ever take place in a uniform infinite line as the incident wave will never reach the termination point which is at infinity. However if we consider the lossless case when there is a reflected wave from an open-circuit termination, then equal power is being returned to the source by the reflected wave and in that case the current is indeed $\pi/2$ out of phase with the voltage (Eqs. (19) and (25)).

If we consider a transmission line with no discontinuities whatsoever, then it will have to be an infinite line and the energy will be getting stored as electric and magnetic fields in its reactive elements further and further along the line. There is no violation of the energy conservation, and since there is no reflected wave to restore energy to the source, the latter would be continuously supplying energy, which gets stored in electric and magnetic fields in more and more inductances and capacitances down the line. Seen this way there does not seem to be any paradox.

The paradox actually had arisen only because we were comparing two sets of solutions which are for quite different situations. One involves only an incident wave (i.e., without any reflected wave) and then the input impedance $Z_i = Z_0$ is a *real* quantity, and the voltages and currents are in phase everywhere along the circuit, with energy getting apparently “spent” as it is getting stored in

the inductors and capacitors down the line as the incident keeps on advancing for ever in an infinite transmission line. The other solution was for the case with a reflected wave, and there the superposition of the incident and reflected waves results in Z_i to have *imaginary* value with no *net* power loss since the source gets the energy back as the reflected wave.

Conclusions

It was shown that while an open-circuit finite ladder network or a transmission line with distributed network has a characteristic impedance Z_0 which is only reactive (imaginary), an infinite ladder network or an infinite transmission line has a finite *real* component of the input impedance. It was shown that the famous paradox of power loss in a lossless infinite transmission line is successfully resolved when one takes into account both the incident and reflected waves. The solution of the paradox lies in the realization that there is an absence of a reflected wave in an infinite transmission line. In a finite transmission line or ladder network, the source still keeps on supplying power as an incident wave but gets it equally back in terms of the reflected wave. Therefore there is no further net power transfer from the source which is consistent with the reactive elements presenting zero net resistance.

However in the case of an infinite ladder network or an infinite transmission line there is no discontinuity to start a reflected wave, thus the source supplies power in a forward

direction, but does not get it back in terms of a reflected wave from the termination point. Therefore there is an apparent net power loss, which actually appears as stored energy in its reactive elements (capacitances and inductances) further down the line. It was also shown that radiation plays absolutely no role in resolving this paradox.

Appendix – Input impedance of a driven LC circuit computed from a superposition of incident and reflected waves

Here we explicitly demonstrate that a driven LC circuit can be treated as an open-circuit 1-block ladder network having incident and reflected waves and from their superposition, the voltages and currents, and in particular, input impedance of the LC circuit can be calculated for all driving frequencies. We denote by V_0, I_0 and V_1, I_1 the voltages and currents at the input (AA) and termination (BB) respectively, and which are related (Fig. 3(a)) by $V_0 - V_1 = j\omega LI_0/2$, $I_0 = j\omega CV_1$, where ω is the frequency at which the circuit is being driven by, say, a generator at the input end AA. The input impedance $Z_i = V_0/I_0$ is given by,

$$Z_i = j\omega L/2 + 1/(j\omega C). \quad (26)$$

Denoting voltages and currents for the incident and reflected waves by V', I' and V'', I''

respectively, the boundary conditions at open end BB in Fig. (3a) imply $V_1'' = V_1'$ and $I_1'' = -I_1'$, the minus sign arising because the reflected current is out of phase with the incident wave by an angle π , so as to make the net current $I_1 = I_1' + I_1'' = 0$. However to evaluate I_1' , we need to isolate the incident wave and which can be done by terminating the circuit in its characteristic impedance Z_0 (Fig. 3(b)). The propagation factor for the incident wave from Eq. (17) is,

$$\frac{V_1'}{V_0'} = \frac{I_1'}{I_0'} = 1 - \omega^2 LC/2 - j\sqrt{\omega^2 LC}\sqrt{1 - \omega^2 LC/4}, \quad (27)$$

with $V_0'/I_0' = V_1'/I_1' = Z_0$. As demonstrated in 5.1.1, the incident wave is not affected by the termination impedance. The difference it makes is only in the reflected wave. For instance, in the open circuit case (Fig. 3(a)) there also exists a reflected wave, while there is no reflected wave present when the circuit is terminated in its characteristic impedance Z_0 (Fig. 3(b)).

For the reflected wave in Fig. 3(a) one can write the propagation factor as,

$$\frac{V_0''}{V_1''} = \frac{I_0''}{I_1''} = 1 - \omega^2 LC/2 - j\sqrt{\omega^2 LC}\sqrt{1 - \omega^2 LC/4}. \quad (28)$$

Equation (27) can be rewritten as,

$$\frac{V_0'}{V_1'} = \frac{I_0'}{I_1'} = 1 - \omega^2 LC/2 + j\sqrt{\omega^2 LC}\sqrt{1 - \omega^2 LC/4}. \quad (29)$$

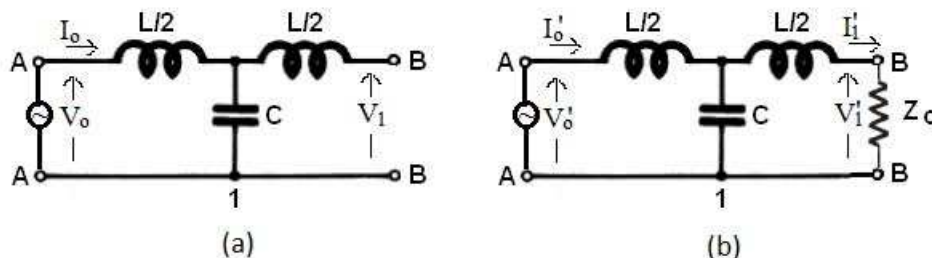


Figure 3: A driven LC Circuit or a single-block network (a) open-circuited (b) terminated in its characteristic impedance Z_0 and thereby carrying only the incident voltage and current with no reflection at the end BB.

From Eqs. (28) and (29) we get for the voltage V_0 and current I_0 as the superposition of the incident and reflected waves,

$$V_0 = V'_0 + V''_0 = 2V'_1(1 - \omega^2 LC/2) \quad (30)$$

$$I_0 = I'_0 + I''_0 = 2I'_1 j\sqrt{\omega^2 LC} \sqrt{1 - \omega^2 LC/4}. \quad (31)$$

Therefore we get the input impedance Z_i as,

$$Z_i = \frac{V_0}{I_0} = \frac{V'_1}{I'_1} \frac{1 - \omega^2 LC/2}{j\sqrt{\omega^2 LC} \sqrt{1 - \omega^2 LC/4}}. \quad (32)$$

Using $V'_1/I'_1 = Z_0 = \sqrt{L/C - \omega^2 L^2/4} = \sqrt{L/C} \sqrt{1 - \omega^2 LC/4}$, we get $Z_i = j\omega L/2 + 1/(j\omega C)$, which of course is the expected result (Eq. (26)). The input impedance is *imaginary* for all driving frequencies.

Acknowledgements

I thank Prof. S. C. Dutta Roy of IIT Delhi for his comments and suggestions on the manuscript.

References

- [1] J. D. Ryder, Network, Lines and Fields, 2nd ed., (Prentice Hall, NJ, 1955), Ch. 6
- [2] E. D. Jordan and K. G. Balmain, Electromagnetic Waves and Radiating Systems, 2nd ed., (Prentice Hall, NJ, 1968), Ch. 7
- [3] S. Ramo, J. R. Whinnery and T. V. Duzer, Fields and Waves in Communication Electronics, (Wiley, NY, 1965), Ch. 1
- [4] R. P. Feynman, Lectures on Physics, Vol. 2, (Addison-Wesley, Reading, MA, 1964), Ch. 22
- [5] S. C. Dutta Roy, Proc. Indian Natn. Sci. Acad. 73 (1), (2007) 33-36
- [6] S. J. van Enk, Am. J. Phys., 68, (2000) 854-856

- [7] A. M. Dykhne, A. A. Snarskiĭ and M. I. Zhenirovskiĭ, *Physics - Uspekhi*, 47 (8), (2004) 821-828
- [8] A. U. Keskin, D. Pazarci and C. Acar, *Am. J. Phys.*, 73, (2005) 881-88
- [9] C. Ucak and C. Acar, *Eur. J. Phys.*, 28, (2007) 321-329
- [10] C. Ucak and K. Yegin, *Eur. J. Phys.*, 29, (2008) 1201-1209
- [11] H. Krivine and A. Lesne, *Am. J. Phys.*, 71, (2003) 31-33
- [12] A. G. Ramm and L. Weaver, *Int. J. Appl. Math. Sci.*, 1, (2004) 111-116

Mermin-Wagner Theorem and its implications

Debnarayan Jana¹

¹Department of Physics
University of Calcutta
92 A PC Road, Kolkata- 700009, India.
djphy@caluniv.ac.in

(Submitted: 16-05-2016)

Abstract

In this article, we would like to discuss the implications of Mermin-Wagner theorem, a well known theorem used in condensed matter physics to rule out the possibility of spontaneous magnetisation at non-zero finite temperature in one and two dimension for some class of model Hamiltonian having continuous symmetry. This similarity of the absence of spontaneous magnetization can be invoked in other branches of condensed matter physics. Finally, we will try to shed some light on the debate related to the stability of 2D graphene sheet and magnetism of 1d finite linear chain consisting of Co atoms in the context of this theorem.

Keywords: Continuous Symmetry, Spontaneous Magnetization, Short-range Interaction, Magnetic Anisotropy

1. Introduction

In recent years, there is considerable interest in looking up some intriguing effects of external magnetic field on the properties of various emerging materials and magnetic phase transition. This is also to be noted

that the effect of magnetic field on materials can be understood in the framework of quantum mechanics only. The word phase according to Gibbs[1] refers to a state of matter which is uniform both in chemical composition as well as in physically. The role of uniformity is important. For example,

the density or the conductivity must be same throughout the system in addition to the uniformity associated with the chemical composition. In condensed matter physics, the occurrence of a phase transition is connected to the failure of one of the phases to exhibit a certain symmetry property associated with the underlying Hamiltonian. As an example, crystals by their very lattice structures, break the translational symmetry encountered in the continuum description of fluids. On the otherhand, ferromagnets, are not invariant under rotations (denoted as $O(3)$) in spin-space, although the underlying Heisenberg Hamiltonian describing the system may well be. In case of superfluids and superconductors, a breaking of gauge invariance ($U(1)$ continuous) occurs. It is known that a single spin is unable to show a phase transition in the thermodynamic sense. Thus, the suitable interaction is necessary for the occurrence of phase transition. If the nearest-neighbor spins in a system possess a lower energy if they are oriented parallel to each other than if they are anti-parallel, then, the system is known as ferromagnet (FM) and is characterized by a positive value of the strength of interaction ($J > 0$). There exists other class of materials in which neighboring spins align in opposite directions. These systems are called antiferromagnets (AFMs) and are eventually characterized by the negative value of J . In the absence of an external applied magnetic field, ferromagnetic materials exhibit a cooperative phenomenon known as spontaneous magnetization at low enough temperatures. However, when the material is heated, the magnetization decreases and

vanishes at a critical temperature, known as the Curie temperature. Beyond this critical temperature (at higher temperatures), the material possesses no net magnetization and becomes paramagnet (PM). The PM spin state is less ordered but more symmetric compared to FM one as there is no restriction on the alignment of spins. The continuous $O(3)$ symmetry associated with PM is spontaneously broken [2, 3] in FM state when the temperature becomes below the critical transition temperature and thus establishing more ordered but less symmetry in the spin state. This transition is categorised as second order or continuous phase transition as the first derivative of the free energy continuously changes with temperature while the second derivatives either diverge or change discontinuously. Hohenberg, Mermin, and Wagner, in a series of papers [3, 4], proved in 1960's that one or two dimensional (2d) systems with a continuous symmetry cannot have a broken symmetry at finite temperature. Hohenberg considers *superfluid problem* while Mermin-Wagner took magnetism associated with one or two dimensional system having continuous symmetry. Ironically, the title of the paper by Hohenberg was "Existence of long range order in one and two dimensions". The non-existence of spontaneous magnetization in one dimensional ferromagnetic linear chain was rigorously proved by Dyson [5]. Sidney Coleman [6] showed the absence of spontaneous symmetry breaking (SSB) in $d \leq 2$ dimensional systems in quantum field theory. Later, Añãños [7] has generalized Coleman-Mermin-Wagner theorem in ϕ^6

scalar field theory in $(2 + 1)$ dimension at finite temperature. Many intriguing aspects of Mermin-Wagner (MW) theorem have been comprehensively reviewed by Gelfert and Nolting [8]. A quantum version of quantum rotors in two dimensions has been described in detail by Kelbert and Suhov [9]. The versatile aspects of MW theorem in two dimensional statistical models have been discussed in the literature [10, 11, 12, 13, 14, 15]. Ghosh [16] more specifically, recovered the MW theorem for the Hubbard model, an important key model in many body systems.

The paper is organized as follows. Introducing the two important models in statistical mechanics, we proceed to convey the reader the simplest proof of MW theorem via scaling theory in section 3. After discussing the proof, We demonstrate a field theoretic approach to this theorem in section 4. We also present another approach to MW theorem via Landau-Ginzburg model in section 5. In section 6, the implications are indicated. Finally, in section 7, we give our conclusions.

2. Ising vs Heisenberg Model

Ising and Heisenberg are two important key models in statistical physics. In both models, spins are attached to the lattice points having nearest neighbour interaction. The simplest Ising type model in an external magnetic field

H_{ext} can be written as

$$H = - \sum_{\langle ij \rangle} J_{ij} S_i S_j - H_{ext} \sum_i S_i \quad (1)$$

with J_{ij} is the strength of interaction between the nearest neighbour spin pairs i and j . However, the spins in Ising model cannot be rotated rather only flipped by a sign ($S_i \rightarrow -S_i$). In other words, Ising model constrains the spins to point either up or down. But the Heisenberg model defined as

$$H = - \sum_{\langle ij \rangle} J_{ij} \vec{S}_i \cdot \vec{S}_j - \vec{H}_{ext} \cdot \sum_i \vec{S}_i \quad (2)$$

allows the spins to point any direction leading to $O(3)$ symmetry with respect to short range spin interaction. Note that $O(3)$ has nothing to do with translation or point group symmetry which will generally be discrete in nature. Mathematically speaking, if $H(\vec{S}_1, \vec{S}_2, \dots, \vec{S}_N)$, then $H(R\vec{S}_i) = H(\vec{S}_i)$ where R belongs to $O(3)$. Note that classical model uses classical spin (a fixed length vector) instead of the quantum mechanical spin operator.

3. Statement of the MW theorem and A simple proof

The theorem is related with low dimensional (one and two) system having continuous symmetry associated with the model Hamiltonian. It is not applicable to discrete symmetry such as Z_2 associated with Ising

spin Hamiltonian. The statement of the theorem goes as follows: In one and two dimensions, continuous symmetries cannot be spontaneously broken at finite temperature in isotropic systems with sufficiently short-range interactions. Stated in another way, at any non-zero finite temperature, a one or two dimensional (2d) isotropic spin- S Heisenberg model with finite-range exchange interaction can be neither FM nor AFM.

In other words, the spontaneous magnetization at finite non-zero temperature is not possible for one dimensional ferromagnetic system. Thus, FM exists with short range interaction for the isotropic case only in three spatial dimensions. The long range ordered state that breaks a continuous symmetry spontaneously at non-zero finite temperature in low dimensions is impossible. The fluctuations in lower dimension are strong enough to destroy the SSB. It is essential to note that the theorem is restricted only to the non-existence of spontaneous magnetization. It does not however, necessarily exclude other types of phase transitions such as the divergence associated with susceptibility or correlation length. In fact, in 2d XY model, there exists a phase transition known as Kosterlitz-Thouless [17] one in a topological sense. This result is a universal in the sense that it can be applicable not only to magnets but also to solids, superfluids and other systems having a broken continuous symmetry. This theorem is also valid for arbitrary spin S . For $T = 0$, the corresponding rigorous statements have been given by Pitaevskii and Stringari [18]. This

important theorem is mostly proved via the rigorous Bogolibov inequality [2, 19]. Some interesting applications of Bogolibov's inequality in equilibrium statistical mechanics have been demonstrated by Mermin [20].

However, in this pedagogical approach, we would like to have an alternative approach the problem from a simple scaling theory given by Bloch [21] long before Mermin-Wagner came into the picture. The elementary excitations in ordered distribution of magnetic moments (whether parallel or antiparallel) are known as magnons. Like phonons, they obey Bose-Einstein distributions as there are no restrictions on the number of magnons in any of the quantized energy levels connected with spin waves. The magnetization M at any finite non-zero finite temperature (T) of a ferromagnetic system can be defined as

$$M(T) = M(0) - \Delta M(T) \quad (3)$$

where $\Delta M(T)$ indicates the contribution of thermally excited magnons in spin waves. Using Bose-Einstein distribution, (Noting that for both phonons as well as magnons, the chemical potential is zero) $\Delta M(T)$ can be expressed as

$$\Delta M(T) = \int_0^\infty \frac{N(E)dE}{\exp(\beta E) - 1} \quad (4)$$

with $\beta = (k_B T)^{-1}$ and $N(E)$ is the *spin* density of states. At this junction, it is important to point out the argument leading to the proof of MW theorem depends critically on the dependence of *spin* density of states

$N(E)$ on the spatial arbitrary dimension d for a given elementary excitation of magnons in the form $E(k) \sim k^s$ ($s > 0$). Noting that $N(k)d^d k = N(E)dE$, it is easy to establish that

$$N(E) \sim E^{(d-s)/s} \quad (5)$$

Now, for ferromagnetic spin waves [2] ($E(k) \sim k^2$) in 2d, it is easy to convince that spin density of states $N(E) \sim \text{constant}$, independent of energy E . Hence, the contribution of thermally excited spin waves at low temperature can be written

$$\Delta M(T) = \text{const.} T \int_0^\infty \frac{dx}{e^x - 1} \quad (6)$$

A careful look into the above integral (for small x) reveals that the integral diverges logarithmically. This indicates that $\Delta M(T)$ diverges for finite T signalling a breakdown of magnetic order for $T > 0$. The essential point to note that at any finite temperature, it is quite easy to excite spinwaves thus destroying the magnetic order for ferromagnetic system in two spatial dimensions.

But what happens at one dimension? In this case, it is evident from equation (5) that $N(E) \sim E^{-1/2}$. Therefore, the equation (6) takes the form

$$\Delta M(T) = \text{const.} T^{-1/2} \int_0^\infty \frac{dx}{x(e^x - 1)} \quad (7)$$

that diverges in the lower integral limit indicating the absence of FM or spontaneous magnetization at finite non-zero temperature in one spatial dimension. However, in three spatial dimensions, the similar analysis

yields a finite non-zero value of $\Delta M(T)$ and the order is restored. The extension to the case of AFM is immediate by introducing a staggered field and staggered magnetization as done by Mermin and Wagner [4]. For 2d antiferromagnetic system, $N(E) \propto E$ and the scaling arguments result a divergent integral of *staggered* magnetization for small value of x . This completes the proof for FM and AFM systems. The proof of absence of AFM in low dimensional systems has also been beautifully illustrated by Keffler et al. [22]. For $d = 3$, the above analysis for FM system points out $M(T) = M(0) - AT^{3/2}$ (the famous Bloch's $T^{3/2}$ law).

The theorem is strictly valid for $T > 0$. It does not rule out the possibility for the system to exhibit SSB and ordering even at $T = 0$. The theorem also points out an important aspect regarding the lower critical dimension of a model used in statistical mechanics. The lower critical dimension of a model is defined as the space dimension at and below which there is no SSB and no transition to an ordered state. Thus, the theorem predicts that the lower critical dimension for the Heisenberg model is two while that of the Ising (Z_2) model is one.

However, few comments are necessary to justify the important terms associated with MW theorem.

- *Isotropic Interactions:* The theorem is valid only for the isotropic Heisenberg model. Suppose, the interaction between two spins at neighboring vertices i and j takes the form in 2d (XY model)

$-J\vec{S}_i \cdot \vec{S}_j$. The isotropy is essential in the sense that if we replace the above interaction by $-J_1 S_x(1)S_y(1) - J_2 S_x(2)S_y(2)$, with different constants J_1 and J_2 , then there would be spontaneous magnetization at low temperatures. This can be visualized as follows. In presence of ferromagnetic anisotropy, the dispersion relation [23] takes the form

$$E(k) = A + Ck^2, A = \text{const.} \quad (8)$$

signalling a gap at $k = 0$ in contrast to gapless excitations of isotropic FM system. This is however expected in the sense that in presence of an anisotropy ($J_1 \neq J_2$) even a rotation of the magnetic moment in the limit of $k \rightarrow 0$ requires a finite energy. The presence of such a gap is reflected in the exponential dependence of specific heat and magnetization at low enough temperature in contrast to power law dependence of the same physical quantity on T for isotropic one. The equation (6) now reduces to

$$\Delta M(T) = \text{const.} T \int_A^\infty \frac{dx}{e^x - 1} \quad (9)$$

Thus, we notice that the lower boundary is now shifted ($A > 0$) the integral does not diverge at this boundary and the magnetic order is thus stabilised by anisotropy. It has been shown by Bander and Mills [24] that even arbitrarily small anisotropies can restore long-range order in the system. This explains the existence of a number of 2d Heisenberg FMs and AFMs like K_2CuF_4 .

- *Sufficiently Short Range:* The MW theorem is strictly valid for such systems for which the exchange integral J decreases sufficiently fast with increasing distance r for N spins so that the quantity

$$Q = \lim_{N \rightarrow \infty} \frac{1}{N} \int_0^\infty r^2 J(r) Dr \quad (10)$$

remains finite. In Mermin-Wagner's paper [4], the above short-range condition is stated as

$$\sum_r r^2 |J|_r < +\infty \quad (11)$$

For interactions with a finite range or with an exponential decay, the condition in equation (11) is easily satisfied. For interactions dictated by power law decay such as $|J|_r \sim r^{-\alpha}$, the condition (11) is satisfied only when $\alpha > d + 2$, where d is the space dimensionality (i.e., $\alpha > 4$ for $d = 2$ or $\alpha > 3$ for $d = 1$) Note that in metallic magnetic systems, the exchange interactions are of the Ruderman-Kittel- Kasuya- Yosida (RKKY) type [2], which have a long-range oscillatory behavior for large r as $J(r) \sim \frac{\cos(q_0 r + \phi)}{r^d}$. It is clear that the RKKY interactions do not satisfy the criterion of short rangedness as given in equation (11). As a result, no conclusion on the magnetism related to 1d and 2d RKKY systems can be obtained from MW theorem. However, later on Bruno [25], extended the result of Mermin and Wagner to Heisenberg and XY systems

with a long-range interaction not satisfying the condition (11). In terms of energy $E(k) = \sum_r J(r)(1 - e^{i\vec{k}\cdot\vec{r}}) = E(-k)$ MW theorem is restated in the first Brillouin zone (BZ) in $d = 1$ or 2 dimensional Heisenberg and XY model as

$$\frac{1}{\Lambda_d} \int \frac{d^d k}{E(k)} = +\infty \quad (12)$$

where Λ_d is the volume of the first BZ. For long range interaction, the dispersion relation [26] as $E(k) \sim \sqrt{k}$, then the spin density of states at two dimension turns out as $N(E) \sim E^3$. In that situation, the equation (6) turns out

$$\Delta M(T) = const.T^2 \int_A^\infty \frac{x^3 dx}{e^x - 1} \quad (13)$$

And it is easily seen that the integral does not diverge at the lower limit. Thus, the magnetic order is stabilized for order such long range interaction even at 2d.

The reason for the observation of the magnetism or long range order is due to the presence of an essential magneto-crystalline anisotropy and the dipolar interaction (although very small indeed) [27, 28]. They are just enough in breaking the claims of MW theorem. Even the layered crystal structures are generally anisotropic in nature. The origin of anisotropy lies in the relativistic spin-orbit interaction. The isotropy in Heisenberg model can be destroyed by either by making $J_x \neq J_y \neq J_z$ or adding

terms of the form $-K \sum_i (S_i^z)^2$. The typical dipolar term looks as

$$H_{dipol} = -\frac{1}{2} \sum_{i \neq j} Q_{ij}^{\alpha\beta} \vec{S}_i^\alpha \cdot \vec{S}_j^\beta$$

$$Q_{ij}^{\alpha\beta} = (g\mu_B)^2 \frac{3R_{ij}^\alpha R_{ij}^\beta - \delta_{\alpha\beta} R_{ij}^2}{R_{ij}^5} \quad (14)$$

For easy axis, $K > 0$, because of the presence of gap at $k = 0$ as noted in equation (8), there is no divergence of $\Delta M(T)$ at low T in 2d indicating the presence of long range order. However, for easy plane anisotropy ($K < 0$), there is no spin gap and the dispersion looks like $E(k) \sim K$ and again $\Delta M(T)$ becomes divergent in 2d.

- *d as a continuous parameter*: Suppose we assume the spatial dimension d as a continuous parameter with $d = 2 + \epsilon$ with $\epsilon \ll 1$. Again, assuming the conventional ferromagnetic magnon dispersion relation $E \sim k^2$, we find $N(E) \sim k^{\epsilon/2}$. Therefore, the equation (5) becomes

$$\Delta M(T) = const.T^{\epsilon/2} \int_0^\infty \frac{x^{\epsilon/2} dx}{e^x - 1} \quad (15)$$

In the limit of small but finite ϵ , the above integral does not diverge. This immediately indicates that not the strict two-dimensionality, but rather slightly higher dimension such as a film with finite thickness may stabilize the order.

It is interesting to note that although the assumptions (isotropic and short-range) is usually not strictly fulfilled, it is hard to confirm MW theorem in a real system. Nevertheless, this theorem provides an important benchmark and gives a qualitative explanation why the ordering temperature T_c is usually reduced for thinner films [29]. Bensch et al.[30] have found a strong correlation between the onset of FM and of magnetic anisotropy in epitaxial Fe films on GaAs (001). The study involving magnetic and related structural issues for ultrathin Fe films grown epitaxially as wedge structures onto Ag(100) and Cu(100) shows [31] that thermal fluctuations in a nearly isotropic 2d does not diverge at the lower limit. Heisenberg system can suppress long-range order and/or give rise to domain structures with unusual characteristics.

It is interesting to note that power law decay is much slower than exponential decay which is indeed characteristic of short rangers present in the system. On the otherhand, power law implies that there is no length scale in the problem. The system is thus scale-invariant leading to quasi-long range order.

In one dimension, there exists an isotropic discrete quantum Heisenberg model having $SU(3)$ symmetry where it has been shown [32, 33] to possess a long range order in thermodynamic sense. The arguments depend critically on the effective radius of the exchange interaction between spins. In this model, the energy contribution exceeds the entropic contribution.

A general mechanism of random field induced order occurring in system having continuous symmetry has been proposed by Wehr et al.[34] and illustrated in 2d classical ferromagnetic XY model in a random uniaxial field. A non-zero spontaneous magnetization has been shown to persist even for small $T > 0$.

4. Field Theoretic approach to the Theorem

Let us start with classical XY 2d model to derive the correlation function. The model is intrinsically defined on 2d square lattice and the spin has only two components namely S_x and S_y . Thus, the model Hamiltonian with $J > 0$ is defined as

$$H = -J \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j = -J \sum_{\langle ij \rangle} \cos(\phi_i - \phi_j) \quad (16)$$

The angular bracket $\langle ij \rangle$ denotes that the interaction is restricted to nearest neighbor only. The system described by the above Hamiltonian is rotationally invariant ($O(2)$) or in terms of ϕ as translationally invariant $\phi_i \rightarrow \phi_i + c(\text{constant})$. It is clear that the energy is minimized when all the spins point in the same direction hence breaking spontaneously $O(2)$ in the ground state. The question is whether this order does indeed survive at non-zero finite temperature. To account for this, we would like to define the correlation function as

$$C(\vec{r} - \vec{r}') = \left\langle e^{i(\phi(\vec{r}) - \phi(\vec{r}'))} \right\rangle \quad (17)$$

Note that here $\langle \dots \rangle$ indicates the usual thermal average adopted in canonical distribution. Note that the correlation function is 1 at zero temperature when all the spins orient in the same direction. We claim that a long range order is established in the system at non-zero finite temperature when this correlation function is non-zero at large distances. Similarly, in case of disordered state when the two spins at large distances are uncorrelated, this correlation function goes to zero after finite correlation length. In the ordered state, the fluctuations between the neighboring spins being small, we can approximate the above Hamiltonian defined in (16) as

$$H = E_0 + \frac{J}{2} \sum_{\langle ij \rangle} (\phi_i - \phi_j)^2 \quad (18)$$

Since, the Hamiltonian is quadratic in nature, it is easy to calculate the correlation function. However, we would like to have a continuum field theory of this Hamiltonian in the limit when the correlation length is significantly larger than the lattice constant (a) of the underlying lattice. Thus, the field theoretic version of the Hamiltonian [35] looks as

$$H \approx \frac{J}{2} \int d^2r (\nabla \phi(\vec{r}))^2 \quad (19)$$

Assuming the Fourier decomposition of $\phi(\vec{r})$ as

$$\phi(\vec{r}) = \frac{1}{2\pi} \int d^2k \phi(\vec{k}) e^{i\vec{k} \cdot \vec{r}} \quad (20)$$

we find the continuum Hamiltonian reduces to

$$H = \frac{1}{2} \int E(k) d^2k \phi(\vec{k}) \phi(-\vec{k})$$

$$= \frac{1}{2} \int E(k) d^2k |\phi(k)|^2 \quad (21)$$

with $E(k) = Jk^2$. In the last step, we have used that fact $\phi(\vec{r})$ is real. With the above form of the Hamiltonian, we can compute the correlation of $\phi(\vec{k})$ and $\phi(\vec{k}')$ as

$$\begin{aligned} \langle \phi(\vec{k}) \phi(\vec{k}') \rangle &= \frac{\int \mathcal{D}\phi \phi(\vec{k}) \phi(\vec{k}') e^{-\beta H}}{\int \mathcal{D}\phi e^{-\beta H}} \\ &= \frac{\delta(\vec{k} + \vec{k}')}{\beta E(k)} \end{aligned} \quad (22)$$

where $\beta = (k_B T)^{-1}$. It is interesting to note that $\langle \phi(\vec{k}) \rangle = 0$ indicating the absence of spontaneous magnetization according to MW theorem at any finite non-zero temperature. Since the Hamiltonian is quadratic or Gaussian, the correlation function as defined in equation (17) reduces to

$$C(\vec{r} - \vec{r}') = e^{-\frac{1}{2} \langle (\phi(\vec{r}) - \phi(\vec{r}'))^2 \rangle} \quad (23)$$

Thus, the calculation of correlation function simply reduces to the calculation of expectation value of $(\phi(\vec{r}) - \phi(\vec{r}'))^2$. Therefore,

$$\begin{aligned} \langle (\phi(\vec{r}) - \phi(\vec{r}'))^2 \rangle &= \int \int \frac{d^2k d^2k'}{(2\pi)^2} \\ &\quad \left(e^{i\vec{k} \cdot \vec{r}} - e^{i\vec{k}' \cdot \vec{r}'} \right) \\ &\quad \left(e^{i\vec{k}' \cdot \vec{r}} - e^{i\vec{k} \cdot \vec{r}'} \right) \langle \phi(\vec{k}) \phi(\vec{k}') \rangle \end{aligned} \quad (24)$$

Substituting the value of $\langle \phi(\vec{k}) \phi(\vec{k}') \rangle$ from (22), we find

$$\langle (\phi(\vec{r}) - \phi(\vec{r}'))^2 \rangle = \frac{1}{2\beta\pi^2} \int d^2k \frac{1 - \cos(\delta\vec{r} \cdot \vec{k})}{E(k)} \quad (25)$$

where $\delta\vec{r} = \vec{r} - \vec{r}'$. In the limit of $k \gg (\delta r)^{-1}$, we notice a logarithmic divergence ($\frac{1}{J\beta\pi} \int_{(\delta r)^{-1}}^{\frac{dk}{k}}$) at higher momenta. Considering the original lattice model, if we put in hand the high momentum cut-off $\alpha \propto a^{-1}$, we find that

$$\langle (\phi(\vec{r}) - \phi(\vec{r}'))^2 \rangle = \frac{1}{J\beta\pi} \log(\alpha\delta r) \quad (26)$$

and the correlation function turns out as

$$C(\vec{r} - \vec{r}') \propto (\alpha\delta r)^{-\eta(T)} \quad (27)$$

with $\eta(T) = \frac{T}{2\pi J}$. The result eventually indicates the correlation function between distant spins vanishes and the system is not ordered at any finite non-zero temperature. However, one important thing to notice here is that the correlation function decays in a power law fashion without any length scale. The correlation length is in fact infinite in the sense of second order phase transition. However, instead of an isolated point in second order phase transition, the phenomenon occurs over a region of parameters. Thus, we call it as quasi-long range behavior. At high enough temperature, this model also shows a topological phase transition known Kosterlitz-Thouless [17] phase transition different from the symmetry breaking one. Note that the integral defined in equation (25) is perfectly finite for $d > 2$ ensuring the presence of magnetization and hence the breaking of spontaneous symmetry $O(n)$.

5.MW theorem From Landau-Ginzburg (LG) Model

Near the critical point T_C , the free energy in the Landau-Ginzburg Model[23] can be written in terms of order parameter (magnetization $\vec{M}(\vec{r})$) as

$$F = \int d^d r \left[\frac{1}{2} a(T) (M(\vec{r}))^2 + \frac{g}{2} (\vec{\nabla} M(\vec{r}))^2 + \frac{1}{4} b (M(\vec{r}))^4 - \vec{M}(\vec{r}) \cdot \vec{H}_{ext} \right] \quad (28)$$

Higher powers of $\vec{\nabla} M$ can be only neglected in the limit of small spatial fluctuations. It is important to note that the above expansion is not valid for small value of T_C because in that case quantum fluctuations are important than the usual thermal fluctuations. The uniform magnetization (neglecting $\vec{\nabla} M$ term) at any temperature T can be determined from

$$a(T)M + bM^3 = H_{ext} \quad (29)$$

Thus, the equilibrium value of magnetization M_0 in the absence of the external magnetic field is simply $M_0^2 = -\frac{a(T)}{b}$ for $T < T_C$ and is zero for $T > T_C$. If we now write $M(\vec{r}) = M_0 + m(\vec{r})$, where $m(\vec{r})$ is the small fluctuation over the equilibrium value M_0 , then one can write the thermal average of small fluctuations for various q (assuming Fourier transform $M(r) = \sum_q M(q) e^{i\vec{q}\cdot\vec{r}}$) in the absence of the external field H_{ext} as

$$\langle |m(q)|^2 \rangle \sim \frac{k_B T}{a(T) + gq^2 + 3bM_0^2} \quad (30)$$

The above formula indicates the large fluctuations at small value of q . From this, we can obtain

$$\begin{aligned}
 \langle M(r)^2 \rangle &= M_0^2 + \langle m(r)^2 \rangle \\
 &= -\frac{a(T)}{b} + \sum_q \langle |m(q)|^2 \rangle \\
 &= -\frac{a(T)}{b} + \\
 &\quad B \sum_q \frac{k_B T}{a(T) + gq^2 + 3bM_0^2}
 \end{aligned} \tag{31}$$

with B being a constant. In 1d as well as in 2d, the above integral diverges near T_C and the fluctuations become larger than M_0 indicating the absence of long range order at $T \neq 0$,

6. Implication of MW theorem

We can use the above logic to check the stability of infinite 2d solid. If we approximate the deviations of the potential from the equilibrium as harmonic, then the potential looks in terms of the displacement vector \vec{u}_i of i -th atom in a square 2d lattice as

$$\frac{K}{2} \sum_{\langle ij \rangle} (u_i - u_j)^2 \tag{32}$$

This form of the potential is similar to that adopted in XY model.

This indicates that the relative displacement vector between distant sites is wildly

fluctuating leading to an instability of infinite 2d square lattice structure [36].

A 2d lattice is simply a crystal lattice periodical in two dimensions and confined in the third direction. Nature however tries to find the equilibrium by folding the low dimensional crystal as done in fullerene or DNA molecules. MW theorem applies only to *infinite* 2d crystals while stable films of finite sizes have sufficient thicknesses to show magnetism below a finite non-zero critical transition temperature T_c , generally lower than their bulk counterpart. In recent years, the existence and stability of free-standing graphene [37] sheets have drawn again the attention of MW theorem. Sometimes, it has been argued that the free-standing 2d crystal would be disrupted by thermodynamic forces and hence, 2d materials were expected not to exist. Even the ultraflat graphene grown on mica surfaces has invoked rippling to explain the thermodynamic stability of free standing graphene sheets. Sometimes, it has been pointed out that the low-amplitude ripples protect graphene from disruption by perturbing it away from exact flatness thereby making the MW theorem technically inapplicable to the situation. In fact, time to time, it has been observed that there has been basic misreading about this MW theorem. As has been clearly explained above, that it talks about the long wavelength thermal fluctuations so called elastic deformation in 2d structures. This has nothing to do in fact with the disrupting structure. Note that in infinite 3d crystal, thermal fluctuations displace atoms from their mean equilibrium positions centered about the

lattice points. The MW strictly proves that in an infinite 2d crystal, thermal fluctuations will destroy long range alignment with a lattice. However, graphene (including other 2d structures) can maintain this crystalline order except in the sense that they cannot maintain strict periodicity in the limit of infinite size.

In other words, MW theorem is not about the stability of *finite* structures but about the crucial and important property of long range elastic deformation of *inifinite* systems. Most recently, the stability of graphene and graphene like other 2D materials (Silicene, Germanene etc) has been argued [38] on the basis of transverse short-range displacements of appropriate atoms. The distortions produced in these exotic 2D materials although being small can be however described in a framework of Ising model with competing interactions on in 2D hexagonal lattice and predict various patterns and buckling transitions to be observed in Raman and infrared experiments. In fact, the preference of the buckled structures of silicene and germanene, are consistent with the MW theorem due to the vanishing of the effective bending for a very broad range of electron-phonon interactions [39].

It is known that a macroscopic piece of Co remains ferromagnetic below its Curie temperature 1388 K. Can one engineer magnetism in atomic scale (nanoscale)[40]? Or in other words, what will happen if we consider a finite nanowire consisting of few Co atoms on platinum (Pt) surface? But the measurements of Gambardella and colleagues

[41, 42, 43, 44] have shown unambiguously by two methods that the chain is indeed a ferromagnet with a Curie temperature of about 15 K. Is MW theorem applicable here or in error? In fact, the measurements pointed out the presence of a giant magnetic anisotropy [42, 44] existing in the system thus rendering MW theorem to be invalid here. Another point is important here about the finite length of the chain (~ 50 Co atoms). The original MW theorem assumes the chain to be infinitely long while the system prepared by Gambardella and colleagues is of finite length. These two aspects clearly show the presence of finite magnetism in this one-dimensional system. A simple argument due to Landau [45, 46] can be given to show that the finite chains are stable magnets [41, 43].

Let us consider a chain consisting of $(N-1)$ with nearest- of N moments described by the Ising Hamiltonian $H = -J \sum_{i=1}^N S_i^z S_{i+1}^z$ with neighbour exchange coupling energy $J > 0$ (FM interaction). The ground state energy of the system can be computed when all the spins are either up ($S_i = +1$) or down ($S_i = -1$) and is thus doubly degenerate. Thus the energy is given as $E_0 = -J(N-1)$ and the situation corresponds to the case where all the moments are aligned. The lowest-lying excitations are those in which a single break occurs at any one of the N sites of parallel aligned states. In practice, however, there are only $N-1$ such excited states, all of them having the same energy $E = E_0 + 2J$. At a finite non-zero temperature T the change in free energy due to these excitations

$$\Delta G = 2J - k_B T \ln(N-1) \quad (33)$$

In the limit $N \rightarrow \infty$, we have $\Delta G < 0$ at any finite temperature and the ferromagnetic state becomes unstable against thermal fluctuations. For $(N - 1) < e^{\frac{2J}{k_B T}}$, however, it is possible to have energetically stable ferromagnetic order. As an explicit order of magnitude of a critical transition temperature of 50 atoms in linear chain having $2J = 15$ meV [47, 48], we get an estimation of $T_C = 45K$. In fact, the measurements of the magnetization in the above Co monatomic chains agree with this limit. If this chain is stable at room temperature, then it can be used as high density magnetic storage material on computer disk.

7. Conclusions

To conclude, we have discussed a simple scaling proof of Mermin-Wagner Theorem and indicated the important roles placed by the isotropic and long range nature of the interaction. The field theoretical calculations to show the validity of this MW theorem have also been included. We have also argued the validity of this MW theorem to graphene and 1d finite linear chain consisting of Co atoms.

Acknowledgments

I hope that the presentiaion will be helpful for graduate students working in any branch of theoretical physics. The author would like to acknowledge Mr. Partha Nandy and Mr. Suman Chowdhury for asking several interesting questions related to this subject which

prompted me to write this paper.

References

- [1] J. Cardy, *Scaling and renormalization in statistical physics*, (Cambridge: Cambridge University Press, 2002), ISBN 9780521499590.
- [2] W. Nolting and A. Ramakanth, *Quantum Theory of Magnetism* (Springer, 2009)
- [3] P. C. Hohenberg, *Phys. Rev.*, **158**, 383 (1971).
- [4] N. D. Mermin and H. Wagner, *Phys. Rev. Lett.*, **17**, 1133 (1966). N. D. Mermin, *J. Math. Phys.*, **8**, 1061 (1967). See for an interesting issue regarding short range interaction; *Phys. Rev. Lett.*, **87**, 137303 (2001) and *Phys. Rev. Lett.*, **107**, 107201 (2011). For details. N.D. Mermin, *Rev. Mod. Phys.*, **51**, 591 (1979).
- [5] F. J. Dyson, *Commun. Math. Phys.*, **12**, 212 (1969)
- [6] S. Coleman, *Commun. Math. Phys.*, **31**, 259 (1973).
- [7] G. N. J. Añaños, *J. Math. Phys.*, **47**, 012301 (2006)
- [8] A. Gelfert and W. Nolting, *J. Phys.: Condens. Matter*, **13**, R505 (2001).
- [9] M. Kelbert and Y. Suhov, *J. Math. Phys.*, **54**, 033301 (2013).

- [10] R. L. Dobrushin and S. B. Shlosman, *Comm. Math. Phys.*, **42**, 31 (1975).
- [11] C. E. Pfister, *Comm. Math. Phys.*, **79**, 181 (1981).
- [12] J. Fröhlich, C. E. Pfister, *Comm. Math. Phys.*, **81**, 277 (1981).
- [13] A. Klein, L. J. Landau, D. S. Shucker, *J. Stat. Phys.*, **26**, 505 (1981).
- [14] C. A. Bonato, J. F. Perez, A. Klein, *J. Stat. Phys.*, **29**, 159 (1982).
- [15] D. Ioffe, S. B. Shlosman and Y. Velenik, *Comm. Math. Phys.*, **226**, 433 (2002)
- [16] D. K. Ghosh, *Phys. Rev. Lett.*, **27**, 1584 (1967).
- [17] J. M. Kosterlitz and D. J. Thouless, *J. Phys. C*, **5**, L124 (1972); *ibid*, **6**, 1181 (1973); *ibid*, **7**, 1046 (1974).
- [18] L. P. Pitaevskii and S. Stringari, *Jour. Low Temp. Phys.*, **85**, 377 (1991)
- [19] G. Roepstorff, *Comm. Math. Phys.*, **53**, 143 (1977).
- [20] N. D. Mermin, *J. Phys. Soc. Jpn*, **26**(Suppl), 203 (1969)).
- [21] F. Bloch, *Z. Phys*, **61**, 206 (1930).
- [22] F. Keffler, H. Kaplana and Y. Yafet, *Am. J. Phys.*, **21**, 250 (1953).
- [23] E. M. Lifshitz and L. P. Pitaevskii, *Statistical Physics*, Part 2 (Butterworth Heinemann, Oxford, 1980).
- [24] M. Bander and D. L. Mills, *Phys. Rev. B*, **38**, 12015 (1988).
- [25] P. Bruno, *Phys. Rev. Lett.*, **87**, 137203 (2001)
- [26] O. Grechnev, *Theoretical Studies of Two- Dimensional Magnetism and Chemical Bonding*, (Ph.D. Thesis, Uppsala University, Sweden, 2005, ISBN 91-554-6164-6).
- [27] V. Yu. Irkhin, A. A. Katanin and M. I. Katsnelson, *Phys. Rev. B*, **60**, 1082 (1999).
- [28] A. Grechnev, V. Yu. Irkhin, M. I. Katsnelson and O. Eriksson, *Phys. Rev. B*, **71**, 024427 (2005).
- [29] Schneider et al., *Phys. Rev. Lett.*, **64**, 1059 (1990).
- [30] F. Bensch, R. Moosbühler, and G. Bayreuther, *J. Appl. Phys.*, **91**, 8754 (2002)
- [31] S. D. Bader, Dongqi Li, and Z. Q. Qiu, *J. Appl. Phys.*, **76**, 6418 (1994).
- [32] D. J. Thouless, *Phys. Rev.*, **187**, 732 (1969).
- [33] Y. G. Rudoy and O.A. Kotelnikova, *J. Mag. Mag. Mat*, **324**, 3605 (2012).
- [34] J. Wehr, A. Niederberger, L. Sanchez-palenia and M. Lewenstein, *Phys. Rev. B*, **74**, 224448 (2006)

- [35] C. Itzykson and J. M. Drouffe, *Statistical Field Theory*, (Cambridge University Press, 1989), Vol I.
- [36] N. D. Mermin, *Phys. Rev.*, **176**, 250 (1968).
- [37] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov and A. K. Geim, *Rev. Mod. Phys.*, **81**, 109 (2009).
- [38] A. O'Hare, F. V. Kusmartsev and K. I. Kugel, *Nano Lett.*, **12**, 1045 (2012).
- [39] Wei Wei, Ying Dai, Baibiao Huang and Timo Jacob, *Phys. Chem. Chem. Phys.*, **15**, 8789 (2013).
- [40] J. Kübler, *J. Phys.: Condens. Matter*, **15**, V21 (2003).
- [41] P. Gambardella, M. Blanc, H. Brune, K. Kuhnke and K. Kern, *Phys. Rev. B*, **61**, 2254 (2000).
- [42] P. Gambardella, A. Dallmeyer, K. Maiti, M. C. Malagoli, W. Eberhardt, K. Kern and C. Carbone, *Nature*, **416**, 301 (2002).
- [43] P. Gambardella, *J. Phys.: Condens. Matter*, **15**, S2533 (2003).
- [44] P. Gambardella, A. Dallmeyer, K. Maiti, M. C. Malagoli, S. Rusponi, P. Ohresser, W. Eberhardt, C. Carbone, and K. Kern, *Phys. Rev. Lett.*, **93**, 077203 (2004).
- [45] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Vol. **5**, 482 (Pergamon, London, 1959).
- [46] M. Plischke and B. Bergerson, *Equilibrium Statistical Mechanics*, (Prentice-Hall International, NJ, 1989).
- [47] M. Pratzner, H. Elmers, M. Bode, O. Pietzsch, A. Kubetzka and R. Wiesendanger, *Phys. Rev. Lett.*, **87**, 127201 (2001).
- [48] S. Frôta-Pessoa and R. M. J. Kudrnovský, *Phys. Rev. B*, **62**, 5293 (2000).

The genesis of the internal resistance of a battery – a physical perspective

Ashok K. Singal

Astronomy and Astrophysics Division
Physical Research Laboratory
Navrangpura, Ahmedabad - 380 009, India.
asingal@prl.res.in

(Submitted 14-11-2015)

Abstract

The standard exposition of the internal resistance of a battery, that a battery comprises a source of emf in series with an internal resistance, as given in engineering and physics text-books, is lacking in proper explanation. It is treated merely as an experimental fact, and not something that should follow from logic. The battery has a tendency to maintain electric potential difference across its terminals equal to its chemical potential, and in an open circuit, when no current flows, these two do match. However in a closed circuit, a drop in electric potential across the battery terminals is inevitable for a steady flow of electric current throughout the circuit, because the chemical reactions driving the electric current within the battery can proceed only if the electric potential at its terminals differs from the chemical potential. It is shown that for small voltage changes, the current passing through the battery is linearly proportional to the change in potential from the open-circuit value (i.e., its chemical potential), giving rise to a semblance of an internal resistance in series with the external resistance. It follows that a battery *has to have* an internal resistance in order to function as a power source. It is also shown that Thevenin's theorem does not make our results superfluous, in fact our results are presupposed in its derivation.

Introduction

In almost all physics or engineering undergraduate text-books [1, 2, 3, 4], the internal resistance of a battery is introduced more or less as a factual statement that a battery comprises a source of emf in series with an internal resistance R_i (Fig. (1a)), which is the resistance of the electrolyte of the battery. In general there is not much exposition as to the genesis of the internal resistance of the battery and more specifically why it needs to be put outside the battery in series with the external resistance. A student soon learns to live with it and, at most taking it as an experimental fact, moves on. But a feeling remains that something is lacking. After all the internal resistance is due to the constituents of the battery within it, therefore the word “internal resistance” conjures up a vision like that of a resistance internal to the battery like in Fig. (1b), or equivalently where the emf of the battery is across its internal resistance. Then with a finite voltage at its ends (positive and negative electrodes of the battery) one expects from Ohm’s law that there should be a finite current flowing through the resistor within the battery (even if the external circuit were open) as long as a finite electric potential difference exists across the resistor. Further, in an actual circuit when the circuit is closed and an electric current does flow through the internal resistance, it is in a direction from a lower electric potential (negative electrode) to a higher one (positive electrode) within the battery (Fig. (1c)), contrary to that expected from Ohm’s law for

a normal resistor where the electric current should flow from higher to a lower electric potential which is not seen within the battery. Additionally, as the internal resistance is supposedly that of the electrolyte residing in-between the two electrodes of the battery, how come the internal resistance is shown to exist not between the two electrodes as in Fig. (1b), but – somewhat mysteriously – is instead put in series with the external resistance outside the battery Fig. (1c)? Moreover, why should it be causing a drop in potential from the open-circuit value when the internal resistance itself is a part and parcel of the battery system, giving rise to that potential? Here we should clarify that we are not doubting the truth of the long-known experimental facts (for example, we know experimentally that Fig. (1a) and Fig. (1c) are factually correct while Fig. (1b) is not), we are only attempting to logically examine these facts from a simple physical perspective.

EMF, voltage and current

A battery or cell is an electrochemical device [5] that converts chemical energy to electric energy by driving an electric current through a circuit containing a load (an external resistance). Historically, a single chemical source of emf was called a cell and a set of interconnected cells was called a battery, however, it is now common practice to refer to even a single cell as a battery, as is done here too. Batteries could be of different kinds, for example, disposable ones designed to be used only once and the rechargeable ones, which can

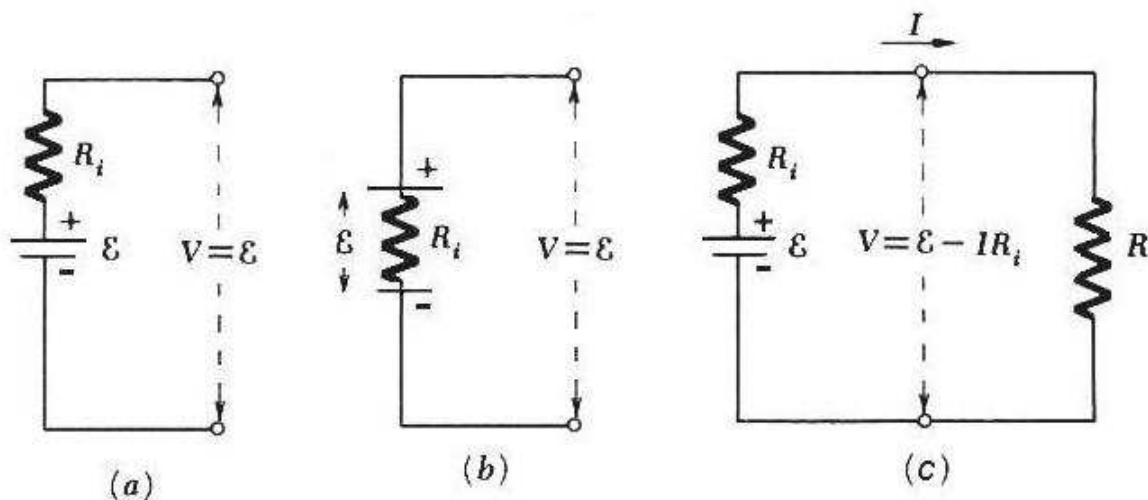


Figure 1: The voltages and currents in a battery of emf \mathcal{E} for (a) an open circuit with internal resistance R_i in series outside the battery (b) R_i “inside” the battery and (c) a closed circuit with R_i in series with the external resistance.

be used more than once. A common example of the former is the zinc-carbon battery, often used in torch-lights. The rechargeable batteries include lead-acid batteries used in automobiles while others like nickel-cadmium or lithium-ion are used in mobile phones and laptop computers. A battery in general, consists of two electrodes of different material immersed in an electrolyte, which could be a fluid or a moist paste. The electrolyte interacts chemically with the electrodes and due to their chemical reactions a push is exerted on the positive charges towards one terminal, called the positive electrode, and on the negative charges towards the other terminal, called the negative electrode. Irrespective of the make of a battery (its type, size, volume, the nature of the electrodes and the

electrolyte and the details of their chemical reactions etc.), a battery ultimately causes a separation of positive and negative charges, giving rise to an electric potential across the battery. The chemical potential, which is the line integral of the force per unit charge due to chemical reactions (from the negative electrode to the positive electrode), is called the emf \mathcal{E} of the battery (historically called electromotive force which actually is a misnomer as \mathcal{E} is not a force but is instead a potential difference between the two electrodes). To a first approximation we can write the effective force due to chemical reactions on a charge e as $F_c = e\mathcal{E}/d$, where d is the distance between the two electrodes. It is this force F_c due to the chemical reactions that pushes positive charges towards the positive

electrode and the negative charges towards the negative electrode inside the battery or cell, giving rise to an electric potential difference between the two electrodes. In general the change in electric potential within the battery may not be linear and is mostly localized at the electrode–electrolyte interfaces, but from energetic point of view what finally matters is the net potential difference V between the two electrodes. This in turn gives rise to an electric field $E = -V/d$ within the battery which exerts on every charge e an electric force eE in a direction opposite to the force F_c due to the chemical reactions. As a result the net force on a charge pushing it towards its respective electrode within the battery becomes,

$$F = F_c + eE = e \frac{\mathcal{E} - V}{d}. \quad (1)$$

As long as V is smaller than \mathcal{E} , the force $F > 0$ and the charges will continue to move inside the battery toward their respective electrodes, with more and more charges getting deposited there. However with increasing V , F will decrease, reducing the current flow inside the battery, until the electrodes achieves a voltage difference $V = \mathcal{E}$. Then from Eq. (1), $F = 0$ and the charge movement reduces to zero inside the battery. Thus in an open circuit the battery will acquire across its terminals a voltage V equal to its emf \mathcal{E} , i.e. its chemical potential (Fig. (1a)), with no net force on the charges and hence no electric current within the battery in spite of the electric potential V across its terminals.

Now let a load (an external resistor R) be connected across the battery, closing the

circuit. Immediately an electric current will start from the positive terminal towards the negative one through the external resistance. Actually the electric current in a circuit is due to the flow of electrons in a direction opposite to that of the current conventionally shown in a circuit (Fig. (1c)), but it does not alter the physics of the problem. The external current causes a deficiency of some negative charges at the negative electrode as well as neutralizes some positive charges at the positive electrode, the reduction of charges causing a slight drop in voltage from the initial open-circuit value $V = \mathcal{E}$. This means that the electric field within the battery will now be less than the open-circuit value (i.e., $V < \mathcal{E}$ or from Eq. (1), $F = F_c + eE > 0$) and the electric force will not fully cancel the force F_c on the charges within the battery due to the chemical reactions. This in turn will cause the charges to move according to Eq. (1) giving rise to a positive current from the negative electrode to the positive electrode inside the battery. Initially as the current within the battery may be less than that in the external circuit, the charges getting replenished at the terminals will be less in quantity than those getting depleted by the external current flow, therefore the voltage V will be falling still further. And as $\mathcal{E} - V$ increases, this should give rise to not only a still higher current within the battery, it will also cause a drop, even if only slight, in the external current as the voltage V across the external resistor drops. A stage however, will be reached very soon when the internal current within the battery becomes equal to that in the external circuit. Now onwards there will

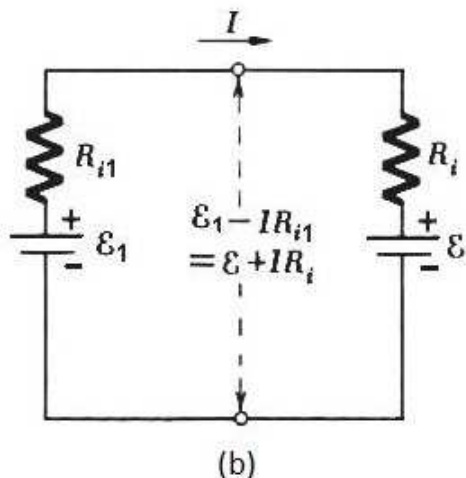


Figure 2: The voltages and currents in a battery of emf \mathcal{E} when recharging with another battery of emf $\mathcal{E}_1 > \mathcal{E}$.

be no further change in the voltage at the battery terminals and a state of equilibrium has been reached. However it will remain a constant struggle for the battery, through the internal current, to keep replenishing charges being depleted at its terminals by the external current. Thus we see that a current will be flowing from a lower electric potential to the higher one within the battery because of the larger push on the charges in that direction by the force $F_c = e\mathcal{E}/d$ due to the chemical reactions than the opposing force by the electric field eV/d from a higher to a lower potential.

The internal resistance

In a closed circuit the electric current flowing within the battery is due to the chemical reactions, which will take place only if the voltage across the battery V is different from the chemical potential \mathcal{E} . A steady state means the internal current within the battery must be equal to the current $I = V/R$ through the external resistance R . The internal resistance could be defined by $R_i = 1/|dI/dV|_{V=\mathcal{E}}$. It should be noted that the current increases when V decreases. Assuming a constant dI/dV in a certain range of V around \mathcal{E} and noting that $I = 0$ when $V = \mathcal{E}$, we get $R_i = (\mathcal{E} - V)/I$. From this we could write $V = \mathcal{E} - IR_i$, which justifies representing the battery as a source of emf \mathcal{E} with its internal resistance R_i in series (Fig. (1c)). The current $I = V/R$ is then given by $I = \mathcal{E}/(R + R_i)$. One could even have a reverse current through the battery when V across the battery is made higher than \mathcal{E} . For that another source of emf, say \mathcal{E}_1 and with an internal resistance R_{i1} , so that $\mathcal{E}_1 > \mathcal{E}$ of the battery in question, is connected across its terminals (Fig. (2)). This is done, for example, to recharge the lead-acid battery or other rechargeable batteries. The magnitude of the reverse current through the battery will now be given by $I = (V - \mathcal{E})/R_i$, as the recharging voltage $V = \mathcal{E}_1 - IR_{i1}$ is larger than \mathcal{E} , then $I = (\mathcal{E}_1 - \mathcal{E})/(R_i + R)$. The reverse current means that the positive charges move towards the negative electrode while the negative charges move towards the positive electrode, thereby reversing the chemical

reaction and recharging the battery. In the case of non-rechargeable battery no reverse current takes place and we could say it has a discontinuity in its internal resistance at $V = \mathcal{E}_{+0}$. (One is in general, cautioned against attempts to recharge non-rechargeable batteries as these could explode).

Now $e\mathcal{E}$ is the amount of chemical energy expended as work on a charge e in transporting it from one electrode to the other. Out of this, an amount eV is spent against the electric field, which ultimately gets delivered to the external load, the remaining energy $e(\mathcal{E} - V)$ represents the ohmic losses within the battery. Thus $(\mathcal{E} - V)I = I^2R_i$ are the power losses in the battery as expected from a resistance R_i lying outside the battery in series.

The actual value of the internal resistance of a cell may depend upon a combination of various factors. If the effective cross-section areas of the electrodes are large, more current may flow through the battery even for the same $\mathcal{E} - V$ change, implying a lower R_i value. Similarly a larger separation between the electrodes would imply a smaller push on the charges even for the same $\mathcal{E} - V$ change (Eq. 1), resulting in a smaller current, implying a higher R_i value. The nature of the constituents (electrodes and the electrolyte) of a battery also matter as a better conducting electrolyte means a higher current for the same $\mathcal{E} - V$ and thereby a smaller R_i . Further as with usage the density of chemical components within the battery may decrease, it would lead to an increase in the internal resistance.

Thevenin's theorem

Thevenin's theorem [6] states that any two-terminal network containing energy sources (generators) and impedances can be replaced with an equivalent circuit consisting of a voltage source in series with an impedance. Thus at a first look it may appear to preempt all our above discussion, which may in fact appear redundant. But a careful look at the proof of the Thevenin's theorem shows that our above results are rather presupposed there. In the proof offered [6] one may have batteries/generators and impedances in series or in parallels or in other complicated distributions but to begin with one always has the *internal impedances*, if any, of the batteries/generators always in series with them. Therefore Thevenin's theorem does not make our results superfluous, in fact our results are made use of in its derivation.

Conclusions

We have shown that due to the tendency of the battery to attain a voltage across its terminals equal to the chemical potential \mathcal{E} , a finite drop in voltage from the open circuit value \mathcal{E} is essential for a steady current in the circuit because then and only then will there be chemical reactions taking place so that a current flows within the battery. Thus a drop in the voltage is essential for a steady state current implying the existence of a finite internal resistance in any practical battery, which can be justifiably represented as a source of emf \mathcal{E} with a resistance R_i in

series. Therefore a battery has to have an internal resistance in order to function as a power source. Further we have shown that Thevenin's theorem does not make our results superfluous, in fact our results are made use of in its derivation.

Acknowledgements

It was due to the nagging doubts expressed by Tanmay Singal, then an undergraduate student, in the standard text-book explanation for the internal resistance of a battery that prompted me to look for a more compelling physically arguments. It is hoped these would be beneficial to others as well in similar predicaments.

References

- [1] D. Halliday, R. Resnick and J. Walker, Fundamentals of Physics, 9th ed., Wiley, NJ, 2011, Ch. 27
- [2] D. C. Giancoli, Physics for Scientists and Engineers - With Modern Physics, 4th ed., Pearson Prentice Hall, NJ, 2009, Ch. 26
- [3] R. A. Serway and J. W. Jewett Jr., Physics for Scientists and Engineers - With Modern Physics, 8th ed., Brooks/Cole, Belmont, 2010, Ch. 28
- [4] M. E. Van Valkenburg, Network Analysis, 3rd ed., Prentice-Hall, NJ, 1974, Ch. 2
- [5] E. M. Purcell, Electricity and Magnetism - Berkeley Phys. Course vol. 2, 2nd ed., McGraw-Hill, New York, 1985, Ch. 4
- [6] J. D. Ryder, Network, Lines and Fields, 2nd ed., Prentice Hall, NJ, 1955, Ch. 1

Determination of The Dimensions of the Two Co-Axial Cylindrical Cavities Hidden Inside A Mechanical Black Box

Bhupati Chakrabarti¹, Shirish Pathare², Saurabhee Huli³

¹23, Ibrahimpur Road, Kolkata 700032.

bhupati2005@yahoo.co.in

²Homi Bhabha Centre for Science Education

V. N. Purav Marg, Mankhurd - 400088. Mumbai. India.

shirish@hbcse.tifr.res.in

³Department of Physics,

St.Xaviers College, Mumbai – 400 001.

saurabhee.huli@gmail.com

(Submitted: 06 -10 - 2016)

Abstract

Experiments devised to explore the hidden mechanical components in an arrangement by external measurements are known as mechanical black box (MBB). The philosophy behind any black box experiments is non-destructive testing of the hidden components supported by logical reasoning. This paper presents an experimental problem from the Physics Olympiad programme, where the students were asked to figure out the dimensions of two co-axial cylindrical cavities hidden inside a mechanical black box (MBB), by measuring length, mass and time. This paper discusses in detail the situation of two coaxial cavities of the same diameter but different lengths. In addition, possible variations of the problems are suggested. Finally, challenges and student learning in relation to this problem are discussed.

1. Introduction

Mechanical black boxes are essentially an arrangement which consist of some hidden mechanical components. These are designed in

such a way that by performing some suitable external non-destructive experiments the hidden components may be determined.

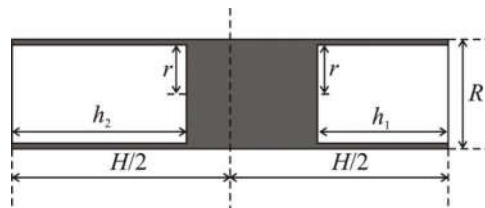


Figure 1: Schematic diagram of the Mechanical Black Box



Figure 2: Photograph of the mechanical black box

In this experiment, MBB has been made out from an iron (density 7.9 g cm^{-3}) bar with square cross section (rectangular parallelepiped) of side $R(= 3.8 \text{ cm})$ and having a length $H(= 24.0 \text{ cm})$. The weight of this solid piece (without cavities) is (M_s) 2737.8 g . Two coaxial cylindrical cavities are drilled out from the two ends of the parallelepiped. Both of these drills have same diameter (3.2 cm) but one is having a length of 13.0 cm while the other has a length of 5.0 cm . The open ends of the cavities were covered with thin light bakelite sheets. Paper scales with least count of 0.1 cm was pasted along the length of the square bar on all its sides. The box (with cavities)

now weighs about (M_{MBB}) 1639.0 g indicating that a mass of about 1100.0 g has been drilled out. Figures 1 and 2 show the schematic diagram and the actual photograph of the MBB respectively. The task is now to find out the radius r of the cylindrical drills and their respective lengths h_1 and h_2 .

2. The problem

The problem consists of three unknown quantities r, h_1, h_2 which needs three independent equations to solve. The density of iron has been supplied ($\rho_{\text{iron}} = 7.9 \text{ g cm}^{-3}$).

The actual mass of the MBB (M_{MBB}) is supplied. This gives us the first equation that comes from the removed mass (m). By subtracting the actual mass of the MBB from this projected mass of the solid box (M_s) one gets the removed mass m (masses m_1 and m_2 removed to form cavities), we get

$$M_s - M_{\text{MBB}} = m = m_1 + m_2 \quad (1)$$

$$m = \pi r^2 (h_1 + h_2) \rho_{\text{iron}} \quad (2)$$

$$h_1 + h_2 = \frac{m}{\pi r^2 \rho_{\text{iron}}} = k(\text{say}) \quad (2a)$$

Next the MBB is made to undergo torsional oscillation about the axis along its length after suspending it from a suitable steel wire (Fig.5c). The time period of its oscillation will be governed by the moment of inertia of the MBB (I_{MBB}) about this axis. Moment of inertia of the

box (I_s), had it been completely solid, will be greater than I_{MBB} . The contribution to the moment of inertia from the two cavities has caused the moment of inertia to reduce from I_s to I_{MBB} . I_s can be calculated by making measurements of the dimensions of the MBB.

So we may write

$$I_s - I_{MBB} = \frac{1}{2}mr^2 \tag{3}$$

$$r = \sqrt{\frac{2(I_s - I_{MBB})}{m}} \tag{3a}$$

Determination of I_{MBB} :

By oscillating the MBB from the point of suspension, we obtain T_{MBB} . I_{MBB} corresponding to T_{MBB} can be obtained from a calibration curve between moment of inertia against square of the time period of oscillation. To develop this calibration curve, following assembly was used.

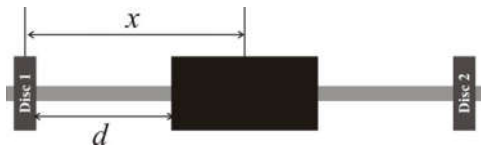


Figure 3: Assembly for calibration curve

In this assembly, the change in the position of discs will give different moment of inertia (Fig.3). When this assembly is given torsional oscillations about an axis perpendicular to the axis of the central cylinder, it gives corresponding time period of oscillation. Thus by repeating this exercise for different positions of discs (i.e. different moments of inertia), a calibration curve

between the moment of inertia and the square of the period of oscillation can be obtained.

Next the geometric midpoint of the MBB is kept over a knife edge. Since the removed masses are not equal, the MBB naturally remains in a tilted position with larger cavity side going up as more mass has been removed from that part. An external mass m_3 is placed over the MBB at a suitable distance z from the geometric midpoint so as to bring back the balancing equilibrium over the knife edge (Fig. 4).

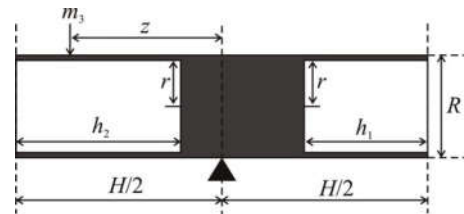


Figure 4: Moments about the geometric midpoint

Consider the cavity with length h_2 and radius r . The MBB has a length H and the square cross section has each side equal to R (Fig.4).

The moment of force acting on that half of MBB with the cavity of length h_2 about the midpoint is given by:

$$\left(\frac{H}{2} R^2 \rho \left(\frac{H}{4} \right) - \pi r^2 \rho h_2 \left[\frac{H}{2} - \frac{h_2}{2} \right] \right) g$$

Similarly the moment of force acting on that half of MBB with the cavity of length h_1 about the geometric midpoint is given by:

$$\left(\frac{H}{2} R^2 \rho \left(\frac{H}{4} \right) - \pi r^2 \rho h_1 \left[\frac{H}{2} - \frac{h_1}{2} \right] \right) g$$

Taking moment about the geometric mid point

$$\left(\frac{H}{2}R^2\rho\left(\frac{H}{4}\right) - \pi r^2\rho h_2\left[\frac{H}{2} - \frac{h_2}{2}\right]\right)g + m_3gz = \left(\frac{H}{2}R^2\rho\left(\frac{H}{4}\right) - \pi r^2\rho h_1\left[\frac{H}{2} - \frac{h_1}{2}\right]\right)g$$

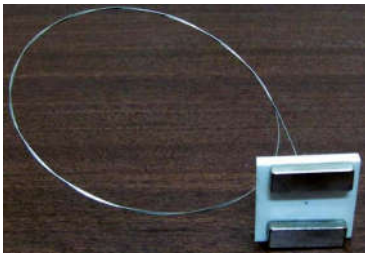
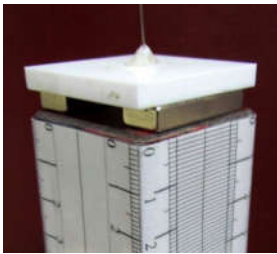




After rearranging the terms we get

$$m_3 = (h_2 - h_1) \left[\left(\frac{H}{2} - \frac{k}{2} \right) \frac{\pi r^2 \rho}{z} \right] \tag{4}$$

From the plot of m_3 against $\left[\left(\frac{H}{2} - \frac{k}{2} \right) \frac{\pi r^2 \rho}{z} \right]$ we get a straight line (Fig.8) and the slope of the graph is $(h_2 - h_1)$.

This slope obtained can be added with equation (2a) to get the values of h_1 and h_2 .

3. Description of apparatus

		
<p>Figure 5a: Steel wire with magnetic attachment</p>	<p>Figure 5b: Magnetic attachment placed symmetrically on MBB</p>	<p>Figure 5c: MBB clamped to the magnetic attachment</p>
		
<p>Figure 5d: Iron cylinder attached to the suspension wire</p>	<p>Figure 5e: Knife edge with support</p>	<p>Figure 5f: Magnetic stand on the MBB</p>

The suspension wire is already attached to the square acrylic piece (9.0 g). Two strong bar magnets are attached to this acrylic piece (Fig.5a). Each of the magnets weighs 11.1 g. The mechanical black box should be attached to this attachment (Fig. 5b and 5c). The iron cylinder also needs to be placed at the centre of the two magnets(Fig.5d). The iron cylinder with rods

protruding from its ends and slotted weights slid along them can be used to get values of periodic time of oscillation corresponding to various values of moment of inertia of the system and for drawing the corresponding graph. The two 50.0 g slotted weights are placed at suitable distances. Markings have been made on the rod to adjust the positions of these weights. The knife edge

mounted on a supporting plate has been used to balance the MBB (Fig.5e). Magnetic stand (a small magnet inserted at the bottom of the stand) could be placed upright on the MBB (Fig.5f). Slotted weights of 50.0 g and ring weights

of 5.0 g can be slid along the rod of the stand. The weights used for bringing in the equilibrium could be measured with the help of a digital weighing machine.

4. Experiment

A] Determination of I_{MBB}

The MBB is clamped to the magnetic attachment (Fig.5c) and it is subjected to torsional oscillations.

$$T_{MBB} = 3.820 \text{ s}$$

Then the MBB is removed and the iron cylinder is attached to the magnetic attachment (Fig.5d). Two 50.0 g slotted weights are used for suitably changing the moment of inertia of the assembly (Fig.6). These masses are treated as point masses for the calculation. The calibration graph (Fig.7) is

obtained by suitably changing the moment of inertia of the assembly (table 1).



Figure 6: Assembly for obtaining calibration graph

$$I = I_{att} + I_{rod} + I_{cyl} + 2I_{disc} + 2m_{disc}x^2$$

where, I is the moment of inertia of the assembly, I_{att} is the moment of inertia of the attachment which is negligible, I_{rod} is the moment of inertia of the rod, I_{cyl} is the moment of inertia of the cylinder, I_{disc} is the moment of inertia of the disc (50.0 g slotted weight).

d/cm	x/cm	$2mx^2 / g \cdot cm^2$	Time period for 5 oscillations				T/s	T^2/s^2
			t_1/s	t_2/s	t_3/s	t/s		
Without disc		0	12.40	12.38	12.32	12.37	2.473	6.117
0.0	2.925	855.6	15.16	15.16	15.06	15.13	3.025	9.153
1.0	3.925	1540.6	16.59	16.68	16.57	16.61	3.323	11.04
2.0	4.925	2425.6	18.41	18.44	18.49	18.45	3.689	13.61
3.0	5.925	3510.6	20.31	20.35	20.34	20.33	4.067	16.54
4.0	6.925	4795.6	22.34	22.40	22.37	22.37	4.474	20.02
5.0	7.925	6280.6	24.41	24.47	24.52	24.47	4.893	23.95
6.0	8.925	7965.6	26.88	26.91	26.88	26.89	5.378	28.92

Table 1: Observations for the periodic time T for oscillating system

$$T^2 = \frac{4\pi^2}{c} [I_{rod} + I_{cyl} + 2I_{disc} + 2m_{disc}x^2]$$

where, c is torsion constant.

$$T^2 = \frac{4\pi^2}{c} [2m_{disc}x^2] + \frac{4\pi^2}{c} [I_{rod} + I_{cyl} + 2I_{disc}]$$

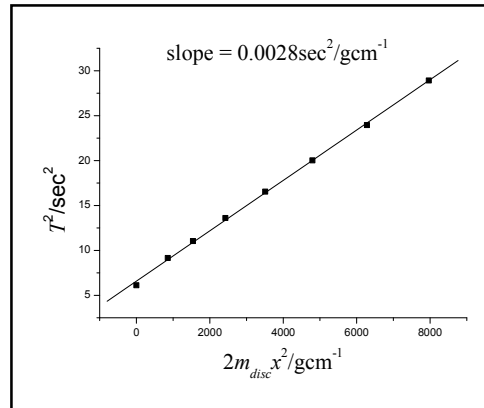


Figure 7: graph of T^2 vs the moment of inertia of different combinations

From the slope of the graph, torsion constant c is obtained. I_{MBB} is obtained by using this value of c .

$$\frac{4\pi^2}{c} = 0.0028$$

$$c = 1.409 \times 10^4 \text{ dyne} \cdot \text{cm}$$

$$\therefore I_{MBB} = \frac{T_{MBB}^2}{0.0028} = \frac{14.59}{0.0028} = 5210.7 \text{ g} \cdot \text{cm}^2$$

B] Calculation of the radius of the cavity

Using the value of I_{MBB} and c , the radius of the cavity, r is determined.

a] Calculation of removed mass due to cavities:

Removed mass forming cavity 1: $m_1 = \pi r^2 h_1 \rho$

Removed mass forming cavity 2: $m_2 = \pi r^2 h_2 \rho$

$$m_1 + m_2 = m = \pi r^2 h_1 \rho + \pi r^2 h_2 \rho$$

The MBB has a length $H = 24.0$ cm and width $R = 3.8$ cm

Volume of the MBB (had it been completely solid) can be given as:

$$V_s = R^2 \times H = (3.8)^2 \times 24.0 = 346.6 \text{ cm}^3$$

Density of Material of MBB is: $\rho = 7.9 \text{ g} \cdot \text{cm}^{-3}$

Hence, mass of MBB had it been completely solid would be,

$$M_s = \rho V_s = 7.9 \times 346.6 = 2737.8 \text{ g}$$

Mass of the MBB (given): $M_{MBB} = 1639.0$ g

Removed mass:

$$m = M_s - M_{MBB} = 2737.8 - 1639.0 = 1098.8 \text{ g}$$

b) Calculation of radius:

The moment of inertia of MBB (had it been completely solid) would have been given by:

$$I_s = \frac{M_s}{12} [R^2 + R^2] = \frac{2737.8 \times [3.8^2 + 3.8^2]}{12}$$

$$I_s = 6589.0 \text{ g} \cdot \text{cm}^2$$

Moment of inertia of the removed mass can be obtained:

$$I_{RM} = I_s - I_{MBB}$$

$$I_{RM} = 6589.0 - 5210.7 = 1378.3 \text{ g} \cdot \text{cm}^2$$

But, $I_{RM} = \frac{mr^2}{2}$

C) Determination of the lengths of the cavities, h_1 and h_2

The MBB is balanced by keeping its geometrical midpoint on the knife edge and adjusting the magnitude of mass m_3 at suitable points along the MBB (Fig.4 and 5f). The mass m_3 is placed on the magnetic stand that can stand erect on the MBB because of the magnet at the bottom. The stand is

$$\therefore r^2 = \frac{2 \times I_{RM}}{m} = \frac{2 \times 1378.3}{1098.8} = 2.509 \text{ cm}^2$$

$$r = \underline{\underline{1.584 \text{ cm}}}$$

[Actual value $r = 1.6 \text{ cm}$]

$$\therefore k = \frac{m}{\pi r^2 \rho} = \frac{1098.8}{3.14 \times 7.9 \times r^2} = \frac{44.30}{r^2}$$

Substituting the value of r in equation (2a)

$$k = \frac{44.30}{(1.584)^2} = 17.66 \text{ cm}$$

$$\therefore h_1 + h_2 = 17.66 \text{ cm} \quad (5)$$

shifted and the z changed and the m_3 is accordingly chosen. All the values of m_3 are measured by the digital weighing machine (Table 2).

Moments about the midpoint: (Midpoint of MBB = 12.0 cm).

z / cm	$1/z / \text{cm}^{-1}$	$\left[\left(\frac{H}{2} - \frac{k}{2} \right) \frac{\pi r^2 \rho}{z} \right] / \text{g} \cdot \text{cm}^{-1}$	m_3 / g
11.5	0.0870	17.165	104.5
11.0	0.0909	17.935	113.6
10.5	0.0952	18.783	119.2
10.0	0.1000	19.730	130.1
9.5	0.1053	20.776	136.0
9.0	0.1111	21.920	147.2
8.5	0.1176	23.202	157.6
8.0	0.1250	24.662	172.4
7.5	0.1333	26.300	182.7
7.0	0.1429	28.194	202.6

Table 2: Observations of mass m_3 for its different positions, z , from the midpoint

Graph: m_3 vs $\left[\left(\frac{H}{2} - \frac{k}{2} \right) \frac{\pi r^2 \rho}{z} \right]$

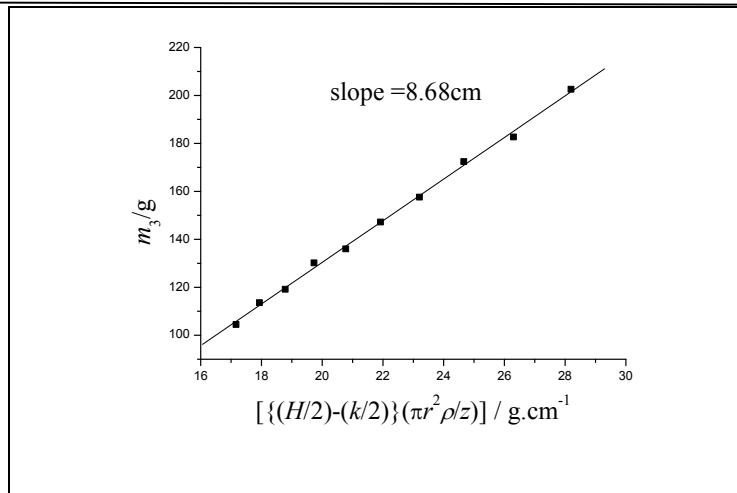


Figure 8: Graph for the determination of $(h_2 - h_1)$

Hence, using equation (4), the slope of the graph(Fig.8) can now be written as,

$$\therefore h_2 - h_1 = 8.68 \text{ cm} \quad (6)$$

Adding equations (5) and (6) we get $2h_2 = 26.34$ cm,

$$h_2 = 13.2 \pm 1.2 \text{ cm}$$

Subtracting equation (6) from (5) we get $2h_1 = 8.98$ cm,

$$h_1 = 4.5 \pm 1.2 \text{ cm}$$

Actual values of h_1 is 5.0 cm and that of h_2 is 13.0 cm.

5. Discussion and Conclusion

a. Possible variations of the problem:

The value of the density of the metal from which the MBB has been made (iron in this case) is very critical for the calculations. A small change in the actual value creates a huge deviation in the experimental values of the unknown quantities. So we decided to supply the density of the materials to the students while they performed the experiment. However the determination of density from a sample material by one of the many well-known and reasonably reliable methods could have been the part of the experiment.

This MBB had three unknown parameters in the form of the radius r of the cylindrical cavities and the two heights (h_1, h_2) of the two cavities. This experiment involves most of the measurements normally required in a mechanics experiment. The MBB was not an expensive one as it was made of iron but it was indeed a bit heavy and one needs to be careful while handling the experimental MBB. While designing the MBB certain aspects are necessary to look after. For example, the removed mass in this case was actually about 40% of the solid piece of metal. The diameter of the cavities was more than 80% of the side of the square cross sections. All these produce a perceptible change in

the moment of inertia of the MBB compared to the solid metal piece of similar external dimensions. This is necessary as we need to have a distinct difference in the time period of torsional oscillation of the MBB and the corresponding solid mass for reliable calculations.

In principle by keeping the number of unknown parameters to three one can have other designs of similar MBBs.

1. Suppose the MBB is constructed with one single co axial cavity that does not start from the one end of the MBB but a few centimeters inside from either of the ends of the MBB then similar experiment permits us to find out three unknowns in the form of r , h and x where x is the distance of the cavity from one end of the MBB. This experiment not only gives us the dimension of the coaxial cylindrical cavity but its location inside the MBB as well.
2. In another design the MBB may contain one single drill starting from one of its ends and the drill may be stuffed with some suitable solid material of uniform density. In this problem the first two unknowns are radius r and length h of the cylinder. The third unknown is the density of the material that has completely filled up the cavity. By performing the experiments described here and by forming three equations the three unknowns can be found out. However the stuffing in of the cavity with another material brings down the value of the amount of reduced mass. To have a significant change in the moment of inertia, proper size of the cavity and material of suitable density need to be chosen.

b. *Importance of the black box experiment in student learning:*

The black box experiments are open-ended experiments in which students are not aware of end results. Moreover such experiments go beyond mere verification of a principle or determination of a constant.

Students are provided with measured and relevant information and hints with respect to the parts of the black box and the testing to be done. They are rigorously tested on their present content knowledge and the skill to apply that knowledge in the given situation to solve the problem. Many times the content knowledge required is quite elementary. For example in a simplest electrical black box, students are supposed to take observations and analyze it with their understanding of I-V characteristics of the resistors, diodes and their combinations. For an optical black box, they are expected to know how a lens, a grating or similar optical elements respond to the given source. Similarly mechanical black box experiments test their basic understanding of the problem through the equations that they develop on centre of mass, oscillatory motions, etc.

Apart from the formulation of the equations, the most important aspect which can be tested is the students' confidence in their own observations. Another is the skill of making meticulous observations at all required possibilities which indicates inquisitiveness on the students' part. Once the data is collected, critically looking through the data and logically arriving at the required solution is equally significant.

The mechanical black box presented here can be modified to different difficulty levels. In order to guide students through different stages, we have given students many hints at different parts of the problem. If these hints are removed then it can become a real challenge to arrive at the values of the cavities. In this experiment, a student arrives at an equation which looks complicated. The

challenge it to linearize the equation which will enable the students to plot suitable graphs and obtain desired quantities either from slope or from the intercept.

This experiment also hints the student to plot a calibration graph between moment of inertia and period of oscillation. This is an indirect method of getting moment of inertia of irregular object about an axis. Through linearization the students should

know that it is the term $m_{\text{disc}}x^2$ only which is important in getting the calibration equation and that the calculation of moment of inertia of the other objects like central cylinder, connecting rod, etc is not required.

Though such kind of testing makes black box experiment a favourite choice for Olympiad examinations, the scope of these experiments can be extended to every undergraduate level.

Acknowledgement

The authors wish to thank Prof. D.A. Desai and Prof. R. M. Dharkar for fruitful suggestions. We would also like to thank the Indian Physics Olympiad students who patiently performed the

experiment. Authors would also like to express the Physics Olympiad programme for funding this activity.

References

- [1] Experimental Question 2004 International Physics Olympiad 2004, Pohang, South Korea <http://ipho.phy.ntnu.edu.tw/problems-and-solutions.html#2004>
- [2] Experimental Question 2011 International Physics Olympiad 2011, Bangkok, Thailand <http://ipho.phy.ntnu.edu.tw/problems-and-solutions 5.html#2011>

- [3] Worsnop B L and Flint H T 1927 Advanced Practical Physics for Students (London: Methuen) p 61
- [4] Chakrabarti B., Pathare, S. Huli S., Nachane M. (2013), Experimental determination of unknown masses and their positions in a mechanical black box, Physics Education, **48** (4), 477-483.

Optical Imaging of Metallic and Semiconductor Nanostructures at Sub-Wavelength Regime

A. K. Sivadasan¹, Kishore K. Madapu¹ and Prajit Dhara²

¹Nanomaterials Characterization and Sensors Section, Surface and Nanoscience Division, Indira Gandhi Centre for Atomic Research, Homi Bhabha National Institute, Kalpakkam-603102, India

²1st Year Student, Dept. of Electrical and Electronics Engineering Birla Institute of Technology and Science-Pilani, Pilani Campus Pilani- 333031, India

(Submitted: 17-02-2017)

Abstract

The near-field scanning optical microscopy (NSOM) is not only a tool for imaging of objects in the sub-wavelength limit but also a prominent characteristic tool for understanding the intrinsic properties of the nanostructures. The effect of strong localized surface plasmon resonance absorption of excitation laser in the NSOM images for Au nanoparticles is observed. The role of electronic transitions from different native defect related energy states of AlGaIn are also discussed in understanding the NSOM images for the semiconductor nanowire.

1. Introduction

The study of light-matter interaction in the near-field regime at the vicinity of nanostructures is a very interesting as well as challenging task for the scientific community. Abbe's diffraction limit prevents conventional optical microscopes to possess a spatial resolution beyond the value of $\sim\lambda/2$ (sub-wavelength limit), where λ is the wavelength of excitation light with a maximum numerical aperture value of unity for the probing objective. Thus, even for the visible light of $\lambda=400$ nm cannot image nanostructures of size below 200 nm. The near-field scanning optical microscopy (NSOM) assisted with the help of plasmonics is a unique tool to understand the light-matter interaction in the near field regime for optical imaging of nanostructures in the sub-wavelength limit. The light passing through the metal coated tip of NSOM probe with a circular aperture of diameter around few nanometers at the apex is capable of

surpassing the diffraction limit [1]. In the near-field regime, the evanescent field emitting from the

NSOM probe is not diffraction limited. Hence, it facilitates optical and spectroscopic imaging of objects with nanometer level spatial resolution. Light-matter interactions in metallic nanostructures have opened to a new branch of surface plasmon (SP) based photonics, known as plasmonics. The SPs are originated due to the collective oscillation of the free electrons about the fixed positive charge centers in the surface of metal nanostructures with a frequency of the oscillation of electrons, also known as plasma frequency, $\omega_p = (n_e e^2 / m_{eff} \epsilon_0)^{1/2}$, where, n_e is the density, m_{eff} is the effective mass, e is the charge of an electron and ϵ_0 is the permittivity of free space [1-3]. The coupling of the incident electromagnetic waves to the coherent oscillation of free-electron plasma near the metal surface is known as a surface plasmon polariton (SPP) and it is a propagating

surface wave at the continuous metal–dielectric interface. The electromagnetic field perpendicular to the metal surface decays exponentially and is known as evanescent wave providing sub-wavelength confinement near to the metal surface. Matching of the incident excitation frequency (ω) of electromagnetic wave with the plasmon frequency (ω_p) of the electrons in metal nanostructures, leads to an enhanced and spatially localized light–matter interaction, known as surface plasmon resonance (SPR) [1-3].

AlGaN is an intrinsically *n*-type semiconductor and one of the most prominent candidates among the group III nitride community with wide, direct and tunable band gap from 3.4 to 6.2 eV. Therefore, the group III nitrides including the ternary alloy of AlGaN nanostructures find tremendous applications in short wavelength and high frequency optoelectronic devices including light emitting diodes, displays and optical communications [4]. Consequently, by considering the above mentioned importance of AlGaN nanostructures in optoelectronic applications as well as semiconductor industries, it is also very important to understand the interaction of AlGaN nanowire (NW) with visible light.

In the present report, we have investigated the light–matter interaction of metallic Au nanoparticle (NP) catalysts (diameter \sim 50–150 nm) along with semiconductor AlGaN NW (diameter \sim 120 nm) grown *via* vapor liquid solid (VLS) mechanism in the near–field regime by using NSOM technique with external laser excitation of 532 nm (2.33 eV). The variations in contrast and absorption phenomena observed in the NSOM images of Au NPs are understood by considering the plasmonic effects of metallic nanostructures. In order to understand the light–matter interaction of AlGaN NW, we invoked the different energy states related to native defects originating due to the unavoidable incorporation of C and O in the material during the growth process.

2. Experimental section

The semiconductor AlGaN NWs along with Au NPs were synthesized using chemical vapor deposition technique *via* VLS growth mechanism. The detailed synthesis and basic characterizations of the sample is available in one of our earlier reports [5].

2.1 Atomic force microscopy

The atomic force microscopy (AFM) is one of the types of scanning probe microscopic (SPM) system used for the study of topography related information of a sample with an order of atomic scale spatial resolution. The AFM probes consist with a sharp tip of the order of 100 Å used for probing the tip–sample interactions [6]. There are several interactions possible to contribute the deflection/natural frequency of an AFM cantilever. The common force associated with AFM interaction is inter atomic force called as the van der Waals force and it varies with distance between the tip and the sample (FIG. 1(a)). The three modes of operations, available in the AFM setup (contact, non–contact and intermittent), can be selected by choosing the three different regimes of forces between tip and sample [6]. The major components commonly involved in the SPM systems are shown (schematic, FIG. 1(b)). The SPM system used in the present study works based on tuning fork feedback mechanism. The change in natural frequency of the tuning fork with respect to the tip/sample interactions are considered as a feedback parameter to measure the tip/sample force to map the surface modulation or topography. Thus, the frequency of the tuning fork changes, as the surface morphology changes with the variation of force felt, providing an image of the surface.

The AFM used for the studies on AlGaN NWs, reported in the present studies, is from a SPM system (Nanonics, MultiView 4000; Multiprobe imaging system). The MultiView 4000 uses normal force tuning fork technology with a high *Q*-factor and phase feedback to allow the control of probe/sample separation. Tuning forks in normal force mode with phase or amplitude feedback

permit high performance and ease of operation for AFM imaging in intermittent mode.

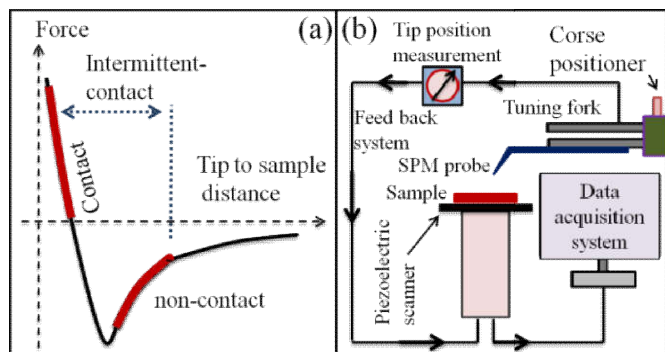


FIG. 1: (a) Interaction force Vs distance between tip and sample. (b) Block diagram of general SPM system.

The AFM tips are glass-based probes. Unlike standard piezo scanners that keep probes separated, the 3D flat scanner with excellent AFM resolution, large vertical (axial) displacement of up to 100 μm for sample scanning and up to 30 μm for tip scanning, is used in this system. The atomic force microscopic (AFM) images are recorded by using a 20 nm tip configured with a tuning fork feedback mechanism (MultiView 4000; Nanonics, Israel).

2.2 Near-field scanning optical microscopy

The NSOM is a microscopic technique used to investigate the light-matter interaction of nanostructures in the sub diffraction regime by using the advantages of evanescent waves (or confined light) which surpass the conventional far-field resolution limit. The generation of evanescent waves can be achieved with the help of either by plasmonics or by the use of sub-wavelength apertures coated with noble metals such as Au, Ag or combination of both. The evanescent waves emanated from the apertures/probe with higher momenta *i.e.*, lower wave lengths and velocities compared to that of normal light can be used for achieving the high resolution by placing the detector very close (near-field regime, smaller than wavelength λ) to the sample specimen surface. This allows us to record the light-matter interactions with high spatial, spectral and

temporal resolution [1-3]. In this technique, the resolution of the image is determined by the size and geometry of the aperture probe and not by the λ of the excitation light. The NSOM provides simultaneous measurements of the topography and optoelectronic properties of nanostructures with high spatial resolution in the sub-wavelength regime. The detailed schematic of experimental set up used for recording the light-matter interaction of AlGaIn single NW is shown (FIG. 2). The NSOM imaging of nanostructures was used to understand the interaction with 532 nm laser (~ 2.33 eV).

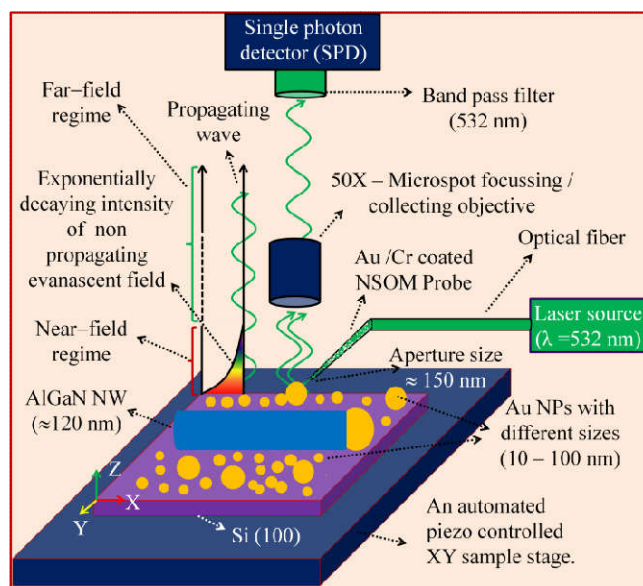


FIG. 2: The schematic experimental setup for NSOM imaging of AlGaIn NW and Au NPs.

In order to perform an NSOM experiment, with near-field excitation and far-field collection configuration (FIG. 2), a point light source emanated through the probe at the near-field regime was scanned over the surface of the sample specimen with tuning fork feedback and the propagating optical signal emitted from the sample surface due to the dipole radiation was detected at the far-field. A band pass filter (532 nm) was employed to extract the excitation laser after the light-matter interaction and before reaching the light to the single photon detector (SPD) in the far-field configuration. We used an optical fiber with a circular aperture and metal (Au/Cr) coated probe with a tip apex (aperture) diameter of 150 nm for near-field excitation of laser light. The optical fiber

coated with Cr (buffer layer) and Au (over layer) to avoid leakage of optical power, was used to enhance the optical transmission and confine the light to the sample surface. The experimental setup (SPM with coupled Raman spectroscopy system) involved in the AFM/NSOM imaging system is shown in the figure 3.

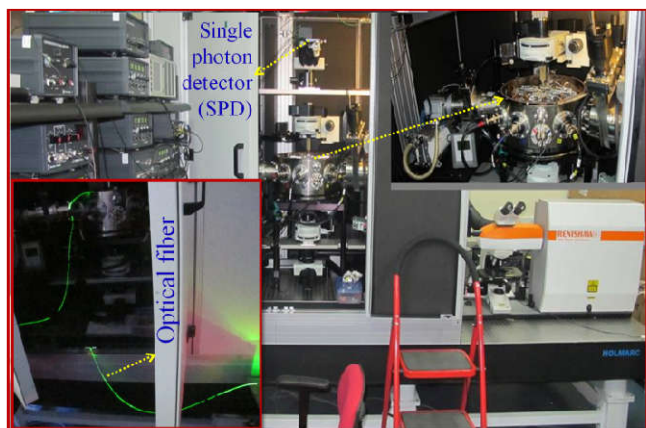


FIG. 3: The experimental set up; SPM (MultiView 4000; Nanonics, Israel) coupled Raman spectroscopy (Renishaw, UK) system, used for the AFM/NSOM imaging of the sample. The inset shows an optical fiber carrying laser light of 532 nm (2.33 eV) to NSOM probe with an aperture of 150 nm.

The same instrument with NSOM configuration is used to understand the near-field light-matter interactions of nanostructures with visible laser 532 nm (~ 2.33 eV). The scanning was performed either using the translational movement of NSOM probe (description for FIG. 2) or by motorized XY sample stage with very precise spatial resolution controlled by inbuilt sensors and piezo-drivers. A band pass filter (532 nm) was used to extract the scattered light, post light-matter interaction, before entering the SPD in the far field configuration. The same probe was used as an AFM tip for simultaneous scanning of the topography (description for FIG. 1(b)), along with the NSOM image of the sample with the tuning fork feedback mechanism.

3. Results and discussions

The morphological shape, size and distribution of mono-dispersed AlGaIn NWs are shown in the AFM topographic image (FIG. 4(a)). The high resolution AFM image shows cylindrical shaped NWs with very smooth surface morphology along with Au NP catalyst at the tip (FIG. 4(b)).

Uniformly sized and mono-dispersed NWs with average diameter of 120 nm were observed. The well separated Au NPs, which participated in the VLS growth process of the NWs, were uniformly sized (~ 150 nm) (FIG. 4(c)). The Au NPs with diameters around 50–100 nm, not participating in the growth process, were also found to be distributed uniformly over the substrate (FIG. 4(a)).

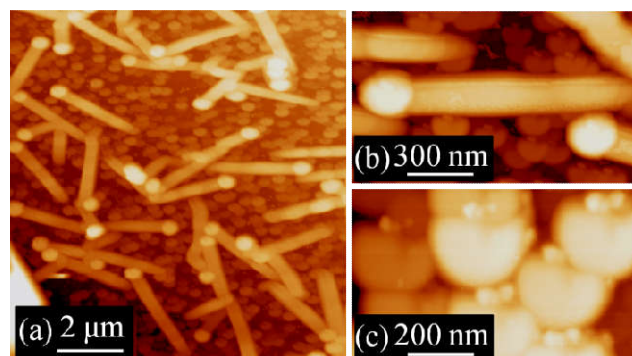


FIG. 4: AFM topographic images (a) mono-dispersed AlGaIn nanowires and high resolution image of (b) single AlGaIn nanowire (c) Au nanoparticles

Since the diameter of Au NPs (~ 100 nm) as well as AlGaIn NW (~ 120 nm) is far below the diffraction limit for the excitation wavelength (532 nm), one needs to shorten the wavelength down to the sub-diffraction regime to obtain highly resolved optical images. Using metal coated NSOM probe, it is possible to produce evanescent waves with momentum higher than that of the original excitation wavelength $\lambda_0 = 2\pi/k_0(\omega)$ with wave vector of $k_0(\omega) = \omega/c$, where c is the velocity of light. Therefore, the evanescent waves emanating from the NSOM probe aperture possess group of wave vectors higher than the original excitation laser as $k_{ev}(\omega) = \omega/v$, with different velocities ($v < c$) slower than the excitation wave velocity ($v < c$) [1-3]. Therefore, the NSOM measurements are advantageous in providing the super-resolution along with localization of intense electric fields. At the same time, it conserve the excitation energy and hence the frequency. Thus, it offers the possibility of optical as well as spectroscopic imaging in the sub-wavelength regime. So, from the scanned images using NSOM technique, we can understand even the intrinsic properties of a sample as revealed by its electronic or vibronic characteristics of the material.

The observed NSOM images (FIG. 5) of Au NPs shows a strong SPR related absorption in two dimension (2D) and 3D. The high resolution topographic AFM image of the Au NPs shows (2D and 3D images in FIGs. 5(a) and 5(b), respectively) smooth and spherical shape of the NP with a diameter of ~ 100 nm.

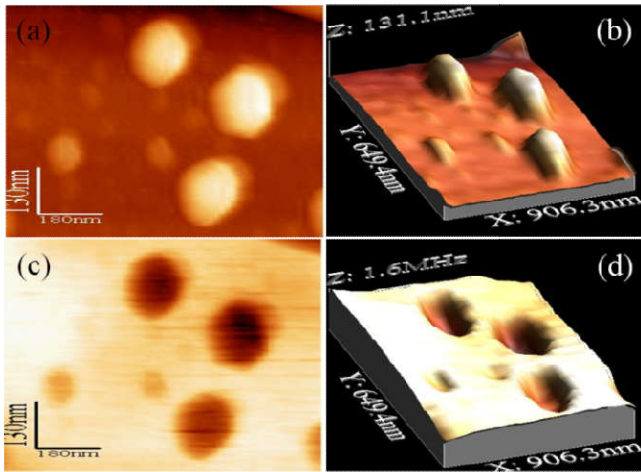


FIG. 5: AFM topographic images of Au nanoparticles in (a) 2D and (b) 3D. The NSOM images of Au nanoparticles in (a) 2D and (b) 3D.

The NSOM image of Au NPs shows (2D and 3D images in FIGs. 5(c) and 5(d), respectively) a strong absorption of electromagnetic waves. The significant absorption of light with wavelength 532 nm by Au NP is because of the fact that, the SPR peak value ~ 540 nm for Au NPs matches the excitation wavelength. At resonance, the incident electromagnetic waves coupled with collective oscillation of electrons, can produce SPPs which are perpendicular to the surface of the Au NP. The frequency dependent wave vector of SPP can be expressed in terms of frequency dependent dielectric constants of metal ($\epsilon_m = \epsilon_m' + i\epsilon_m''$) and surrounding dielectric material ($\epsilon_d = 1$, for air or vacuum), $k_{spp}(\omega) = \frac{\omega}{c} \sqrt{\frac{\epsilon_d \cdot \epsilon_m}{\epsilon_d + \epsilon_m}}$. Therefore, the effective

wavelength of the SPP is $\lambda_{spp} = 2\pi/k_{spp}$ [1-3]. The SPPs of different wavelengths, lower than the excitation, can propagate through the surface of Au NPs up to a propagation length which depends on the complex dielectric constants of the metal and dielectric medium [1-3]. Once the SPP propagates through the surface of Au NP and crosses the

metallic region, then the electromagnetic wave may decouple from the SPP and it can be converted to a propagating wave. The intensity of the absorption is influenced by the frequency dependent polarizability of the Au NPs and it can vary with respect to the size of the Au NPs. Thus, because of the variation of different sizes of the Au NPs, it is possible to observe them with relatively different absorption intensities (FIGs. 5(c) and 5(d)). Apart from the formation of SPP, some portion of the absorbed excitation laser may also participate in lattice phonon generations leading to heating as well as inter-band transitions of Au NPs [7].

The NSOM images obtained as a result of near-field light-matter interaction is shown for AlGaIn single NW along with Au NPs of various sizes (FIG. 6). The high resolution topographic AFM image of the single NW shows (2D and 3D images in FIGs. 6(c) and 6(d), respectively) smooth and cylindrical shape, as observed in the FESEM images. The NSOM images of AlGaIn single NW as well as catalyst Au NPs are also observed (2D and 3D images in FIGs. 6(c) and 6(d), respectively). The reported room temperature band gap of our AlGaIn NWs is 3.55 eV [5], which is higher than the excitation energy of 2.33 eV. Therefore, a complete transmission of light through the AlGaIn NW is expected. Surprisingly, we observed a prominent absorption of light along the AlGaIn NW, as shown in the NSOM images (2D and 3D images in FIGs. 6(c) and 6(d), respectively).

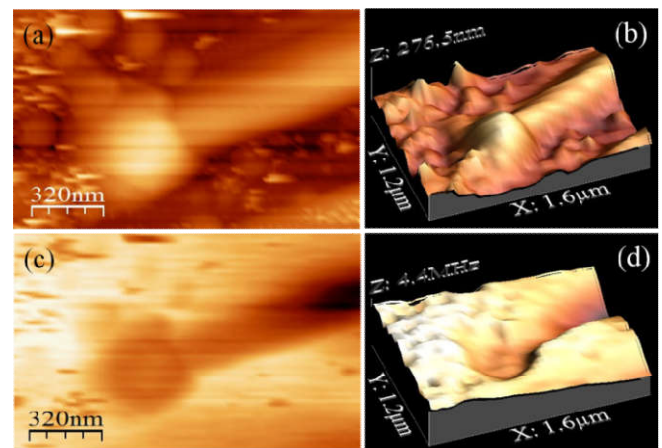


FIG. 6: AFM topographic images of mono-dispersed AlGaIn nanowire in (a) 2D and (b) 3D. The NSOM images of AlGaIn nanowire in (a) 2D and (b) 3D.

This absorption of light by the semiconductor AlGaIn NW is observed because of the presence of native defects originating due to the unavoidable incorporation of C and O in the material, which may create energy levels below 2.33 eV.

In conclusion, we envisage the use of near field scanning optical microscopy (NSOM) technique for direct understanding of light-matter interaction of metallic as well as semiconductor nanostructures of sub-wavelength limited dimension in the near-field regime. The NSOM images of metallic Au nanoparticles with diameter ~ 100 nm shows a strong surface plasmon resonance related absorption

References:

- [1] L. Novotny and B. Hecht. *Principles of Nano-Optics*. 1st ed. Cambridge: Cambridge University Press, Cambridge Books Online (2006)
- [2] S. Kawata, Y. Inouye and P. Verma, *Nature Photonics*, **3**, 388 (2009)
- [3] W. L. Barnes, A. Dereux and T. W. Ebbesen *Nature*, **424**, 824 (2003)
- [4] S. Zhao, H. P. Nguyen, M. G. Kibria and Z. Mi, *Prog. Quant. Electron.* **44**, 14 (2015)

of excitation laser with an energy of 2.33 eV (532 nm) due to the formation of surface plasmon polaritons as well as the localized surface plasmon resonance near to the surface of the Au nanoparticles. The isolated single semiconductor AlGaIn nanowire with a diameter ~ 120 nm shows a strong absorption of visible light due to the electronic transitions originated from the native defect related energy levels.

Acknowledgements

We thank, S. Dhara of SND, IGCAR for his valuable guidance and helpful discussions.

- [5] A. K. Sivadasan, A. Patsha, S. Polaki, S. Amirthapandian, S. Dhara, A. Bhattacharya, B. K. Panigrahi and A. K. Tyagi, *Cryst. Growth Des.* **15**, 1311(2015)
- [6] R. Howland, L. Benatar, *A Practical Guide: To Scanning Probe Microscopy*, DIANE Publishing Company (1998)
- [7] M. R. Beversluis, A. Bouhelier and L. Novotny, *Phys. Rev. B.* **68**, 115433 (2003)

Electrical Circuits as the Nerve Cells

M. R. KHOSHBIN-E-KHOSHNAZAR

*Physics Department, Curriculum Development Center,
Organization for Research and Educational Planning,
P.O.Box 15855-363, Tehran, Iran*

E.mail: *khoshbin@talif.sch.ir*

(Submitted: 02-07 - 2016)

Abstract :

The cell for transmitting and processing in the nerve systems and brains is nerve cell(neuron).Nerve cells conduct the necessary currents via the movement of ions. They communicate via electrochemical waves called action potentials. All of the important functional properties of the nerve cells can be represented by electrical circuits. This is very interesting subject and should be considered in introductory Physics and biology textbooks. In this article we review on this topic

1. The Resting Membrane of the Neuron

Nerve cells via nerve fibers transmit electrical impulses to the muscles or to and from the brain. Nerve cells at rest are permeable to Na^+ and Cl^- ions in addition to K^+ ions. How can the concentration gradients for these three ions be maintained permanently across the membrane of a single cell, and how do these three gradients interact to determine the cell's resting membrane potential? To answer these questions, we first examine only diffusion of

K^+ . Let us consider a cell having only K^+ channels with concentration gradients for the three ions. Under these conditions the resting membrane potential is determined solely by the K^+ concentration gradient and will be calculated by

Nernst Equation. What is Nernst Equation? Simply put, as K^+ ions are present at a high concentration inside the cell, and Na^+ and Cl^- have a high concentration outside the

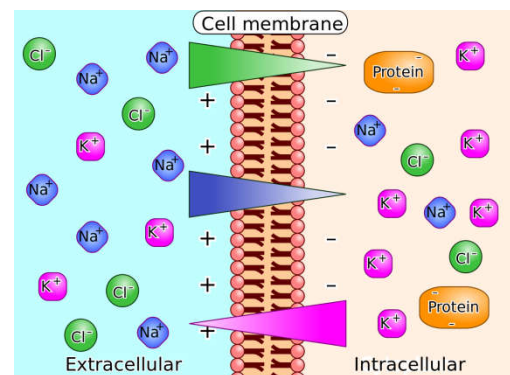


Figure.1 Sodium- potassium pump and diffusion(Adapted from nl.wikipedia.org.)

cell, K^+ ions tend to diffuse from inside to outside the cell, along their chemical concentration gradient. As a result, the outside of the membrane accumulates positive charges and the inside negative charges. The diffusion of K^+ gives rise to an electrical potential difference that causes a *self-limiting* separation of charges: positive outside, negative inside. This potential difference tends to oppose the further efflux of K^+ . Thus, ions are subject to two forces driving them across the membrane: First, a chemical driving force that depends on the concentration gradient across the membrane and secondly, an electrical driving force that depends on the electrical potential difference across the membrane. When the nerve cell is at rest, the electrical force driving K^+ into the cell exactly balances the chemical force driving K^+ ions out of the cell. The value of resting membrane potential in this case is called the potassium equilibrium potential, E_K . However, the equilibrium potential for any ion X can be calculated from an equation derived in 1888 from basic thermodynamic principles by the German physical chemist Walter Nernst:

$$E_x = (RT/ZF) \ln([X_o]/[X_i])$$

where R is the gas constant, T the temperature (in Kelvin), Z the valance of the ion, F the Faraday constant, and $[X_o]$ and $[X_i]$ are the concentrations of the ion outside and inside the cell, respectively. RT/F would be 25 mV at room temperature and the constant for converting natural logarithms to base 10 logarithms is 2.3. So, the Nernst equation can also be written as:

$$E_x \cong (58/Z) \log ([X_o]/[X_i])$$

Thus for K^+ , with $X_i = 400 \text{ mol/m}^3$ and $X_o = 20 \text{ mol/m}^3$, we obtain $E_K = -75.46 \text{ mV}$ and in a similar way

$$E_{Na} = 54.78 \text{ mV (with } X_i = 50 \text{ mol/m}^3, X_o = 440 \text{ mol/m}^3) \text{ and}$$

$$E_{Cl} = -59.87 \text{ mV (with } X_i = 52 \text{ mol/m}^3, X_o = 560 \text{ mol/m}^3).$$

2. Equivalent Circuit Model for Neuron's Resting Membrane Potential

An electrical circuit consisting only of conductors or resistors (the ion channels), batteries (the concentration gradients) and capacitors (the ability of the membrane to store charge) can explain the functional properties of a neuron [1]. In an equivalent circuit, K^+ channels can be represented as a resistor or conductor of ionic current with conductance of g_K . Then the gradients represented as the current through the K^+ channels would be given by Ohm's law: $i_K = g_K U$, where U is the membrane potential. Since there is a K^+ concentration gradient, there will be a chemical force driving K^+ across the membrane. In the equivalent circuit this chemical force is represented by a battery, whose electromotive force is given by the Nernst potential for K^+ . The K^+ current that flows solely because of its concentration gradient is given by $i_K = -g_K E_K$. The negative sign is required, because a negative equilibrium potential produces a positive current. Thus, for a real neuron that has both a membrane voltage and K^+ concentration gradient, the net K^+ current is given by the sum of the currents due to two electrical and chemical driving forces: $i_K = (g_K U) - (g_K E_K) = g_K (U - E_K)$. The term $U - E_K$ is called the electrochemical driving force that determines the direction of the ionic current.

If a few resting Na^+ channels are added to the membrane, it becomes slightly permeable to Na^+ . Two forces act on Na^+ to drive it into the cell. First, Na^+ is more concentrated outside than inside and therefore it tends to flow into the cell down its chemical concentration gradient. Second, Na^+ is

driven into the cell by the negative electrical potential difference across the membrane. The influx of positive charge (Na^+) *depolarizes* the cell, but only slightly from the K^+ equilibrium potential (≈ -75 mV). The new membrane potential does not come close to the Na^+ equilibrium potential of $\approx +55$ mV because there are many more resting K^+ channels than Na^+ channels in the membrane. As soon as the membrane potential begins to *depolarize* from the value of the K^+ equilibrium potential, K^+ flux is no longer in equilibrium across the membrane. The reduction in the negative electrical force driving K^+ into the cell means that there will be a net efflux of K^+ out of the cell, tending to counteract the Na^+ influx. So far we have ignored the contribution of Cl^- to the resting potential, even though many nerve cells have Cl^- channels that are open in the resting membrane. This simplification is valid for nerve cells that do not have a mechanism for active transport of Cl^-

against an electrochemical gradient. In these cells the resting potential is ultimately determined by K^+ and Na^+ fluxes, because the intracellular concentration of K^+ and Na^+ are fixed by the $\text{Na}^+ - \text{K}^+$ pump (active transport), where Cl^- concentration inside the cell is affected only by passive forces due to the electrical potential and concentration gradient. Thus, the movement of Cl^- ions tends toward equilibrium across the membrane, so that E_{Cl} is equal to the resting potential, and there is no net Cl^- flux at rest.

However, we shall initially ignore Cl^- channels and begin with just two types of channels, K^+ and Na^+ . Moreover, since the membrane potential is constant in the resting state, we have $dU/dt = 0$ and the net current must be equal to zero: $I_{\text{Na}} + I_{\text{K}} = 0$. In the other hand, the total potential difference is the sum of the potential differences across E 's across g 's:

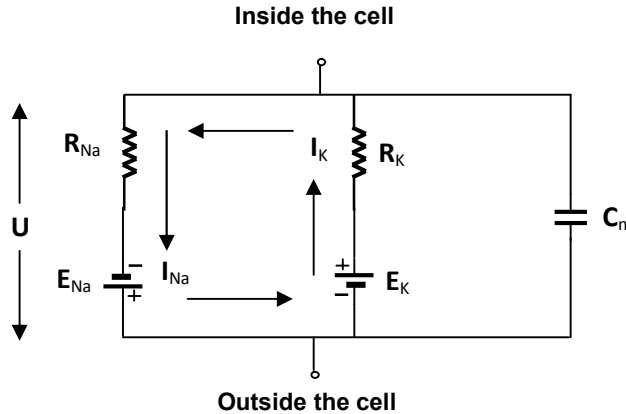


Figure.2 Electrical equivalent circuit which only includes the Na^+ and K^+ .

$$U = E_{\text{Na}} + I_{\text{Na}}/g_{\text{Na}} \quad , \quad U = E_{\text{K}} + I_{\text{K}}/g_{\text{K}} \quad ,$$

or

$$I_{\text{Na}} = g_{\text{Na}}(U - E_{\text{Na}}) \quad , \quad I_{\text{K}} = g_{\text{K}}(U - E_{\text{K}})$$

Now, by substituting I_{Na} , I_{K} in charge conservation: $I_{\text{Na}} + I_{\text{K}} = 0$ and solving for U , we have

$$U = [(E_{\text{Na}} g_{\text{Na}}) + (E_{\text{K}} g_{\text{K}})] / (g_{\text{Na}} + g_{\text{K}})$$

Using the values of $g_{Na} = .5 \times 10^{-6} S$, $E_{Na} = +54.78 mV$, $g_K = 10 \times 10^{-6} S$ and $E_K = -75.46 mV$, we conclude $U = -69.26 mV$.

We can derive a more general equation for U from an equivalent circuit that includes the Cl^- pathway with associated Nernst potential (battery):

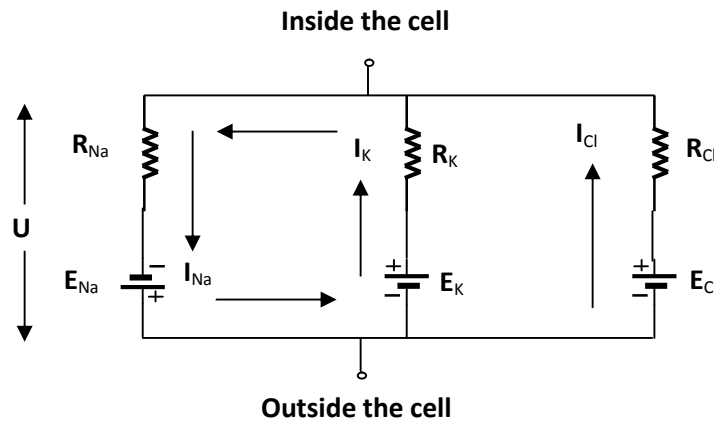


Figure 3. Electrical equivalent circuit which includes the Cl^- pathway and omits the capacitor.

$$U = [(E_{Na} g_{Na}) + (E_K g_K) + (E_{Cl} g_{Cl})] / (g_{Na} + g_K + g_{Cl})$$

$= 2.5 \times 10^{-6} S$, we obtain $U = -69.26 mV$ (rounded off to two places) which is only slightly more negative than the previous value.

Note, that since no net current flows through the Cl^- channels we have $E_{Cl} = -69.26 mV$ and by using g_{Cl}

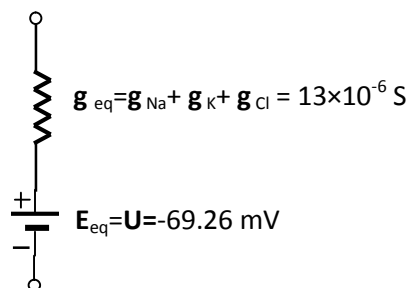


Figure 4. Simplified equivalent circuit of resting channels can be represented by a single equivalent conductance and a battery (no current flows through the Cl^-).

However, it has been shown that chloride is actively pumped out of some (but not all) cells. If we

assume this situation for all cells, then we should use Nernst Potential instead of E_{Cl} (i.e., $E_{Cl} = -59.87$ mV) and obtain $U = -67.54$ mV. Anyway, at rest, the membrane potential is close to Nernst potential for K^+ , the ion that the membrane is most permeable to.

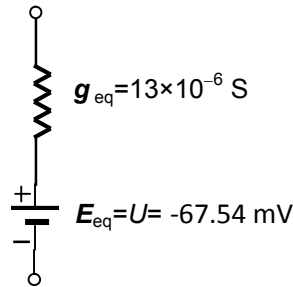


Figure 5. Simplified equivalent circuit, when current flows through the Cl^- .

3. Hodgkin - Huxley Model

As we showed in the previous section, we can represent neurons by RC circuits. As we know, a capacitor is short-circuits at $t = 0$ and open-circuit at $t = \infty$. Actually, $t = \infty$ relates to the membrane potential at rest. However, what happens in intermediate situations? Hodgkin and Huxley [2] had to change the concentration of sodium and potassium of the giant axon of the squid in voltage clamp experiments, in order to see how each affected the action potential [of course, in the 1960s, methods for blocking channels pharmacologically had been developed. Tetrodotoxin (TTX) blocks Na^+ channels and Tetraethylammonium (TEA) blocks K^+ channels].

Let us now translate their consideration into mathematical equations [3]. In these situations dU/dt is not equal to zero and capacitive current is $I_C = CdU/dt$. The conservation of electric charge on the membrane implies that input current $I(t)$ which injected into the cell splits into I_C and further components I_j which pass through the ion channels. Thus

$$I(t) = I_C(t) + \sum_j I_j(t), \quad \text{or}$$

$$C \frac{dU}{dt} = -\sum_j I_j(t) + I(t)$$

where the sum runs over all ion channels. The probability that a channel is open or closed is described by additional variables m , n and h . The K^+ gates are controlled by n (activation channel) and Na^+ gates are controlled by m (activation channel) and h (inactivation channel). Hodgkin and Huxley formulated the three current components as

$$\sum_j I_j(t) = m^3 h g_{Na}(U - E_{Na}) + n^4 g_K(U - E_K) + g_{Cl}(U - E_{Cl})$$

where g 's are maximum conductance when all channels are open. The three gating variables evolve according to the following differential equations

$$\frac{dm}{dt} = \alpha_m(U)(1 - m) - \beta_m(U)m$$

$$\frac{dn}{dt} = \alpha_n(U)(1 - n) - \beta_n(U)n$$

$$\frac{dh}{dt} = \alpha_h(U)(1 - h) - \beta_h(U)h$$

where the various functions of α and β are given, respectively, in Table 1.

Table 1. The parameters of the Hodgkin-Huxley equation.

x	$\alpha_x (U/mV)$	$\beta_x (U/mV)$
n	$(0.1-0.01U)/[\exp(1-0.1U) - 1]$	$0.125\exp(-U/80)$
m	$(2.5-0.1U)/[\exp(2.5-0.1U) - 1]$	$4\exp(-U/18)$
h	$0.07\exp(-U/20)$	$1/[\exp(3-0.1U)+1]$

As with good numerical problems, the dynamics of the Hodgkin-Huxley model can be studied for different types of inputs such as constant input, step input, sinusoidal input and most important and realistic situation: time-dependent input, in turn. In addition we can also study neuronal refractoriness via Hodgkin-Huxley model [3]. For example, at a certain time the model is stimulated by a short current pulse that is sufficiently strong to excite a spike. We'll find that a second current pulse of the *same* amplitude by certain interval time could not trigger a second action which is clear evidence of neuronal refractoriness.

Post Script

In this paper we've focused on classical view of neuron action mechanism. But, Stuart Hameroff and some other persons has thrown doubt upon it

References

1. Kandel, E.R., Schwartz, J.H., and Jessel, T.M. editors(2000).*Principles of Neural Science*. McGraw Hill, New York, fourth edition.

[4].Their research led them and other scientists into quantum realm [5] . Human thinking, as many of us know, often fails to respect the principles of classical logic [6].The classical approach is roughly speaking that neuron fires or it does not-comparable to a computer bit,1 and 0.However, it seems neurons are incredibly more complicated. A subject that could be left for possible future communications. Although it is very difficult to see how application of this formalism could produce anything practical in the current situation when we are still trying to understand consensus in ontological and epistemological issues, it will be worth the effort to explore such possible quantum mechanisms.

- Hodgkin, A.L. and Huxley, A.F. (1952). A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve, *Journal of Physiology*, 117, 500-544.
- Grestner, E.R. and Kistler W.M. (2002). *Spiking Neuron Models, Single Neurons, Populations, Plasticity*, Cambridge University Press, 1st edition.

-
4. See e.g. Hameroff S.R. and Penrose R.(2003). Conscious events as orchestrated space-time selections, *NeuroQuantology* 1(1) ,10-25. And
 5. Khoshbin-e-Khoshnazar M.R.(2007), Achilles' Heels of the Orch OR Model, *NeuroQuantology* 5(1),182-185.
 6. Khoshbin-e-Khoshnazar, M.R. (2014).Quantum superposition in the Retina, *NeuroQuantology* 12(1), 97-101.
 7. Khoshbin-e-Khoshnazar, M.R.(2016), Possible Application of Variational Cluster Expansion in Modeling Correlated Neurons, *Journal of Consciousness Exploration & Research* 7(1), 122-125.
 8. Buchman, M. Quantum Minds: Why we think like quarks, *Newscientists*, Issue 2828.
<https://www.newscientist.com/article/mg21128285.900-quantum-minds-why-we-think-like-quarks/>

