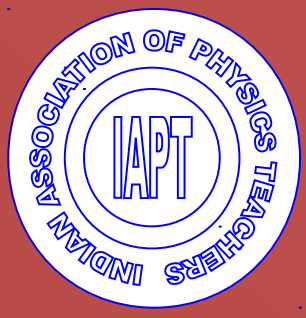
Vol 31 No. 4 Oct - Dec 2015 ISSN 0970-5953



PHYSICS EDUCATION

HAPPY NEW YEAR - 2016

www.physedu.in

Volume 31, Number 4

In this Issue

EDITORIAL

Farewell to Department of Physics

Savitribai Phule Pune University

This is the last issue of Physics Education getting published from Department of Physics, Savitribai Phule Pune University. Next issue onwards, Physics Education will be published from Indian Institute of Science Education and Research (IISER) Pune. This is also the last issue of which I am the Chief Editor. Dr. M. S. Santhanam (IISER Pune) will take over as the Chief Editor of Physics Education from the next issue. As IISER is the national initiative on science education, it is quite natural that IISER takes over the reins of Physics Education and develops it into a platform for new ideas and experiments in learning and teaching physics. Especially we may look forward to the promotion of multimedia inputs and also the integration of this journal into the national endeavor on science education. I am sure that Physics Education will continue to promote interests of undergraduate students and teachers in physics.

If may not be out of place here to put on record a brief history of this journal. Physics Education was started in 1975 as a part of the science education initiative by UGC along with three other journals, with similar names, dealing with Chemistry, Mathematics and Biology. Right from its inception, Physics Education was published from Department of Physics, erstwhile University of Pune. Unfortunately, other three journals started by UGC ceased to publish in a short while, but Physics Education continued as an independent international journal and is contributing in its own way to physics pedagogy. This would not have been possible without the devotion and the meaningful efforts by all the previous editors, Professors A.S. Nigavekar, A. W. Joshi, A. S. Parasnis, S. V. Lawande, P. V. Panat and R. Ramachandran, despite all odds. In the last five years we have made this journal totally web-based and have automated the processing of all the contributed articles, including their allocation to various editors, processing of referee reports, communication with authors and the preparation of each issue in its final form. Only the selection of referees is the prerogative of the editors. This prepares the journal for multimedia inputs and for initiating various forums for interactive contribution by student and other readers. In the last 4/5 years the journal has seen significant increase in contributions from Europe, middle-east, Africa, USA and far east.

Finally, I take this opportunity to express my gratitude towards Prof. R. Ramachandran for his active interest and valuable advice throughout my tenure and the distinguished panel of editors and all the referees who have contributed towards the quality of this journal. I also wish to thank my colleagues from the Physics of Department, Savitribai Phule Pune University for their help and cooperation.

I am sure that under the auspices of IISER, Physics Education will continue to scale new heights and become an instrument of transformation in the area of physics education.

Finally I wish you a very happy reading!

Pramod S. Joag.

Chief Editor, Physics Education
Chief-editor@physedu.in,
pramod@physics.unipune.ac.in

Functional differential equations. 5: Time travel and life

C. K. Raju

ckr@ckraju.net

(Submitted **24-12-2015**)

Abstract

Physics must be non-mechanistic to account for everyday experience. Existing physics becomes non-mechanistic if advanced interactions exist. Advanced interactions are usually eliminated on metaphysical grounds of "causality", but we explain why that is not valid. Admitting advanced interactions involves no hypothesis, but only an acceptance of the most general formulation of physics, using mixed-type FDEs. If advanced interactions are rare, the resulting physics remains approximately mechanistic. The mixed-type FDE model readily resolve various paradoxes of time travel. Specifically, time *machines* are impossible, since realistic time travel implies spontaneity (different from chance). The novel features of this model can be expected to be especially prominent at the microphysical level of biological macromolecules and single particles.

1 Introduction

The twentieth century was marked by two great revolutions in physics: relativity and quantum mechanics. It is common to glorify these revolutions, but our objective is to *understand* them. To that end, we need to ask: what *defects* in the old physics led to these revolutions?

1.1 Relativity and time measurement

The relativity revolution concerned time, as we saw in the previous article in this series.[1] There was a conceptual error about time measurement in Newtonian physics. It did not define equal intervals of time, and hence had no "correct" way to measure time. Consequently, even Newton's first law, by itself, is

not meaningful. Relativity corrected that error and provided a definition of equal intervals of time.

The advantage of trying to understand defects in the old physics is clear: it led to a new theory of gravitation—retarded gravitation theory—which uses FDEs, even in gravitation. This new theory suggested a new way to resolve the problem of galactic rotation curves without the need for either dark matter or *ad hoc* modifications to Newtonian physics. Newtonian gravity remains a first approximation limited to the solar system. But even within the solar system, tiny departures from it would be observable as in the NASA flyby anomaly attributable to the novel gravitational effects of the rotation of the earth.

1.2 FDEs and time asymmetry

However, there is a second fundamental difficulty with time in the old physics: the difficulty with time asymmetry or the "arrow" of time. Actually, however, the difficulty with time asymmetry is part of an even more fundamental difficulty.

Even a child can distinguish between living organism and non-living things, but the equations of the old physics make no such distinction. Hence, those equations ought to apply equally to both: living and non-living.[1] That is, physics *must* be compatible with biology, *and* mundane human experience. But is it?

[1]Indeed, since the old physics provides no way to separate living from non-living, it gives us no way even to articulate any claim that "physics does not apply to living organisms".

Irreversible aging is our most basic experience from birth to death. This observed irreversibility is *contrary* to the old physics, which was time reversible. If the equations of physics are written down using only ODEs and PDEs, the transformation $t \to -t$ does not change those equations. But, on mundane experience, it is impossible to reverse aging, and turn an old man into a baby! Science must be compatible with observation: if physics disagrees with widespread experience, we should correct physics, not dismiss the experience as an illusion.

Retarded FDEs partly correct this error in the old physics, for they model an irreversible physics. Recall that [2, 3] retarded FDEs arise in classical electrodynamics simply by doing the math correctly. That is, we (a) take into account the neglected *coupling* of ODEs and PDEs, required for the many body problem, and (b) use retarded propagators for the solution of PDEs (as in retarded Lienard-Wiechert potentials). Similarly, retarded FDEs are used in the new theory of gravitation explained in the previous article. But why use only retarded propagators?

1.3 Mundane time

Thus, an asymmetry between past and future only partly describes our mundane experience of time. For, on our mundane experience, repeated thousands of time each day, our actions often successfully *create* or "bring about" (a tiny part of) the future cosmos. All the plans we make for the future (including applications for research grants) are premised on this belief

that our actions *now* contribute in deciding the future outcome.

This human creativity (in general, the creativity of all living organisms) is possible only if future is not fully determined from past by physics. That is, apart from the two problems of time measurement and time asymmetry, there is a third problem about time in the old physics: physics is mechanistic and that needs to be corrected.

However, any sort of indeterminism is not acceptable: just as there are many deterministic models, there could be many varieties of indeterminism. This is clarified by the chocolate-ice-cream machine.[4, chp. 8] The machine indeterministically selects a chocolate or ice cream, and slams its choice down our throats. What we will eat is not determined, but that is not the same as choosing between chocolates and ice cream in mundane life.

One way to formalise the difference between mundane time and the time of the old physics is to speak of the topology (or structure) of time, in the sense of temporal logic.[4, chp. 8] The time of daily experience is linear towards the past (we believe we cannot change the past), but branches towards the future (we believe the future is NOT already determined by the past but is influenced by our choices and efforts).

This past-linear future-branching mundane time is *not* the same as the time of physics which is usually assumed to be superlinear (i.e., to have the topology of the real line). That belief arose for a peculiar reason: because of a wrong way of doing calculus, as currently taught in universities. That requires time to be like the real line just to be able to make sense of the differential equations of physics.

Can we make physics compatible with the observed mundane creativity of living organisms? Minimal compatibility with everyday observations requires that we reformulate physics in a non-mechanistic way so that the entire past does *not* fully determine future. FDEs provide an easy way to do that provided we admit also advanced propagators instead of discarding them as "unphysical". This amounts to a minimal change in existing physics. If advanced interactions are rare, and retarded interactions predominate, the model remains time irreversible.

1.4 Causality

Why were advanced propagators excluded as "unphysical"? Physics texts typically justify this on the *metaphysical* grounds of "the principle of causality". That is problematic for various reasons. First, it is confusing, as metaphysics often is.

Thus, the word "causality" is commonly used in diametrically opposite senses. One sense is that of *mechanistic* causality: that the future is determined by the past, just as initial data determines the solution of an ODE, or past data determines the solution of a retarded FDE. The second sense is that of *mundane* causality that human actions are partly responsible for the future: we believe a thief is the cause of a theft. In both cases, future events have an antecedent cause. However, in the case of the thief we believe his

actions were *not* fully determined by *his* past, hence we punish the thief (not his past!).

That is, the single word "causality" has two opposite meanings: (1) that the future is fully determined by the past (hence we exclude advanced propagators) and (2) that human actions, *not* fully determined by the past, are partly responsible for the future (hence we punish thieves). It is an elementary principle of classical logic that from contradictory premises any desired conclusion may be drawn. Hence, the word "causality" with its diametrically opposite meanings is a rich source of confusion, for it can be used to conclude anything we want!

What complicates matters further is that *both* the above contradictory senses of causality are related to deep-rooted religious dogmas. Thus, Aquinas maintained that God rules the world with eternal laws of nature (determinism). On the other hand, the church has long maintained, since Augustine, that God punishes evil-doers in hell. Why should a person be punished (or rewarded) if the future is already determined and hence entirely beyond his control?

Such questions bring out the manifest incoherence in these dogmas about "causality". To "manage" the contradictions, and incoherence, and "save" those dogmas from rejection, there is a vast and confusing discourse on the theology of "free will", which invariably creeps into discussions about time in physics.[5]

However, the very use of the term "free will" should warn us that the discussion has strayed from physics to theology. The issue at hand is whether physics agrees with everyday experience, as it must. Why should theology be essential to mediate a conflict between physics and everyday experience?

To reiterate, if scientific theory does not agree with mundane experience, we should construct a better scientific theory more in accord with experience.

1.5 The tilt in the arrow of time

As already noted, a simple way to do that is to admit *both* retarded and advanced propagators, though in different proportions. That is, we use a convex combination of the two propagators, with weights α and $1-\alpha$, where α, the coefficient of the advanced component, is a small number (say, $\alpha \approx 10^{-10}$). This situation has been described as a "tilt in the arrow of time": *most* physical interactions travel from past to future, but in some rare cases interactions travel from future to past. A "tilt" also raises the interesting possibility of "time travel", in some sense.

Note that a "tilt" does *not* involve any new physical hypothesis. On the contrary, we just *dropped* the hypothesis of (mechanistic) "causality" used to reject advanced propagators. A convex combination of advanced and retarded propagators just gives us the *most general form* of classical electrodynamics and post-relativity physics (including gravitation). We should first study this general form and *then* compare the results of such a study with observations. If the comparison with observations so requires it we can put $\alpha = 0$, to recover (mechanistic) "causality". This approach is obviously better than proceeding

on metaphysical guesswork influenced by religious dogmas.

The immediate mathematical consequence of a "tilt" is this: the resulting equations of motion (of a charged or neutral particle) are mixed-type FDEs. The time asymmetry provided by mixed-type FDEs differs from the time asymmetry provided by retarded FDEs.

2 Advanced FDEs

To understand the difference, let us first consider the fully advanced case $\alpha = 1$. The simplest advanced FDE is of the form

$$\dot{y}(t) = y(t+\tau) \tag{1}$$

where $\tau > 0$ is constant. An advanced FDE such as (1) models a situation where the present state of a system $y(t)$ depends upon its *future* state $y(t+\tau)$.

This equation cannot be solved just by prescribing "initial" data $y(0)$: we can see this by repeating the reasoning used in the retarded case.

However, an advanced FDE cannot be solved even by providing past data. An advanced FDE is the exact time reverse of a retarded FDE: from the theory of retarded FDEs we can obtain the theory of advanced FDEs just by interchanging past and future. That is, to solve (1), we need to prescribe *future* data $y(t), t \geq 0$. We can use this future data to obtain a unique *past* solution, $y(t), t \leq 0$.

We saw that a retarded FDE can be solved forward in time but cannot, in general, be solved backward in time. Symmetrically, an advanced FDE can be solved backward in time but cannot, in general, be solved forward in time. We can see this by slightly modifying the earlier example used for retarded FDEs.

2.1 An example

Consider the FDE

$$y'(t) = a(t)y(t+1), \tag{2}$$

where a is a continuous function which vanishes outside [0, 1], and satisfies

$$\int_{-\infty}^{\infty} a(t)\ dt = \int_{0}^{1} a(t)\ dt = 1. \tag{3}$$

For example,

$$a(t) = \begin{cases} 0 & t \leq 0, \\ 1 - \cos 2\pi t & 0 \leq t \leq 1, \\ 0 & t \geq 1. \end{cases} \tag{4}$$

For $t \geq 1$, the FDE (2) reduces to the ODE $y'(t) = 0$, so that, for $t \geq 1$, $y(t) = k$ for some constant k $(= y(1))$.

Now, for $t \in [0,\ 1]$,

$$\begin{aligned} y(t) &= y(1) - \int_t^1 y'(s)ds \\ &= y(1) - \int_t^1 a(s)y(s-1)ds \\ &= y(1) - y(1)\int_t^1 a(s)ds, \end{aligned} \tag{5}$$

since $y(s-1) \equiv k = y(1)$ for $s \in [0, 1]$. Hence, using (3), $y(0) = 0$, no matter what k was. However, since $a(t) = 0$ for $t \leq 0$, the FDE (2) again reduces to the ODE $y'(t) = 0$, for

$t \leq 0$, so that $y(0) = 0$ implies $y(t) = 0$ for all $t \leq 0$.

Hence, the future of a system modeled by (2) cannot be predicted from a knowledge of the entire past; for if the past data (i.e., values for *all* past times $t \leq 0$) are prescribed using a function ϕ that is different from 0 on $[-\infty,\ 0]$, then (2) admits no forward solutions for $t \geq 1$. If, on the other hand, $\phi \equiv 0$ on $[1,\ \infty]$, then there are an infinity of distinct forward solutions. Fig. 1 shows three such solutions. In either case, knowledge of the entire past furnishes no information about the future.

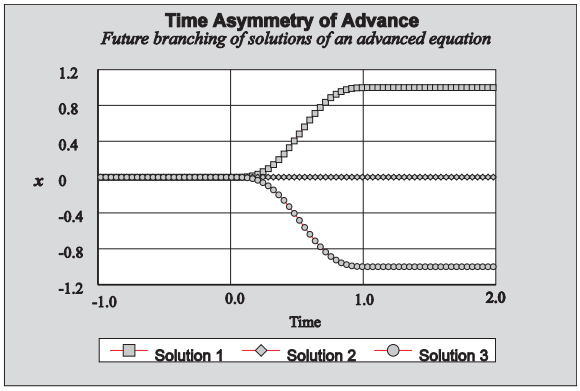


Figure 1: **Asymmetry of advanced FDEs:** Three different solutions of an advanced FDE with the same past have different futures, so that advanced FDEs cannot be solved forward in time. That is, future cannot be inferred from knowledge of past.

With advanced FDEs, multiple futures may collapse to a single past, a situation which may be better described by saying that solutions of advanced FDEs *branch* towards the future.

2.2 Popper's pond paradox

Even this basic knowledge about advanced FDEs easily resolves the pond paradox which so confused Karl Popper. Thus, the retarded solutions of the wave equation correspond to the ripples that spread out when a stone is dropped into a pond. The advanced solutions are the time reverse: they correspond to ripples that spontaneously converge from the edge of the pond. We normally observe the former but not the latter solution.

But suppose someone were to observe this rare occurrence and video record it. A physicist might suspect that the record has been faked; that someone actually video recorded the ripples spreading outward, and then recaptured the video playing it backward. He then falsely claimed that he has actually observed one of those rare events corresponding to advanced waves. How to discriminate between the two possibilities?

In a series of articles in the journal *Nature* long ago [6, 7, 8, 9] Popper suggested that the physicist should ask: by what *process* can one make this (converging ripples) happen? He used Huygens' principle to argue that ripples which arise spontaneously at the edge of the pond could build up into a converging wave *only* if they were coherent. This coherence basically requires a single source; it could happen if we have a perfectly circular pond and the ripples originate from the centre and are reflected back at the edge. Popper opined that coherence could not arise in any other way, except by chance. Further, Popper argued, the probability of coherence happening by chance, across multiple sources, is negligible.

2.3 Resolution of the pond paradox

If we think about it a little, Popper's conclusion is strange because he used metaphysical reasoning to limit physical phenomena. Whether or not advanced interactions exist is a matter of physics, for it should be determined by empirical observations. How did Popper manage to exclude them without reference to physics?

The above understanding of advanced FDEs brings out the error in Popper's reasoning. Phenomena modeled by advanced FDEs *cannot* be explained causally (from the past) just as phenomena modeled by retarded FDEs cannot be explained teleologically (from the future). No "cause" (past data) can explain the spontaneous convergence of advanced waves any more than a "purpose" (future data) can explain a divergent ripple.

Popper's error was the metaphysical stipulation that all phenomena *must* admit a causal explanation, so that there must be a way to make the future happen mechanistically. This is just a disguised and more confusing version of the same old argument for rejecting advanced interactions by invoking causality. (Recall that mechanistic causality was Aquinas' theological dogma, but is now masquerading as an argument about physics in a high-impact journal.)

Popper admitted that my arguments were "strong"[10]; he said he would respond in more detail later, but died before he could do so. In fact, there is no answer to the argument: the existence of advanced interactions must be decided empirically, not by appeal to confusing metaphysics.

Incidentally, we derive the following valuable conclusion from the simple example above. If advanced interactions exist, empirical proof of that would be the existence of some "spontaneous" phenomena which do *not* admit mechanistic explanations from the past.[11]

To better understand the empirical consequences, it is important to study mixed-type FDEs, for, not only Popper, but leading physicists, such as Richard Feynman, have got confused by reasoning intuitively about the issues involved.

2.4 Earlier theories

Thus, advanced electromagnetic radiation was admitted in the Wheeler-Feynman absorber theory of radiation.[12, 13] That theory sought to explain time asymmetric radiation damping starting from time-symmetric propagators (i.e., $\alpha = \frac{1}{2}$). The observed predominance of retarded radiation was explained by putting a condition on the cosmos: namely that it should be totally absorbing.

However, the arguments of Wheeler and Feynman are circular, as I pointed out long ago.[14] Wheeler and Feynman, themselves, did sense the possibility of such circularity and tried to resolve it. Like Popper, they stated that the past motions of particles are uncorrelated. However, in the presence of even a tiny amount of advanced radiation ($\alpha \neq 0$), correlations *will* travel into the past, just as they travel into the future with retarded radiation. So, with time-symmetric propagators

the time asymmetric assumptions made by Wheeler and Feynman (or Popper) of random and uncorrelated past motions were wrong. Unfortunately, some people (e.g. [**?**]) are still using the Wheeler-Feynman absorber theory without addressing that error. My own version of the absorber theory predicted that small amounts of advanced radiation actually exist.[14] Issues concerning the absorber theory were summarised in an earlier article in this journal, also posted on the arxiv.[15]

2.5 The new theory

However, the existence of advanced interactions is now being considered from a fresh point of view,[4] unconnected with any absorber theory. To reiterate, the new point of view is this: it is incorrect to exclude advanced interactions on metaphysical grounds such as causality or Popper's argument. We should instead (1) set up a theory which includes advanced interactions, (2) determine its empirical consequences, and (3) compare those consequences with empirical observations to *then* decide whether or not advanced interactions exist.

From this general perspective advanced interactions and mixed-type FDEs are not limited to electrodynamics. The retarded gravitation theory, outlined earlier,[1] can be easily modified to include advanced interactions. That is, mixed-type FDEs are relevant to the interaction of all particles, not just charged particles.

3 Mixed-type FDEs

Mathematically speaking, a tilt results in mixed-type FDE of the following kind:

$$\dot{y}(t) = \alpha y(t+\tau) + (1-\alpha)y(t-\tau). \qquad (6)$$

(This is a simplified equation, and not the most general one possible.) This describes a situation where the rate of change of y depends upon both the future and past in different proportions. What happens in this situation?

Some general features are obvious. First of all, if the coefficient of the advanced term, α, is small, we can regard it as a perturbation on the retarded FDE model. That is, the retarded FDE model ($\alpha = 0$) would continue to describe the world to a first approximation: the world is approximately mechanistic, and future is approximately decided by the past. Time asymmetry persists, for the reverse situation is not true: future cannot be used to determine the past even as a first approximation.

Since, however, $\alpha \neq 0$, there is an advanced component, hence future cannot be *fully* determined or controlled from the past. Thus, full control is impossible from either past or future.

The curious conclusion, however, is this: even if we prescribe most of *both* past *and* future, that may no longer determine the present!

For example, consider the following mixed type FDE,

$$\dot{y}(t) = a(t)y(t+1) + b(t)y(t-1), \qquad (7)$$

where a has the same properties as above. For b we use the same function used in the retarded FDE case, namely that it has $\int_0^1 b(t)dt = -1$, but now stipulate that it has support on the interval $[2,\ 3]$.

On the interval $(\infty,\ 1]$ we have $b \equiv 0$ so (7) is a pure advanced FDE, and we get future branching solutions from past data as earlier. On $[1,\ \infty)$ we have $a \equiv 0$ so (7) is a pure retarded equation, and we get past branching from future data as before. Combining the reasoning used in the separate cases of retarded and advanced equations, we obtain the solutions below.

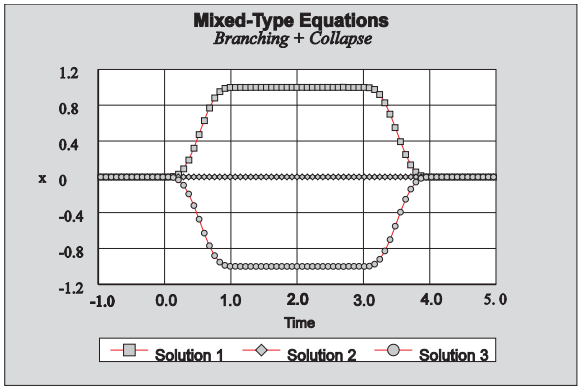


Figure 2: With a realistic mixture of history-dependence and a small amount of anticipation, the past still fails to decide the future. With this model, all phenomena do not admit mechanistic causal explanations, so that spontaneity really is possible. The existence of a small tilt is exactly the condition for time-travel of the second kind.

3.1 Consequences

That is, with a model which uses mixed-type FDEs, with a small advanced component, we have the following immediate consequences.

1. Retarded FDEs, would remain a good first approximation, so past data could still be used to *approximately* determine the future.

2. Time asymmetry persists, since future data cannot similarly be used to determine past, even as a first approximation.

3. Past data fails to decide future exactly. Even if we could prescribe the entire past accurately, that would still not determine the future.

4. There must exist "spontaneous" phenomena which do not admit an explanation from the past ("causal" explanation), in principle.

As is clear, this describes a situation closer to experience.

4 Time travel

The puzzling and counter-intuitive features of such a model are best brought out by the vast confusion in both physics and popular literature (and films) about time travel. Therefore, let us now turn to the question of time travel with a tilt in the arrow of time.[16]

In the popular imagination, time travel is associated with time machines. While physics has not defined life, it is easy enough to define a machine. The defining feature of a *machine*, any machine, is that it can be fully *controlled* (by pressing a button for example). But can a time machine be controlled?

4.1 Two types of time travel

To make matters clearer, time travel may be classified as being of two possible types: (1) with machines, and (2) without machines. Time travel of the first kind visualises time machines which physically transport entire human beings at the press of a button. In time travel of the second kind, only the occasional transfer of small bits of information from future to the past is contemplated, for example through an advanced signal coming from the future. With a "tilt" only time travel of the second kind is contemplated.

4.2 Time machines

However, time machines abound in the science fiction literature, the classic example being H. G. Wells' *Time Machine*. Travel across the galaxy, within the short lifespan of humans, is also visualised in the science fiction literature or films like Star Trek.

Time machines are also found in the scientific literature: for example, the Gödelian time machine based on the closed timelike curves which arise in the Gödel cosmos. (It is believed that given a timelike curve we can construct a rocket which will follow the curve. Such a rocket might require a large amount of energy, but the Gödel cosmos rotates and is not asymptotically flat; hence energy is not well-defined in it.)

Travelling large distances in a time span which is short (compared to light travel times, and the human life span) necessarily involves time travel. Hence, a more recent NASA funded study by Kip Thorne and others [17, 18, 19, 20] explored the possibility of time machines based on general relativity.

Thorne et al. focused on TWISTs: traversable wormholes in space time. A wormhole, such as one made in an apple by a worm, may connect two distant points of spacetime (analogous to the surface of the apple). If the apple is very big, like the cosmos, only a short time may be needed to travel through it compared to the time needed to travel around it (i.e., on the surface of the apple). Such a wormhole is called traversable if the tidal forces within it are not so strong as to kill the traveller. Thorne et al. concluded that such traversable wormholes could be built using negative energy. This possibility is visualised in Carl Sagan's novel and film *Contact*. Thorne even suggested ways of building negative energy, though he was earlier a strong advocate of the positive energy condition[2].

4.3 The tachyonic anti-telephone

This issue of control explicitly arose in the context of the tachyonic anti-telephone.[21]. Tachyons are hypothetical particles which move faster than light; if they exist, they can, in principle, be used to communicate with the past. The tachyonic anti-telephone is a hypothetical device which uses tachyons to allow one to converse with people in the past.

Suppose that, using this device, Shakespeare dictated the script of *Hamlet* to Bacon. Since Bacon came before Shakespeare he has

[2]and had earlier objected to my use of negative energy for gravitational screening

chronological priority, being the first person to actually write down the script of *Hamlet*. So, whom should we rightly regard as the author of *Hamlet*: Shakespeare or Bacon? To resolve this paradox, it was opined that though the cause is in the future, Shakespeare is the one who has *control* over the text of *Hamlet*, therefore he remains the author.

This naive belief that with time machines the future can be used to *control* the past involves a fallacy similar to Popper's pond. The fallacy is to put together two contradictory pictures of time: (1) the notion of time used in physics, according to which the evolution of the cosmos is determined by various equations ("laws of nature"), and (2) the time of our daily experience (mundane time), in which *we* make the future happen. These are contradictory not compatible as already explained. It is elementary that from contradictory hypotheses one may derive any desired conclusion. The solution is to remove the incompatibility, by making physics compatible with mundane time. If we do so by means of a tilt, we find that future cannot be used to control the past. That is, in the above scenario, Shakespeare has little or no control over the play Hamlet.

4.4 Other paradoxes

The resolution of the grandfather paradox and the Augustine-Hawking paradox is similar. In the grandfather paradox we use a time machine (constructed using relativity say) to send Tim back in time. Then we switch from the equations of physics to the (incompatible) mundane view of time, and suppose that acting on his own volition, Tim kills his own grandfather. So how could Tim have been born?

Similarly, the Augustine-Hawking paradox contemplates closed timelike curves in relativity. Then there is a switch to mundane time to say that given such a curve *we* can build a rocketship which will travel around the curve. Then Hawking concludes that the rocketship repeats its history, so we no longer have "free will"! This "free will", Hawking argues, is essential for the belief that we are free to perform any experiment we like. Once again, the contradiction arises from the unstated assumption that the equations of relativity on the one hand decide the future from the past, and on the other hand are compatible with mundane experience (that past does not entirely decide the future, leaving us free to do so). The paradoxical thing is that mundane time plus time travel is now being applied to try to create the past (which is implausible, even on mundane experience, or with a "tilt").

These mistakes made by top physicists and philosophers such as Feynman, Hawking, and Popper show that talking about the future interacting with the past is a tricky matter, made murkier by the constant intrusion of church dogma about "free will" as in Hawking's arguments which directly mimic those of the theologian Augustine, as I have explained at length elsewhere.[5]

4.5 No time machines

In fact, the science fiction scenario of hopping into a time machine and pressing a button to go back into the past assumes the possibility of controlling the time machine from the

future. The consequences of the mixed-type FDE model outlined above make this impossible. Hence, there can be no time machines.

We can understand this conclusion in another way.[16] Just as retarded interactions increase entropy, advanced interactions *decrease* entropy. (On the advanced FDE model one has less information about the past than the future.) So, the mixed-type FDE model may be thought of as a model involving a combination of two sorts of processes: those which increase entropy and those which decrease entropy. Processes which decrease entropy (of the entire cosmos) may exist without conflict with the entropy law (second law of thermodynamics), provided that entropy increasing processes predominate, as they do in the mixed-type FDE model.

However, if entropy decreasing processes could be *mechanically* controlled, at the press of a button, that would allow us to decrease entropy by an unlimited amount. In short, a time *machine* (which can be controlled) would be a perpetual motion machine, and is hence impossible. Time travel can only be of the second kind: without machines. That is, while time travel of the second kind is possible (e.g. with a tilt), time machines are not.

4.6 Empirical evidence for time travel

Finally, let us reiterate the empirical consequences of a tilt. Hawking[22] asserted that if time travel were possible, we would have been invaded by hordes of tourists from the future. That argument naively assumes that time travel means time machines which, as we just saw, is impossible. Hence, going round looking for tourists from the future as evidence for time travel is absolutely the wrong kind of thing to do.

Instead, we should expect to observe some rare events that are spontaneous, and cannot be explained from the entire past.

For example, in the grandfather paradox, the time traveller's chronologically earliest appearance in space time (which is earlier than his biological birth from his mother's womb) would be such a spontaneous event, for it has no possible explanation from the past. (If it can be explained from the past, there is no time travel involved.) However, that is just a figurative example, for, as already explained, time machines are impossible.

Is mundane human creativity (more generally, the creativity of all living organisms), or just the existence of life, an example of such spontaneity? I believe so. If so, that would be empirical proof supporting the mixed-type FDE model against other models.

4.7 Spontaneity vs chance

Note that spontaneity differs from chance. Mathematically, this is readily understood. Spontaneity is modeled by mixed-type FDEs. "Chance" can mean many things: in this context we take it to mean stochastically perturbed retarded FDEs. Though the two models are mathematically very distinct, the solutions have some similarity. One distinction is this: spontaneity leads to reduction of entropy or increase in negentropy. Chance, on the other hand, as in classical thermodynam-

ics, is believed to usually lead to an increase of entropy. Of course, chance too could lead to a decrease in entropy; however, as a rule of the thumb, one expects that the time scale for that to happen would be much longer than the time scale for decrease of entropy with spontaneity.

4.8 Microphysical consequences

Whether or not scientists accept the empirically observed creativity of living organisms, which is our most common experience, we at least now have a model for it. Biology is becoming increasingly important, and mixed-type FDEs are the first model which can account for some basic biological observations.

Since the effects of advanced interactions would be tiny one may imagine that they are especially important at the molecular level, for biological macromolecules, say. Currently, molecular simulation is done using outdated Newtonian many-body theory, and the Coulomb force. It would be interesting to redo this using first retarded FDEs and then mixed-type FDEs.

At the level of single particles, the consequences of the mixed-type FDE model are likely to be even more prominent. We will take up the relation of a tilt to quantum mechanics in more detail in the next and last article in this series. Mundane experience shows that living organisms are somehow able to scale up spontaneity to the macrophysical level. Is this relevant to the current technological problem of scaling up quantum computers?

5 Conclusions

Physics must be compatible with biology and our mundane experience. A simple way to achieve compatibility is to permit a tilt in the arrow of time. This corresponds to using mixed-type FDEs (for both electrodynamics and gravitation). The resulting (non-local) physics is non-mechanistic. With mixed-type FDEs full control of future is not possible from past and much less so is control of the past possible from future. This understanding resolves all paradoxes of time travel. Hence, also, time travel can only be of the second kind: without machines (i.e, time travel is possible, but time machines are not). The correct refutable consequence of realistic time travel is the existence of spontaneous events, not hordes of tourists from the future.

References

[1] C. K. Raju. Functional differential equations-4. retarded gravitation. *Physics Education*, 31(2), April-June 2015. http://www.physedu.in/uploads/publication/19/309/1-Functional-differential-equations-4-Retarded-gravitation-(2).pdf.

[2] C. K. Raju. Functional differential equations. 1: A new paradigm in physics. *Physics Education (India)*, 29(3), July-Sep 2013. http://physedu.in/uploads/publication/11/200/29.3.1FDEs-in-physics-part-1.pdf.

[3] C. K. Raju. Functional differential equations. 2: The classical hydrogen atom. *Physics Education (India)*, 29(3), July-Sep 2013. `http://physedu.in/uploads/publication/11/201/29.3.2FDEs-in-physics-part-2.pdf`.

[4] C. K. Raju. *Time: Towards a Consistent Theory*, volume 65 of *Fundamental Theories of Physics*. Kluwer Academic, Dordrecht, 1994.

[5] C. K. Raju. *The Eleven Pictures of Time: The Physics, Philosophy and Politics of Time Beliefs*. Sage, 2003.

[6] K. R. Popper. The arrow of time. *Nature*, 177:538, 1956.

[7] K. R. Popper. Irreversibility and mechanics. *Nature*, 178:382, 1956.

[8] K. R. Popper. Irreversible processes in physical theory. *Nature*, 179:1297, 1957.

[9] K. R. Popper. Time's arrow and entropy. *Nature*, 207:233–34, 1965.

[10] K. R. Popper. Personal Communication, 4 May 1990. Letter.

[11] C. K. Raju. Time and life: testing a tilt in the arrow of time, Dec 2008. Online essay at `http://fqxi.org/community/forum/topic/347`.

[12] J. A. Wheeler and R. P. Feynman. Interaction with the absorber as the mechanism of radiation. *Rev. Mod. Phys.*, 17:157–81, 1945.

[13] J. A. Wheeler and R. P. Feynman. Classical electrodynamics in terms of direct interparticle action. *Rev. Mod. Phys*, 21:425–3, 1949.

[14] C. K. Raju. Classical time-symmetric electrodynamics. *Journal of Physics A: Math. Gen.*, 13:3303–17, 1980.

[15] C. K. Raju. Electromagnetic time. *Physics Education (India)*, 9(2):119–128, 1992. http://arxiv.org/pdf/0808.0767v1.

[16] C. K. Raju. Time travel and the reality of spontaneity. *Foundations of Physics*, 36:1099–1113, 2006.

[17] M. S. Morris and K. S. Thorne. Wormholes in Spacetime and their use of Interstellar Travel: A Tool for Teaching General Relativity. *Amer. J. Phys.*, 56:3954, 1988.

[18] M. S. Morris, K. S. Thorne, and U. Yurtserver. Phys. *Rev. Lett*, 61, 1988.

[19] S. W. Kim and K. S. Thorne. Phys. *Rev. D*, 44, 1991.

[20] K. S. Thorne. *Black Holes and Time Warps: Einstein's Outrageous Legacy*. W. W. Norton & Co., 1994.

[21] G. A. Benford, D. L. Book, and W. A. Newcomb. The tachyonic antitelephone. *Physical Review D*, 2:263–65, 1970.

[22] S. W. Hawking. Chronology protection conjecture. *Phys. Rev. D*, 46:603–11, 1992.

The Paradox of Two Charged Capacitors – A New Perspective

Ashok K. Singal

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

(Submitted 10-07-2015)

Abstract

It is shown that the famous paradox of two charged capacitors is successfully resolved if all the energy changes in the system are properly considered when some of the charges are transferred from one capacitor to the other. It happens so even when the connecting wire has an identically zero resistance, giving rise to no Ohmic losses in the wire. It is shown that in such a case the "missing energy" goes into the kinetic energy of conducting charges. It is shown that radiation plays no significant role in resolving the paradox. The problem can be formulated and successfully resolved in a novel form, where the capacitance of the system is increased by stretching the plates of the original capacitor, without involving any connecting wires in a circuit. There is an outward self-force due to mutual repulsion among charges stored within each capacitor plate, and the work done by these self-forces during an expansion is indeed equal to the missing energy of the capacitor system.

1 Introduction

In the famous two-capacitor paradox[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] one of the capacitors, say C_1, of capacitance C is initially charged to a voltage V_0 with charge $Q_0 = CV_0$ and energy $U_0 = CV_0^2/2 = Q_0V_0/2 = Q_0^2/(2C)$, while the other similar capacitor, C_2, is initially uncharged, thereby the total energy of

the system being U_0. Both capacitors are assumed be to identical in every respect. Now C_1 is connected to C_2 using a conducting wire, resulting in transfer of some charges from C_1 to C_2. From symmetry each capacitors will end up with charge $Q_0/2$ and voltage $V_0/2$, with energy of each as $CV_0^2/8 = U_0/4$. Therefore the total energy of the system will be $U_0/2$. What happened to the other half of the energy?

Puzzling though this might appear at a first look, the loss of energy is easily explained if we consider the Ohmic losses in the connecting wires. Suppose the connecting wires have a resistance R (Fig. 1), then the charging current will be $(V_0/R)e^{-2t/(RC)}$ and the dissipated energy will be,

$$\int_o^\infty I^2 R\,\mathrm{d}t \;= \int_o^\infty \left(\tfrac{V_0}{R}e^{-2t/(RC)}\right)^2 R\,\mathrm{d}t = \tfrac{CV_0^2}{4} = \tfrac{U_0}{2}. \qquad (1)$$

The above equation is true for any finite value of R. But what happens if there were no Ohmic losses, e.g., if in our ideal hypothetical case the resistance were identically zero (a superconductor!). The total energy in the two capacitors, however, is still half of the initial energy, so where does the remaining energy disappear?

Of course there is nothing special about the two capacitors being identical. In the case the two capacitances C_1 and C_2 are unequal, the initial stored energy $U_0 = Q_0^2/2C_1$ after transfer of charges reduces to $Q_0^2/(2(C_1 + C_2)) = U_0C_1/(C_1 + C_2)$. This implies a loss

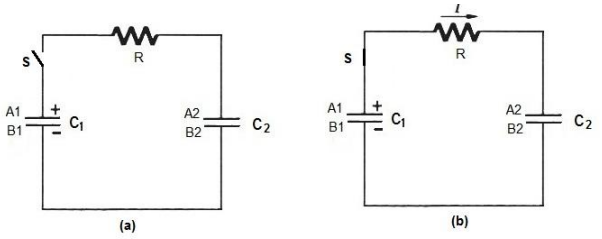


Figure 1: Charging a parallel plate capacitor.

of energy[12]

$$\Delta U = \frac{U_0\, C_2}{C_1 + C_2}. \qquad (2)$$

For equal capacitances ($C_1 = C_2$) the energy loss reduces to $U_0/2$, as derived earlier. Of particular interest is the case for large C_2 ($C_2 \to \infty$), where all stored energy is lost.

Since the charges undergo acceleration while moving from higher to a lower potential in case of zero resistance, can it be that whole of the missing energy appears as radiation from these accelearted charges? The current belief seems to be that the missing energy is radiated away.[13, 14] It should be clarified that here we are not talking of the thermal electromagnetic radiation like in a resistance wire, but of electromagnetic waves radiated from an antenna system. As we will show in Section 4, the present radiation calculations are based on circular arguments. Moreover from maximum possible radiation losses from Larmor's formula we will argue that missing energy cannot be accounted for by radiation losses, and that the radiation hypothesis does not offer a satisfactory resolution of the paradox.

2 Where does the missing energy go?

The missing energy actually goes into the kinetic energy of conducting charges getting transferred from C_1 to C_2 for $R = 0$[12]. Actually one has to be cautious when extremely low resistances are considered. The conductivity of a metal is directly proportional to the characteristic time τ between successive collisions of the charge carriers that results in loss of directional correlation[1, 15]. Drift velocity in the conductor is $qE\tau/m$, where E is the electric field and q is the electric charge and m is the mass of the charge carrier (an electron!). A typical value for τ in the metals is $\approx 10^{-14}$ sec with typical drift velocity usually a fraction of a mm/sec. The resistivity is $\propto 1/\tau$, and low resistivity implies τ is large and then the mean free path λ between collisions ($\propto \tau$) would also be large. In that case there will be fewer collisions and in an extreme case, we could assume that the mean free path λ will be large enough to be longer than the length of the wire or channel joining the two capacitors. This could be termed as $R = 0$ case. Then the conducting charges will steadily gain velocity and kinetic energy as the collisions will be minimal. In that case the charges will not undergo Ohmic losses and when they reach C_2 their kinetic energy will be equal to the potential energy difference during the transfer between two capacitors.

The gain in kinetic energy in the absence of Ohmic losses is easily calculated from the change in potential energy of each charge. The charge gains a velocity increment $\Delta v = qE\Delta t/m$ or $m\Delta v = qE\Delta x/v$ which implies a kinetic energy gain $\Delta(mv^2/2) = q\Delta V$. For a charge transfer Q from C_1 to C_2, the voltage difference between the two becomes $\Delta V = (Q_0 - Q)/C_1 - Q/C_2$. Then the total kinetic energy gained by charges during a total charge transfer Q_2 from C_1 to C_2 is,

$$\int_0^{Q_2} \left(\frac{Q_0 - Q}{C_1} - \frac{Q}{C_2} \right) \mathrm{d}Q = \frac{Q_0 Q_2}{C_1} - \frac{Q_2^2}{2} \left(\frac{1}{C_1} + \frac{1}{C_2} \right). \quad (3)$$

As the voltage difference between C_1 to C_2 becomes zero at the end, then $(Q_0 - Q_2)/C_1 = Q_2/C_2 = Q_0/(C_1 + C_2)$, implying that the total kinetic energy gained by the charges from (3) is $Q_2 Q_0/(2C_1) = Q_2 V_0/2 = U_0 C_2/(C_1 + C_2)$, in agreement with the energy loss ΔU in (2).

When the charges finally get deposited on plates of the capacitor C_2 this kinetic energy should get transferred to the plates of C_2, which as we discuss later, could even be utilized by an external agency, or else the plates of C_2 would get heated. The problem as posed is between two equilibrium states in which the charges are stationary both initially and in the final state. Thus there should be no residual kinetic energy in the system. It implies the charges when finally get deposited on plates of the capacitor C_2, they remain stuck there. This means that all the kinetic energy gained by the moving charges in the absence of Ohmic losses, should get transferred to the plates of C_2, which we assume to be not free to move

(clamped to the lab bench!). There will thus be necessarily inelastic collisions and the plates of C_2 would get heated because of these inelastic collisions. There could also be some partial energy loss in sparks but as we show later the *whole* energy loss cannot be accounted for by the radiation.

Actually $R \to 0$ is only a mathematical idealization which may not hold good when we go below certain very low resistance values. Let us take a material which can turn into a superconductor, say lead. If we lower its temperature, the resistance of the conductor will reduce steadily up to a certain point (7.22 K for lead),[15] below which it may suddenly become zero as the material turns into a superconductor. That means either it will be a normal electrical resistance with Ohmic losses above this turnover point or it will be zero resistance without Ohmic losses below this point. Thus there is a discontinuity in resistance and one does not have $R \to 0$ in limit.

Let us examine the idea of $R \to 0$ in limit in a non–superconductor material. Resistance of a wire is $R = \rho L/A$ where ρ is the resistivity, L is its length and A is the cross section. We cannot increase A beyond certain values (for example, it cannot be larger than the capacitor plate size), so we can decrease ρ or/and L to reduce R. Now $\rho \propto 1/\lambda$, the mean free path, meaning $R \propto L/\lambda$. Usually $L/\lambda \sim 10^7$ for a few cm long wire, however starting from some finite resistance, as we go to lower R, by decreasing ρ and thereby increasing λ or decreasing L, the ratio L/λ will decrease. And near some critical value of resistance, say R_c, λ will approach L, that is the mean free path will become equal to the length of the wire or channel joining the two capacitors. At this stage $1/e$ th fraction of the current carrying charges will pass the length of the wire without suffering any collisions and thus without undergoing Ohmic losses. The remaining charges will of course undergo Ohmic losses due to collisions. It is of course statistically a random process. Let us denote the electric current by the latter as I_1 and that by the collisionless charges as I_2. Then $P_1 = VI_1$ fraction will be the Ohmic losses and $P_2 = VI_2$ fraction will be the power going into the kinetic energy of charges. Thus there will be sharing of power losses between the two processes, with total power loss as $P = P_1 + P_2 = V(I_1 + I_2) = VI$.

Now let us see what will happen as we reduce R. Initially with much higher resistance than R_c, with $L/\lambda \gg 1$, there will be only $P = P_1$, the usual Ohmic losses. As we reach R_c, the Ohmic losses (P_1 fraction) will steadily decrease while the P_2 fraction will increase. For much lower resistance than R_c, there will be almost no collisions, there will be only $P = P_2$, with the conducting charges gaining the kinetic energy in the absence of collisions and the P_1 losses being zero.

In (1) it is implicitly assumed that all charges undergo Ohmic losses however low the collision rates might be (even when $R \to 0$), and accordingly the dissipation losses are calculated. In reality it may not even be proper to still think of resistance below R_c in the usual ohm's law sense, when the collisions will be few and far between. Therefore $R \to 0$ might not be very meaningful much below R_c. Thus the mysterious difference be-

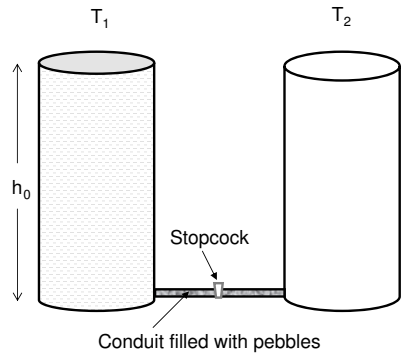


Figure 2: The equivalent case of missing-energy during transfer of water between two tanks of equal storage capacity

tween $R = 0$ and $R \to 0$ cases appears only because in the latter it is implicitly assumed that the charges lose their kinetic energy into Ohmic losses however low their collision rates might be, and accordingly we calculate the dissipation losses in (1), while in an identically zero resistance case, Ohmic losses are not even considered.

3 The equivalent case of water transfer between two tanks

An equivalent example exists in case of a water transfer from one full tank to an identical empty tank under the force of gravity (Fig. 2)[12, 16]. Initially the gravitational potential energy of the water to a height h_0 in tank T_1 is $U_0 = \int_0^{h_0} \rho g A z \mathrm{d}z = \rho g A h_0^2/2 = Q_0 V_0/2$, where ρ is the density of the water, A is the cross-section area of each tank, g is the acceleration due to gravity and z is the vertical distance. Then $Q_0 = \rho A h_0$ is the total quantity (mass) of water and $V_0 = g h_0$ is the gravitational potential. Now we open the stopcock, so that water is transferred from tank T_1 into tank T_2 through a conduit (Fig. 2). However when we consider the friction with the conduit walls and obstructions within (say, pebbles inside the conduit blocking a free flow of water) then the water loses all its kinetic energy during the transfer to tank T_2. At the end with each tank having $Q_0/2$ amount of water up to height $h_0/2$, the potential energy of the water in each tank is $Q_0 V_0/8 = U_0/4$ with the total energy of the system being $U_0/2$, exactly as in the two capacitor case. This is because the water in the upper half of tank T_1 goes into the lower half of tank T_2, then half of the total water mass (i.e., $Q_0/2$) which earlier was at a height between $h_0/2$ and h_0 in T_1 is now at a height between 0 and $h_0/2$ in T_2, thus ending up at an average height lower by $h_0/2$, implying an energy loss of $U_0/2$. If there is no friction with the conduit walls (and no obstructions within either), from Bernoulli's theorem[1] (or from simple energy conversion between potential and kinetic energy) the water would exit with a velocity $v = \sqrt{2g(h_1 - h_2)}$ or $\rho v^2/2 = \rho \Delta V$, at any moment when the heights of water columns in T_1 and T_2 are h_1 and h_2 respectively, with a gravitational potential difference $\Delta V = g(h_1 - h_2)$. The water will thus move in the conduit with a kinetic energy that could be even utilized with a suitable device attached to the conduit (a tiny electric power generator!) otherwise this

energy will be carried to the tank T_2 and ultimately lost as heat there by the time things have settled down.

For unequal tank capacities, let T_1 and T_2 have cross-section A_1 and A_2 respectively. Then the total water that will get transfered from T_1 to T_2 is $Q_2 = Q_0A_2/(A_1 + A_2)$, and the height of water columns in the two tanks will be $h = h_0A_1/(A_1 + A_2)$. That means this much amount of water would have fallen from a height of initial average value $(h + h_0)/2$ in T_1 to a final average value $h/2$ in T_2, implying an average height loss of $h_0/2$ and the loss in potential energy of $Q_2\ gh_0/2 = Q_2V_0/2 = U_0A_2/(A_1 + A_2)$. It also shows readily why the loss of energy in the tank (charged capacitor) system is the total transferred water (charge) Q_2 multiplied by half of the initial potential, i.e., $V_0/2$.

4 Possibility of radiation losses

In the radiation hypothesis the authors in general assume that the power losses (irrespective of the expression for radiation losses, (see e.g., (8), (9) and (10) in [13]) can be written as $V_XI = P_{rad}$ and have thus put $V_X = V_{12}$, where V_{12} is the potential difference between the two capacitors. Thus their assumption directly leads to $P_{rad} = V_{12}I$ and therefore $\int P_{rad}\,\mathrm{d}t = \int V_{12}I\,\mathrm{d}t = \int V_{12}\,\mathrm{d}Q$. From our (3) we know the right hand side is $CV_0^2/4$ irrespective of the time dependence of V. No wonder authors also get $\int P_{rad}\,\mathrm{d}t = CV_0^2/4$, as that is a built–in assumption. This way one is *bound to get* the same final result of energy losses irrespective of any other details of the exact radiation process that might have been assumed (whether it is a magnetic dipole radiation like the authors[13] assumed or some other process), and which could therefore be chosen any arbitrary function of time. In this particular case the authors emphasize that charging/discharging is not instantaneous. But according to this procedure for any arbitrary $P(t)$ one could define radiation resistance as $R_r = P(t)/I^2$, and then writing $V_{12} = IR_r$, one gets $P = V_{12}I$ which no wonder gives $\int P_{rad}\,\mathrm{d}t = CV_0^2/4$, and actually that way one does not really prove anything about the radiation process. It is not the radiation hypothesis that gets confirmed this way, it is only the a priori assumption of equating radiation losses P_{rad} (or losses in any other way!) to $V_{12}I$ which begets the apparently right answer. For this one does not even need to derive any complicated formulae for radiation expressions and it does not prove in any way that the radiation is that of magnetic dipole or some other "multipole". Different assumption about the radiation process (whether it is electric dipole or magnetic dipole or some other multipole) only at most may give a different time dependence of function $V(t)$ or $Q(t)$, but as the time integral of total charge transferred will be $Q_0/2$ and voltage $V_0/2$, one is bound to get the result for *energy dissipated* as $CV_0^2/4 = U_0/2$. Moreover when charging/discharging is not instantaneous, the Ohmic resistance is not identically zero and the lost energy $CV_0^2/4$ should then be distributed between dissipation in R and

radiation. But we find that the energy dissipation is fully satisfied by the Ohmic losses alone (Eq. (1)) even when the resistance reduces in limit to zero ($R \to 0$) and the radiation hypothesis is not at all needed.

It is possible to estimate how much maximum radiation losses can be there. From Larmor's formula[17] we know that the energy radiated by a non-relativistic charge accelerated for a time interval Δt (and thus having gained a velocity $v = a\Delta t$ in the absence of Ohmic losses) is $2q^2a^2\Delta t/3c^3$. For all the energy gained by the charge due to the potential difference to go into radiation implies

$$qV_{12} = \frac{mv^2}{2} = \frac{2q^2a^2\Delta t}{3c^3} = \frac{2q^2av}{3c^3} \qquad (4)$$

or

$$\frac{v}{a} = \Delta t = \frac{4q^2}{3mc^3} \sim \frac{r_e}{c}, \qquad (5)$$

where $r_e = q^2/mc^2$ is the classical electron radius.[17] Thus for *all* of the missing energy $U_0/2$ in the capacitor paradox to appear as radiation is possible *if and only if* the charges move from one capacitor to the other in a time interval of the order in which light travels the classical radius of the electron r_e, which is an impossible condition. In fact the radiation losses, due to the acceleration of the charges will be extremely small and can be made arbitrarily small by making the time over which the charge moves from C_1 to C_2 large enough. For example, an external agency using some electrical probe ("magic tweezers"),[1] could pick up charges one by one from C_1 at a higher potential and deliver them to C_2 at a lower potential at a leisurely rate (quasi-statically) and the difference in the potential energy of these charges can be utilized by the transferring agency. There will be no radiation losses, nor will there be any Ohmic loses. We shall further discuss one such alternate example in the next section.

5 A capacitor is charged without using resistive wires

Instead of charging a capacitor C_2 from C_1 using a wire of zero resistance, we could pose the problem in a different way. Let us suppose that we can expand or stretch the plates of a capacitor quasi-statically so that each plate area becomes double of its previous value, but without changing the plate separation. For simplicity we assume a parallel plate capacitor with dimensions a and b of the capacitor plates much larger than the plate separation, h, so that the electric fields within the capacitor can be considered, with negligible errors, to be uniform as in the case of infinite plates. Let $\sigma_0 = Q_0/A$ be the initial uniform surface charge density on the two oppositely charged plates, with $A = ab$ as the surface area of each plate. Then the electrostatic field is a constant, $4\pi\sigma_0$, in the region between the two plates which thus have a potential difference $V = 4\pi\sigma_0 h$. The field of course is zero everywhere outside. The mutual force of attraction on each plate is $2\pi\sigma_0^2$ per unit area, and the electric potential energy U_0 accumulated in separating the two plates by a distance h is $2\pi\sigma_0^2 A\, h$. The capacity of a parallel plate capacitor is given

by $C = A/(4\pi h)$,[15] and with energy $U_0 = CV^2/2 = Q^2/2C$.

With an expansion of the capacitor plates' areas by a factor of two, the charge density becomes half with the charges now distributed over its double charge capacity. The final energy of the capacitor is now only half of the previous value and the problem returns to the standard two capacitor paradox. The question again rises where has half of the energy gone. Now that there are no connecting wires with their resistance coming into picture, so we do not have to worry about Ohmic losses. There are no radiation losses either. As it is a quasi-static expansion there is no gain in the kinetic energy of current carrier charges. But we still have a problem of the missing energy.

Actually in addition to the force of attraction between two plates of a capacitor, there is also an outward force of repulsion within each capacitor plate. The presence of such self-repulsive forces within the capacitor plates and the work done against them during a Lorentz contraction of the system when the charged capacitor system moves from one inertial frame to another, was first shown explicitly by Singal[18] and accordingly the famous Trouton-Nobel experiment[19] was resolved from energetic points of view.[20] Here we will show by explicit calculations that the energy spent by the capacitor system during expansion is indeed equal to the missing energy, i.e, $CV_0^2/4$.

Adapting the calculations of [18] to our present case, we have calculated these force of self-repulsion in Appendix, where we find the expression for the rate of work done during an expansion of capacitor plates by the forces of self-expulsion as $\mathrm{d}\mathcal{W}/\mathrm{d}\eta = U_0/\eta^2$ (c.f. (16)) with η as the expansion factor.

Now integrating from initial $\eta = 1$ to a final expansion factor η_0, we get the amount of work done by the system during an expansion as $\mathcal{W} = U_0\,[1 - 1/\eta_0]$, which is equivalent to the energy loss ΔU in (2) with the charge capacitance having increased by a factor $\eta = (C_1 + C_2)/C_1$. In particular, for $\eta_0 = 2$, we get the work done during expansion as $\mathcal{W} = U_0/2 = Q_0^2/4C = CV_0^2/4$, which indeed is the energy that were missing in the two equal–capacitor problem.

The above expression for energy change of the capacitor is quite general and it shows that if $\eta_0 \to \infty$, whole of the capacitor energy goes into the expansion of the plates (again this amounts to loss of all stored energy in (2) for $C_2 \to \infty$). We can look at it in another way. If we were to contract the system ($\eta_0 < 1$), then we (an external agency!) have to do work against the forces of electrical self-repulsion within the capacitor plates. In fact the energy stored in the capacitor is nothing but the work done in bringing the charged capacitor plates from an infinite size to finite dimensions which is essentially the work done in moving the charges from infinity (against their electrical forces of mutual repulsion) to the finite-sized plates of a capacitor.

6 Conclusion

We have shown that the famous paradox of two charged capacitors is successfully resolved if one properly considers all the energy

changes in the system. It was shown that the "missing energy" goes into the kinetic energy of conducting charges when the connecting wire has an identically zero resistance. The problem was formulated in an alternate form, without involving connecting wires in a circuit, where the capacitance of the system is increased by stretching the plates of the original capacitor. The paradox was properly resolved by showing that the work done by the outward self-forces, arising due to mutual repulsion among charges stored within each capacitor plate, during an expansion is equal to the missing energy of the capacitor system. It was also shown that radiation plays no significant role in resolving the paradox.

Acknowledgments

I first learnt of this intriguing paradox from a talk by Prof. S. C. Dutta Roy of IIT Delhi in a conference where he exhorted the audience for a successful solution of this yet unresolved problem of many years, and where he also distributed hard copies of the transparencies of his talk to the interested people.

7 Appendix

7.1 Work done during a stretching of the plates of an ideal capacitor

By an ideal capacitor we mean here that the surface charge density is uniform throughout on both plates. We assume that the charges

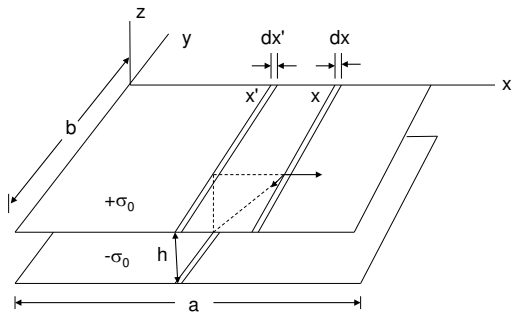


Figure 3: The geometry of the parallel plate capacitor for calculating the forces of self-repulsion within each plate of the capacitor. σ_0 is the surface charge density.

somehow remain "glued" on the surface and the surface charge density decreases as the rubber–like plate surfaces are stretched. Let us assume the plates to be lying in the x-y plane (Fig. 3). The electric field between the plates is parallel to the z-direction. The potential energy of the system as well as the energy in the electrostatic field is $U_0 = 2\pi\sigma_0^2 abh$, where a, b are the plate dimensions and h is the plate separation.

Let us assume that we expand the plate dimensions by say, stretching them along the x-axis. It should be noted that there are electromagnetic forces of *repulsion* on charges *within* each plate, along its surface. We may generally ignore these repulsive forces, but during a stretching of the plates parallel to the plate surface, work will be done by these forces. The forces are indeed small near the plate-centers and become appreciable as we go away from the plate centers, becoming maximum near the plate-edges, and it might

seem that for a and b large enough as compared to h, the effect of these forces should be negligible. But as we will see below, the amount of work done by theses forces during a plate expansion is proportional to the plate dimensions.

As the expansion considered is along the x-axis alone, then only the x-component of the forces of repulsion will be relevant for our purpose. Now the mutual electrostatic force of repulsion between two line charges, each with a linear charge density λ and of a length b, separated by a distance x is easily calculated to be $2\lambda^2(\sqrt{b^2+x^2}-x)/x$.

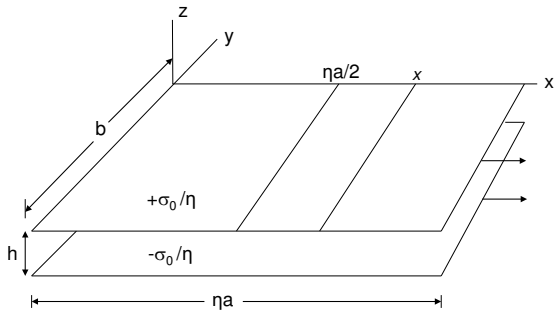


Figure 4: The parallel plate capacitor expanded by a factor η with σ_0/η as the surface charge density.

Accordingly the net force of repulsion on a line charge of linear charge density $\lambda = \sigma_0\,\mathrm{d}x$ lying at x, due to both plates is given by,

$$\mathcal{F}\mathrm{d}x \;=\; 2\sigma_0^2\,\mathrm{d}x\left[\int_0^{2x-a}\mathrm{d}x'\,\frac{\sqrt{b^2+(x-x')^2}-(x-x')}{x-x'}-\int_0^{2x-a}\mathrm{d}x'\,\frac{x-x'}{\sqrt{h^2+(x-x')^2}}\right.$$
$$\left.\cdot\,\frac{\sqrt{b^2+h^2+(x-x')^2}-\sqrt{h^2+(x-x')^2}}{\sqrt{h^2+(x-x')^2}}\right]. \quad (6)$$

Here the second integral term represents the x-component of the force of attraction on the line element at x due to the oppositely charged plate lying at a distance h below (Fig. 3). We have taken the line element at x to be in the right-half of the plate, which experiences a net force towards the +ve x-axis; the left-half of each plate would equally experience a net force along the $-$ve x-axis. Further, only the portion of each plate lying between 0 and $2x-a$ contributes a net force at x, the force due to the remaining portion of each plate gets cancelled because of its symmetry about x.

With a change of variable $x-x'=\xi$, we can write,

$$\mathcal{F} = 2\sigma_0^2\int_{a-x}^{x}\mathrm{d}\xi\;\; g(\xi) \quad (7)$$

where

$$g(\xi) = \frac{\sqrt{b^2+\xi^2}-\xi}{\xi}-\frac{\xi}{h^2+\xi^2}\cdot\left(\sqrt{b^2+h^2+\xi^2}-\sqrt{h^2+\xi^2}\right). \quad (8)$$

Fig. 4 shows a capacitor whose plates have undergone a uniform expansion by a factor η and accordingly the charge density reduced to σ_0/η. Now the charges at x on the expanded plates move an infinitesimal distance $(x-\eta a/2)\mathrm{d}\eta/\eta$ further away with respect to the plate centers, during the change in expansion factor from η to $\eta+\mathrm{d}\eta$.

Then the rate of work being done by the forces of self-interaction, during expansion of *both plates*, is written as ,

$$\mathrm{d}\mathcal{W} = 8\frac{\sigma_0^2}{\eta^3}\mathrm{d}\eta \int_{\eta a/2}^{\eta a} \mathrm{d}x\,(x-\eta a/2)\int_{\eta a-x}^{x} \mathrm{d}\xi\ g(\xi). \tag{9}$$

One factor of 2 in the above expression has entered because an equal work is done on both halves of either plate, while another factor of 2 arose because work is done during expansion of each of the plates.
The rate of work done can be written as

$$\mathrm{d}\mathcal{W} = 4\frac{\sigma_0^2}{\eta^3}\mathrm{d}\eta \int_0^{\eta a} (2x-\eta a) f(x)\,\mathrm{d}x, \tag{10}$$

where

$$f(x) = \int \mathrm{d}\xi\ g(\xi) = \sqrt{x^2+b^2} - x + \sqrt{x^2+h^2} - \sqrt{x^2+b^2+h^2} - b\,\ln\left(\frac{\sqrt{x^2+b^2+h^2}-b}{\sqrt{x^2+h^2}}\cdot\frac{x}{\sqrt{x^2+b^2}-b}\right). \tag{11}$$

With the help of the indefinite integrals,

$$\int \ln\left(\frac{\sqrt{x^2+b^2+h^2}-b}{\sqrt{x^2+h^2}}\right)\mathrm{d}x = x\ln\left(\frac{\sqrt{x^2+b^2+h^2}-b}{\sqrt{x^2+h^2}}\right) + b\ln\left(\sqrt{x^2+b^2+h^2}-x\right) + h\tan^{-1}\frac{bx}{h\sqrt{x^2+b^2+h^2}}, \tag{12}$$

and

$$\int x\,\ln\left(\frac{\sqrt{x^2+b^2+h^2}-b}{\sqrt{x^2+h^2}}\right)\mathrm{d}x = \frac{1}{2}(x^2+h^2).\ln\left(\frac{\sqrt{x^2+b^2+h^2}-b}{\sqrt{x^2+h^2}}\right) - \frac{b}{2}\sqrt{x^2+b^2+h^2}, \tag{13}$$

and after a simplification, we finally get the following expression for the rate of work done during an expansion of the system,

$$\begin{aligned}
\mathrm{d}\mathcal{W} = 4\sigma_0^2 \Bigg[& \frac{2h^2}{3}\left(\sqrt{\eta^2a^2+h^2} - h + \sqrt{b^2+h^2} - \sqrt{\eta^2a^2+b^2+h^2}\right) \\
& -\frac{b^2}{3}\left(\sqrt{b^2+h^2} - b + \sqrt{\eta^2a^2+b^2} - \sqrt{\eta^2a^2+b^2+h^2}\right) \\
& +\frac{\eta^2a^2}{6}\left(\sqrt{\eta^2a^2+h^2} - \eta a + \sqrt{\eta^2a^2+b^2} - \sqrt{\eta^2a^2+b^2+h^2}\right) \\
& +\frac{\eta ab^2}{2}\,\ln\left(\frac{\sqrt{\eta^2a^2+b^2+h^2}-\eta a}{\sqrt{b^2+h^2}}\cdot\frac{b}{\sqrt{\eta^2a^2+b^2}-\eta a}\right) \\
& -\frac{\eta ah^2}{2}\,\ln\left(\frac{\sqrt{\eta^2a^2+b^2+h^2}-\eta a}{\sqrt{b^2+h^2}}\cdot\frac{h}{\sqrt{\eta^2a^2+h^2}-\eta a}\right) \\
& -bh^2\,\ln\left(\frac{\sqrt{\eta^2a^2+b^2+h^2}-b}{\sqrt{\eta^2a^2+h^2}}\cdot\frac{h}{\sqrt{b^2+h^2}-b}\right) \\
& +\eta abh\,\tan^{-1}\frac{\eta ab}{h\sqrt{\eta^2a^2+b^2+h^2}}\Bigg]\frac{\mathrm{d}\eta}{\eta^3}. \qquad (14)
\end{aligned}$$

We can expand this complicated-looking expression in terms of an ascending power series in $h/\eta a$, h/b , $h/\sqrt{\eta^2a^2+b^2}$ as

$$\mathrm{d}\mathcal{W} = 4\sigma_0^2\,\eta abh\left[\frac{\pi}{2} + O(\frac{h}{\eta a},\frac{h}{b},\frac{h}{\sqrt{\eta^2a^2+b^2}})\right]\frac{\mathrm{d}\eta}{\eta^3}, \qquad (15)$$

where $O(\cdots)$ represents the first and higher order power series terms in $h/\eta a, h/b$ etc.

Therefore for $h \ll \eta a, b$, we get,

$$\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}\eta} = \frac{2\pi\sigma_0^2 abh}{\eta^2} = \frac{U_0}{\eta^2}. \qquad (16)$$

References

[1] D. Halliday, R. Resnick and J. Walker, *Fundamentals of Physics*, 5th ed. John Wiley, New Jersey 1997, Ch. 15, 26, 27

[2] R. A. Powel, Am. J. Phys. 47, (1979) 460-462

[3] K. Mita and M. Boufaida, Am. J. Phys. 67, (1999) 737-739

[4] C. Zucker, Am. J. Phys. 23 (1955) 469-469

[5] R.P. Mayer, J.R. Jeffries and G.F. Paulik, IEEE Trans. Education 36 (1993) 307-309

[6] K. Lee, Eur. J. Phys. 30 (2009) 69-74

[7] C. E. Mungan, Eur. J. Phys. 30 (2009) L59-L63

[8] A. M. Abu-Labdeh and S. M. Al-Jaber, J. Electrostatics 66 (2008) 190-192

[9] W. J. O'Connor, Phys. Educ. 32 (1997) 88

[10] S. Mould, Phys. Educ. 33 (1998) 323

[11] A. M. Sommariva, IEE Proceedings - Circuits, Devices and Systems 150, issue 3 (2003) 227-231

[12] D. P. Korfiatis, WSEAS Trans. Circuits and Systems 6 (2007) 76-79

[13] T. B. Boykin, D. Hite and N. Singh, Am. J. Phys. 70 (2002) 415-420

[14] T. C. Choy, Am. J. Phys. 72 (2004) 662-670

[15] E. M. Purcell, Electricity and Magnetism, 2nd ed., McGraw-Hill, New York 1985, Ch. 3, 4, Appendix C

[16] S. Krishnan and M. Rao, Am. J. Phys. 50 (1982), 662

[17] J. D. Jackson, Classical Electrodynamics, 2nd ed., John Wiley, New York 1975, Ch. 14

[18] A. K. Singal, J. Phys. A 25 (1992), 1605–1620

[19] F. T. Trouton and H. R. Noble, Phil. Trans. Roy. Soc. London A 202 (1903), 165-181

[20] A. K. Singal, Am. J. Phys. 61 (1993), 428-433

Euclidean group E(2) and the quantum rotator in three dimensions

Hamideh Rahmati

hrahmati1357@yahoo.com

(*Submitted* 13-07-2015)

Abstract

The paper considers the non-relativistic Schrodinger equation for a particle under the influence of a homogeneous field. The strategy of Hamilton-Jacobi method permits the solutions take the well known results from the classical mechanics. On the other hand, the generators of the Lie group SO(3) for the homogeneous quantum rotator in the large radius limit can be contracted to the Euclidean group E(2) for the motion in the homogeneous field.

Key words: Homogeneous field, Euclidean group E(2), Hamilton-Jacobi method, Contraction of SO(3), Quantum rotator.

Introduction

The motion of a system having two degrees of freedom is said to take place in two dimensions. The simplest motion is the motion of a particle in a homogeneous external field in Cartesian coordinates (x, y, say), with the force F_x, F_y acting on the particle. The potential energy of the particle in the homogeneous field is of the form: $(x, y) = -F_x x - F_y y + constant$; choosing the constant so that $V(x, y) = 0$ for $x = 0, y = 0,$ we have $V(x, y) = -F_x x - F_y y$.

Two interesting cases of the homogeneous fields are an electric field of intensity $\mathcal{E}$ acting on the charge -e with the force $\vec{F} = -e\vec{\mathcal{E}}$, and the other is the gravitation field in the y coordinate with the force $F_x = 0, F_y = -mg$. Projectile motion constitutes a very elementary problem of classical mechanics since the works done by Galoleo Galilei and Sir Issac Newton. The projectile motion is one of the main problems used to teach elementary physics and not well known facts about it appear in physics literature by now. Many of these ideas are presented in a compelling paper by Groetsch [1], Rao [2] and [3]. Also, the motion of the charge in the homogeneous field $\mathcal{E}$ is the elementary problem in the electromagnetic field too [4].

Applications of group theory in physics establish the standard framework for the application of geometric symmetry groups to the treatment of Quantum mechanics systems that possess some geometric symmetry. Contraction is a process to reparameterize the Lie group's parameter space in such a way that the group properties in the Lie algebra remain well defined. The parameter space for the contracted group, Casimir operators, matrix elements of operators are well defined. Also, contraction provides limiting relations among the special functions of mathematical physics.

One of the interesting group contractions is the group SO(3) that is contracted to the Euclidean group E(2). The internal space-time symmetries of massive and massless particles are isomorphic to SO(3) and E(2) respectively. It has been shown that [5] transverse rotational degrees of freedom for massive particles become contracted to gauge degrees of freedom for massless particles.

In this work we investigate the motion of the particle in the homogeneous field in two directions. In particular, it is shown that by using the Hamiltonian-Jacobi method we get the well known results from the classical mechanic. Then, we show that quantum rotator on the sphere S^2 with Lie group SO(3) can be contracted to the motion of the particle with Euclidean group E(2) on the plane R^2.

The paper is organized as follows: we begin by proposing the Schrodinger equation for the motion of a particle in the homogeneous field in two dimensions and the limiting behavior of it, is investigated which goes to the well known results from the classical mechanics. In section 3, we will see that Quantum projectile yields from the contraction of the Quantum rotator and its eigenvalues and eigenfunctions take the form of the motion in the homogeneous field. Finally, we end the paper by conclusion.

2-1: Schrodinger equation for a particle in the homogeneous two dimensional fields

In this section, a particle in the homogeneous field is considered as a point mass *M* that its equation of motion must be solved by the Schrodinger equation. Since, the latter has no spatial extensions, this amount to neglecting issues of shape, orientation and rotation altogether. Furthermore, we suppose lift forces to be negligible too.

This motion has 2 degrees of freedom in the plane and can be considered as a problem with the classical group E(2) [6]. This group consists of all the transformations R_θ on R^2 plane. In other words, the Euclidean group E(2) consists of matrices of the form $\begin{bmatrix} cos\theta & -sin\theta & x \\ sin\theta & cos\theta & y \\ 0 & 0 & 1 \end{bmatrix}$. This group has three infinitesimal generators L_z, P_x, P_y as following:

$$L_z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad P_x = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad P_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \tag{2-1}$$

With the commutation relations:

$$[P_x, P_y] = 0, [L_z, P_x] = 0, [L_z, P_y] = 0. \tag{2-2}$$

The time independent Schrodinger's equation for a particle M in the *X*-*Y* Cartesian plane is:

$$H\psi(x,y) = E\psi(x,y), H = \frac{P_x^2+P_y^2}{2M} + V(x,y), \quad (2\text{-}3)$$

Where by replacing the momentum operators: $P_x = -i\hbar\partial_x$ and $P_y = -i\hbar\partial_y$ we have:

$$-\frac{\hbar^2}{2M}\left(\frac{d^2}{dx^2}+\frac{d^2}{dy^2}\right)\psi(x,y) + V(x,y)\psi(x,y) = E\psi(x,y) \quad (2\text{-}4)$$

We consider the coordinates are completely separable as following:

$$\psi(x,y) = X(x)Y(y), V(x,y) = V(x) + V(y), \quad E = E_x + E_y, \quad (2\text{-}5)$$

So, the Schrodinger equation separates into two terms:

$$\left(-\frac{\hbar^2}{2M}\frac{d^2}{dx^2} + V(x)\right)X(x) = E_x X(x), \quad (2\text{-}6)$$

$$\left(-\frac{\hbar^2}{2M}\frac{d^2}{dy^2} + V(y)\right)Y(y) = E_y Y(y), \quad (2\text{-}7)$$

These equations determine the motion of a particle in the x- and y-coordinates completely. By choosing appropriate variables the equation (2-6), (2-7) are:

$$\left[\frac{d^2}{d\acute{q}^2} + (E_{\acute{q}} - \acute{q})\right]Q(\acute{q}) = 0, \quad q = 1,2. \quad (2\text{-}8)$$

Where $\acute{q} = \frac{q}{l}$ and $l = \left(\frac{\frac{\hbar^2}{2M}}{F}\right)^{\frac{1}{3}}$; also $E_{\acute{q}} = \left(\frac{\hbar^2}{2M}F^2\right)^{\frac{1}{3}}$ and F is the force acting on the particle. Since the potential energy depend on the coordinates, it is clear that the energy levels form a continuous spectrum occupying the whole range of energy values E from $-\infty$ to $+\infty$. None of these eigenvalues is degenerate, and they correspond to motion which is finite. The equation (2-6), (2-7) do not contain the energy parameter. Hence, if we obtain a solution of them, we at once have the eigenfunctions for arbitrary values of the energy. Their solutions, which are finite for all q coordinate is called the Airy function [7].

Since, we'll be using a superposition of several single energy solutions, we must go to the general time dependent Schrodinger's equation, which has solutions like the usual time independent Schrodinger equation. So, by using linear combinations of time dependent solutions, we have:

$$\psi = NAi\left(-\acute{q} + E_{\acute{q}}\right)e^{-iE_{\acute{q}}\frac{t}{\hbar}}, \quad (2\text{-}9)$$

Where N could be determined from the normalization condition.

Two special cases for the potential energy are the gravitational field and the electric field. For gravitational field in the y-direction we have $(x) = 0, V(y) = Mgy$. In this case the differential equation (2-6) is a familiar equation for us, with sinusoidal solution like $sin(kx)$ or $exp(ikx)$, where $k^2 = E_x$ and the complete solution is:

$$\psi = NAi\left(-\acute{y} + E_{\acute{y}}\right)e^{-iE_{\acute{y}}\frac{t}{\hbar}}\, exp(ikx)e^{-iE_x\frac{t}{\hbar}}, \qquad (2\text{-}10)$$

The other case is the electric field in two directions, and the potential energy are: $(x) = -e\mathcal{E}_x x, V(y) = -e\mathcal{E}_y y$, where $\mathcal{E}$ is the electric field and we have two Airy function for the solution in the x- and y-directions.

2-2: The Hamiltonian-Jacobi method and the well known classical results

In this sub section we see that previous solution is completely compatible with the classical solution for semi classical wave function by using the Hamilton-Jacobi method [8]. Our Hamiltonian is:

$$H = \frac{P_x^2 + P_y^2}{2M} + V(x) + V(y) = E_x + E_y. \qquad (2\text{-}11)$$

We obtain the Hamilton-Jacobi equations for S (action) by replacing P_x by $\frac{\partial S_x}{\partial x}$ and P_y by $\frac{\partial S_y}{\partial y}$ in the Hamiltonians (2-7), (2-8). Since our Hamiltonian is separable we obtain two equations:

$$\frac{1}{2M}\left[\left(\frac{\partial S_q}{\partial q}\right)^2 + V(q)\right] + \frac{\partial S_q}{\partial t} = E_q, \quad q = 1,2. \qquad (2\text{-}12)$$

The explicit dependence of S_x and S_y on t is present only in the last terms, so, for S_q we have:

$$S_q = \pm \int \sqrt{2M\left(E_q - V(q)\right)}\, dq - E_q t, \qquad (2\text{-}13)$$

Now, if we choose the particle is in the gravitational field in the y-direction we have for S_x:

$$S_x = \pm\sqrt{2ME_x}x - E_x t, \quad \textit{(2-14)}$$

And for S_y we also have:

$$S_y = \pm\int \sqrt{2M(E_y - Mgy)}\,dy - E_y t,$$

$$= \mp\frac{2}{3g}\sqrt{\frac{2}{M}}(E_y - Mgy)^{\frac{3}{2}} - E_y t. \quad (2\text{-}15)$$

To get out the explicit solutions of x and y coordinates we have:

$$\frac{\partial S_x}{\partial E_x} = \frac{\sqrt{2M}}{2}E_x^{-\frac{1}{2}}x - t = \alpha_1 \Rightarrow$$

$$x = \sqrt{\frac{2E_x}{M}}(\alpha_1 + t), \quad (2\text{-}16)$$

Where α_1 is an arbitrary constant, and we can find it by using initial conditions of the problem. For simplicity at initial time $t = 0$ we take it $\alpha_1 = 0$ and E_x is the kinetic energy $\frac{1}{2}mv_{0x}^2$, so for x equation we have:

$$x = v_{0x}t, \quad (2\text{-}17)$$

And for y coordinate:

$$\frac{\partial S_y}{\partial E_y} = -\frac{1}{Mg}\sqrt{2ME_y - 2M^2gy} - t = \alpha_2 \Rightarrow$$

$$y = \frac{2E_y}{2Mg} - \frac{1}{2}(\alpha_2 + t)^2, \quad (2\text{-}18)$$

Again α_2 at initial time $t = 0$ is a constant and we take it equal zero, and $E_y = \frac{1}{2}Mv_{0y}^2 + Mgy_0$, so, we obtain:

$$y = -\frac{1}{2}gt^2 + v_{0y}t + y_0. \quad (2\text{-}19)$$

On the other hand the action is:

$$S = S_x + S_y =$$

$$\pm\sqrt{2ME_x}x \mp \frac{2}{3g}\sqrt{\frac{2}{M}}(E_y - Mgy)^{\frac{3}{2}} - Et, \quad (2\text{-}20)$$

And semi classical wave function is:

$$\psi = ae^{\frac{i}{\hbar}S} = ae^{\frac{i}{\hbar}\left[\pm\sqrt{2ME_x}x \mp \frac{2}{3g}\sqrt{\frac{2}{M}}(E_y - Mgy)^{\frac{3}{2}} - Et\right]}, \quad (2\text{-}21)$$

Where a is the normalization parameter. Now, we want to compare the relation (2-21) with the wave function (2-10). So, by replacing the asymptotic relation for Airy function, i. e. $Ai(z) = e^{-\frac{2}{3}z^{\frac{3}{2}}}$ and ignoring the normalizing parameters $"N"$ and $"a"$, the right hand side of two relations (2-10) and (2-21) are the same.

For the electric field in two directions we obtain:

$$S_q = \pm \int \sqrt{2M\left(E_q - e\mathcal{E}_q q\right)}\, dq - E_q t,$$

$$= \mp \frac{2}{3}\sqrt{\frac{2M}{e\mathcal{E}_q}}\left(E_q - e\mathcal{E}_q q\right)^{\frac{3}{2}} - E_q t. \qquad (2\text{-}22)$$

By choosing appropriate conditions, we could obtain the well known results for this case too.

3: Quantum rotator and its relation to the Euclidean motion in the gravitational field

A rigid body may be defined in mechanics by the angular momentum operator J, which has the Lie group SO(3) on the sphere S^2. The generators of this group are:

$$J_i = x_j\partial_k - x_k\partial_j \qquad i,j,k = 1,2,3. \qquad (3\text{-}1)$$

The commutation relations are $[J_i, J_j] = \epsilon_{ijk}J_k$.

The Hamiltonian of a homogeneous rigid body is [8]:

$$H = \frac{J_x^2 + J_y^2 + J_z^2}{2I} + V(\theta) = \frac{J^2}{2I} + V(\theta), V(\theta) = MgRsin\theta = J_z\dot{\varphi}sin\theta \qquad (3\text{-}2)$$

Where I is the moment of inertia, the parameter θ is the polar angle and φ is the azimuthal angle. We are familiar with the corresponding eigenvalues and eigenfunctions as follows:

$$\left|\psi\right\rangle \propto P_m^j\left(cos\theta\left(=\frac{y}{R}\right)\right) = \left|{}^{j}_{m}\right\rangle, \quad J^2\left|{}^{j}_{m}\right\rangle = \hbar^2 j(j+1)$$

$$J_z\left|{}^{j}_{m}\right\rangle = m\hbar\left|{}^{j}_{m}\right\rangle. \qquad (3\text{-}3)$$

Under contraction with respect to the subalgebra of rotations about the Z-axis and considering the limit $R \to \infty$, the operators J_x, J_y transform to the operators $-P_y, P_x$ respectively. So, the commutation relations of the contracted group, i. e., the group E(2) are:

$$[J_z, P_x] = -P_y, \ [J_z, P_y] = P_x, \ [P_x, P_y] = 0. \qquad (3\text{-}4)$$

This group consists of rotations about the Z-axis, and displacements of the origin in the x- and y-directions. So, we have:

$$J^2 \to P_x^2 + P_y^2,\ J_z \to P_z,\ \ j(j+1)\hbar^2 \to p^2,\ \left|{}^{j}_{m}\right\rangle \to \left|{}^{p}_{m}\right\rangle,\ \ I \to M, \tag{3-5}$$

And its wave function $P_m^j(cos\theta) = \sqrt{2\pi}e^{im\varphi}Y_m^j(\theta,\varphi)$ is transformed to the $\sqrt{2\pi}e^{im\varphi}J_m(y)$, where $J_m(y)$ is the Bessel function [6]. This answer is equal to the wave function (2-10), because the Airy function could be written in terms of the Bessel functions [7]. For the potential part of the Hamiltonian under contraction we have:

$$V(\theta) \to P_z\dot{\varphi}sin\theta = Mgt\dot{\varphi}sin\theta, \tag{3-6}$$

Where we use the classical mechanics to compute P_z. If we replace $\theta = y/R$ and take the limit $R \to \infty$ for the relation (3-6), we conclude that:

$$V(\theta) \to V(y) = Mgy, \tag{3-7}$$

So, the contracted Hamiltonian is:

$$\hat{H} = \frac{p^2}{2M} + V(y). \tag{3-8}$$

This means that by changing the corresponding space of the motion from S^2 to R^2 by the contraction, the rigid Quantum rotator transformed to the projectile motion.

Lorentz group with respect to the time coordinate yields the homogeneous Galilei group. Contraction of the de Sitter group yields the inhomogeneous Lorentz group and these are because of the great magic power of the contraction!

4: Conclusion

Many physical systems exhibit symmetry. So, it is possible to use group theory and algebra to simplify both the treatment and the understanding of the problem. Central two-body problems, such as the gravitational and coulomb interactions, give rise to systems exhibiting spherical symmetry (two particles) with the Lie algebra so(3) or broken symmetry (planetary systems) with the Lie algebra e(2).

In this paper the motion of a particle in the homogeneous field, such as the gravitational or electric field, is analyzed with the standard Quantum and semi-classical method. It has also indicated that the Quantum mechanics behavior of

the eigenfunctions and eigenvalues transformed to the classical results in the limiting case.

We have shown that the contraction of the group SO(3) takes our rotator on the sphere S^2 to the projectile motion on R^2 plane, i. e., Euler rotations for our rotator under contraction take it to the simple motion on a flat space R^2.

We would like to end our paper with some words from Galileo:

The great book of Nature lies ever open before our eyes and the true philosophy is written in it, but we cannot read it unless we have first learned the language in which it is written. It is written in mathematical language and the characters are triangles, circles and other geometrical figures [9].

5: References

1- C. W. Groetsch, Tartaglia's inverse problem in a resistive medium, *Amer. Math. Monthly,* **103**, 546-551, 1996.

2- K. N. Srinivasa Rao, The motion of a falling particle in a Schwarzschild field, *Ann. Inst. Henri Poincare*, section A, **3**, 227-233, 1966.

3- R. D. H. Warburton and J. Wang, Analysis of asymptotic motion with air resistance using the Lambert W function, *Am. J. Phys*. **72**, 1404-1407, 2004.

4- L. D. Landau and E. M. Lifshitz, *A shorter course of theoretical physics*, **Vol. 1**, first ed., pergamon press, 1972.

5- Y. S. Kim, Group contractions: Inonu Wigner and Einstein*, Int. J. Mod. Phys. A* **12**, 71-78, 1997.

6- R. Gilmor, *Lie Groups, Lie Algebra and Some of Their Applications*, Wiley, New York, 2006.

7- L. D. Landau and E. M. lifshitz, *Quantum mechanics non-relativistic theory*, second ed., **Vol. 3**, pergamon press, 1965.

8- H. Golestein, C. Poole and J. Satko, *Classical Mechanics*, Addison-Wesley, 2000.

9- F. Copleston, *A History of Philosophy*, **Vol. III**, Doubleday, New York, 1963.

Revisiting the Concept of 2-D Bravais Lattices

Jyoti Bhardwaj[1],Vandana Sharda[2], O.S.K.S. Sastri[1,a] & Arbind K.Jha[3]

[1]Department of Physics & Astronomical Sciences
Central University of Himachal Pradesh
Dharamshala 176215, India.

[2]Department of Teacher Education
Central University of Himachal Pradesh
Dharamshala 176215, India.

[3]School of Education
MG International Hindi University
Wardha, Maharashtra 442005,India.
[a]sastri.osks@gmail.com

(Submitted 12- 06- 2015)

Abstract

In this paper, the concept of 2-Dimensional (2-D) Bravais Lattices is being arrived at using a constructivist approach, which is similar to the Socratic method of inquisitive questioning followed up with analysis and activity to obtain a comprehensive understanding of the idea. By posing the question as to how to construct a lattice by repeating a fundamental unit, we have created an activity that involves playing with triangles and quadrilaterals and figuring out which of them could easily tessellate to cover the 2-D chart without leaving any gaps. This lead to the construction of the five possible 2-D Bravais Lattices.

1 Introduction

2-D Bravais lattices were elementary and unrealisable theoretical constructs till the re-

alisation of Graphene in 2004 [1], which was the first 2-D crystalline structure that is stable and has shown tremendous promise in a very short time. Various patterns on wallpapers [2] have exploited the possibility of periodic or symmetrical arrangements in 2 dimensions. Now many more 2-D crystalline structures such as NbS_2, MoS_2 etc. are being realised [3] [4].

On questioning students, it was found that recalling the five 2-D Bravais Lattices, after certain lapse of time post examinations is difficult and putting together all the 14 3-Dimensional (3-D) Bravais Lattices is definitely a daunting task at the undergraduate level.

There is a new found enthusiasm in the Physics community to innovate and explore ways of teaching fundamental concepts, wherein, one tries to systematically identify the alternative conceptions and gaps in learning and develop strategies to overcome them.This process has come to be known as Physics Education Research (PER) [5]. Further, new paradigms in education pedagogy such as Constructivism [6] and Constructionism [7] have paved way for learning concepts (in Physics) through a process of discovery by the students based on their previous knowledge, interaction among the peers and the guidance of the teachers. In this paper, an effort has been made to use PER strategies to develop modules which would help students to construct their own knowledge of 2-D Bravais Lattices, a preliminary concept in Solid State Physics.

Based on Bloom's taxonomy, Sapna et. al. [8] have devised learning/instructional objectives for Bravais Lattices in the context of Crystal Structure.The very first ideas of lattice and basis which form the foundation of crystallography require previous knowledge of symmetry and periodicity. 2-D and 3-D Bravais Lattices are abstractions that result from the classification of all the possible lattice structures.

Here, we are proposing a simple activity to realise the following learning objectives :

1. **Which is the smallest unit that can cover all space to form a lattice?**

2. **What are the distinct types of lattices in 2-D?**

3. **Which is the most convenient way to obtain them?**

2 Methodology

In this paper, we use the constructivist approach wherein we would want to arrive at the learning objectives through a strategy similar to Socratic method of questioning why, what and how followed by analysis, activity and comprehension in a repetitive manner till we gain complete clarity of the desired concept.

The implementation strategy for constructing the knowledge of 2-D Bravais Lattices includes the following steps:

1. firstly capturing the interest of the students by showing them some interesting pictures from the current science,

2. arousing their curiousity by posing a challenge and then raising questions to make them inquisitive,

3. putting them to activity to explore for themselves,

4. encourage them to share their findings with each other,

5. putting together the outcomes, analysing the findings along with their consensus and

6. finally arriving at the comprehensive understanding of the topic and applying it to solve the posed challenge.

2.1 Formulating the Question

Keen observation is the key to asking appropriate questions. We take advantage of current progress in science & technology and propose to look at Scanning Tunneling Microscope (STM) picture of Graphene as shown in Figure 1(a). It is interesting to note that Graphene has a hexagonal structure which is not one of the standard 2-D Bravais Lattices. Posing it as a challenge in the classroom and asking the students to identify the underlying 2-D Bravais Lattice and basis of Graphene, can prove to be a fruitful activity to probe student understanding. This paves way for starting the discussion with the question: ***Why is it that the hexagonal lattice not one of the five 2-D Bravais Lattices?***

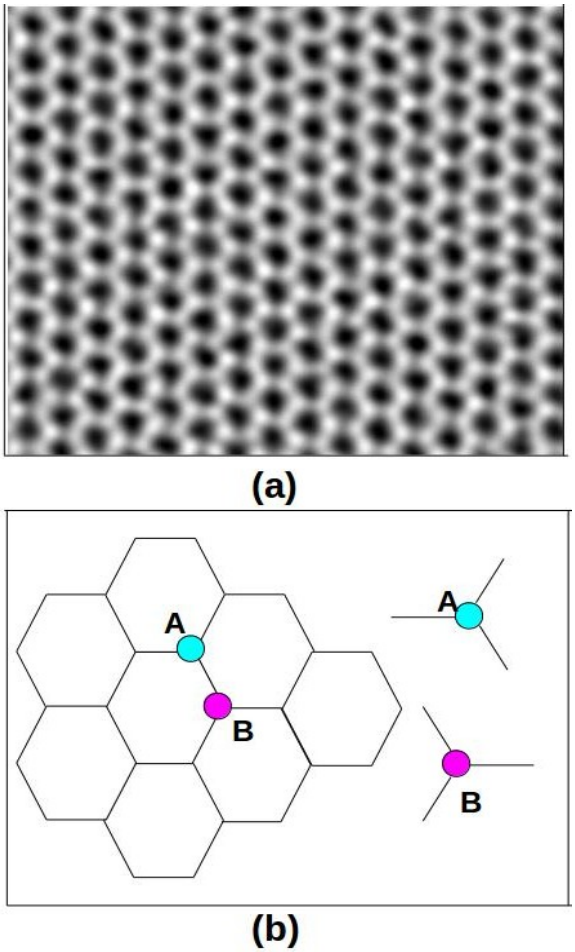


Figure 1: (a)STM image of Graphene. (*source:http://arxiv.org/pdf/1009.4714.pdf*)
(b)In the structure of Graphene, shown are two points A and B. The honeycomb lattice observed is not a Bravais Lattice as the environment around points A and B differ by an angle of 60^o [9] [11]

A lattice is a periodic array of points which ensures a similar environment from any given point all around [9].The vertices of such fundamental units which can build all possible periodic arrangements have come to be

known as Bravais Lattices, named after the person who deduced them [10].
The Graphene structure shown in Figure 1(b), clearly depicts that the environment around points A and B is not same, thus ruling out a hexagonal structure to be a possible Bravais Lattice [11]. This brings us to the following questions:

1. **What are the possible periodic arrangements in 2-D which preserve the property of having exactly same environment at each location?**

2. **What are the various Bravais Lattices, how to obtain them and where do we begin?**

2.2 Building the Activity

The clue is to identify the smallest unit area that can be enclosed by the minimum number of lattice points, which when repeated covers all 2-D space without leaving any gap. Minimum three non-collinear points are required to cover an area in 2-D and joining them gives us a triangle, which can be chosen as the smallest unit for the purpose. From previous knowledge, we know that there are only seven different types of triangles possible as shown in column 1 of Figure 2. The activity proposed here involves trying to cover a chart paper using these 7 types of triangles by a combination of reflection, rotation and/or translation operations. The entire class can be divided into various groups, each of them working with either 1 or

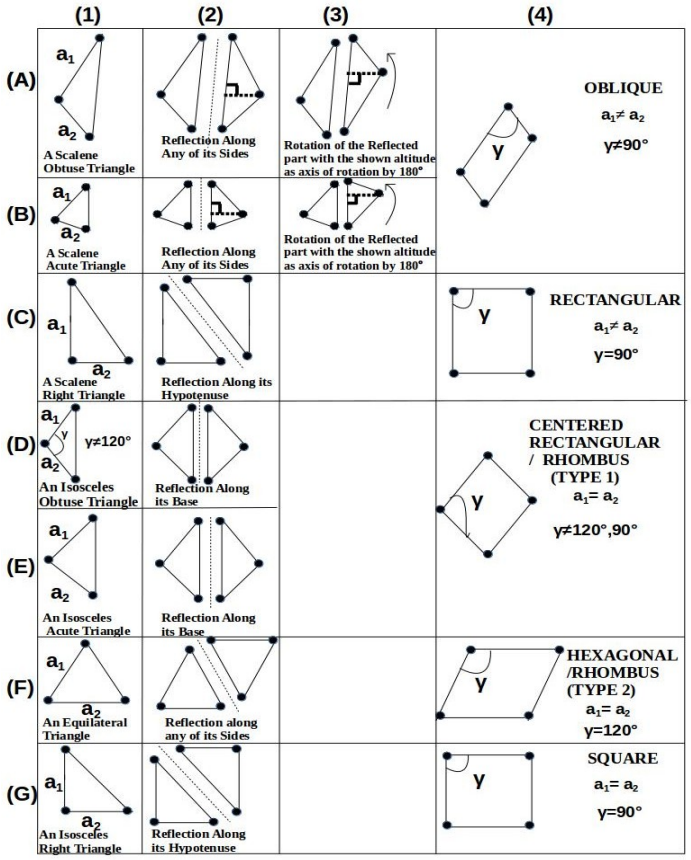


Figure 2: Depicting how five 2-D Bravais Lattices can be obtained using various types of triangles (classified on the basis of sides and angles). Though there may be many other possibilities of obtaining the lattices through triangles, but the (apparent) first intuitive method which incorporates basic symmetry operations (reflection/rotation around a convenient axis or both) has been employed. The dots (or the vertices of the triangles) depict the position of the lattice points in the lattice they constitute. While columns depict the type of triangles and the basic transformations employed, the rows constitute how various symmetry operations on different types of triangles lead to the formation of five 2-D Bravais lattices.

2 types of triangle cut-outs to fill the charts provided to them. The outcome of the activity can then be discussed in the classroom.

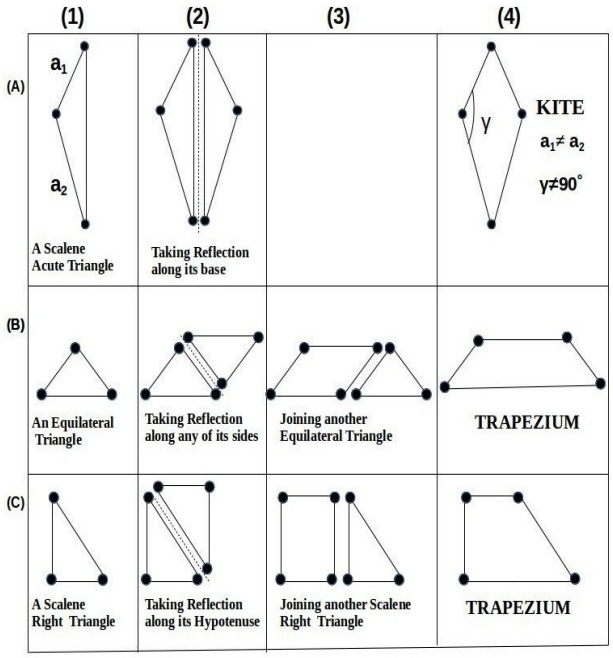


Figure 3: Depicting how triangles can form quadrilaterals which are not Bravais lattices (i.e. when they are tessellated over a given space, they will not be able to cover it wihout leaving gaps)

3 Discussion of Results

Triangles meet the requirement of the first learning objective as they are the smallest unit that can cover all space to form a lattice. Covering all space using triangles required all three operations of reflection, rotation and translation. This is a tedious process. On looking carefully at the output charts, we find that quadrilaterals are the ones which can be tessellated(i.e. translating a unit of area in order to cover a given space without leaving any gaps) along two directions to cover all space. Column 4 of Figure 2 shows those quadrilaterals that can be tessellated to cover all space. Those that do not satisfy the requirement of covering all space without gaps are given in column 4 of Figure 3.

4 Comprehension

The activity shows that there are only five possible quadrilaterals

1. **Oblique** ($a1 \neq a2, \gamma \neq 90^o$),
2. **Rectangular** ($a1 \neq a2, \gamma = 90^o$),
3. **Rhombus,Type-1 (obtained using isosceles obtuse/acute triangles giving** $a1 = a2, \gamma \neq 120^o/90^o$)
4. **Rhombus,Type-2 (obtained using equilateral triangle giving** $a1 = a2, \gamma = 120^o/60^o$)
5. **Square** ($a1 = a2, \gamma = 90^o$),

These are the appropriate fundamental units which when tessellated would cover all the space and could be thought of as primitive unit cells in 2-D. The two rhombuses can be alternatively visualised as centered regular hexagonal lattice and centered rectangular lattice as shown in Figure 4. In standard textbooks [9] [12] though, the connotation used for Rhombus, Type-2 is hexagonal

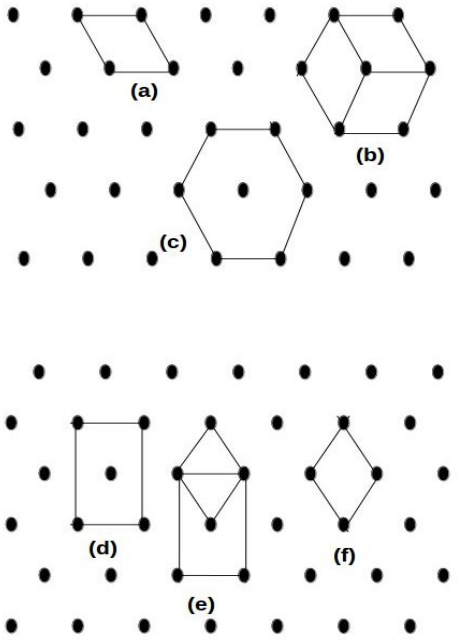


Figure 4: Figures showing equivalence conditions for two of the Bravais Lattices.
(a)shows a Rhombus, Type-2 (adjacent angles 60^o and 120^o) amidst the lattice
(b)shows how the Rhombus, Type-2 is reflected and rotated to obtain the centered regular hexagonal structure
(c)shows the Centered Regular Hexagonal amidst the lattice.
(d)shows a centered rectangle amidst the lattice.
(e)shows how the Rhombus, Type-1 is also inherent in the centered rectangle.
(f)shows the Bravais Lattice, Rhombus, Type-1.

which is not correct and can confuse students. These are non-primitive unit cells but are more conventionally employed and are typically part of the five 2-D Bravais Lattices as they are known and are shown in Figure 5 for sake of completion. Now, coming back to

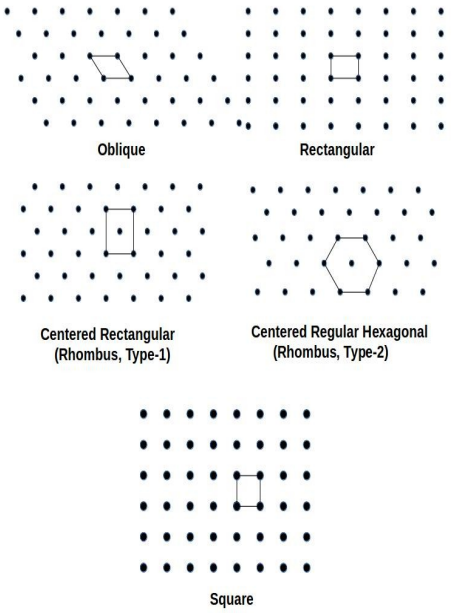


Figure 5: The five 2-D Bravais Lattices.

Graphene, we notice that it is a **Rhombus, Type-1** ($a1 = a2, \gamma \neq 120^o/90^o$) lattice with two C-atoms forming a basis as shown in Figure 6.

5 Conclusion

We have developed an activity based strategy involving symmetry oopreations on triangles which are the smallest unit that could cover all space. Based on the obseravtions and analysis of the outcomes, quadrilaterals were found to be more convenient as they involve only tessellation.

The combination of all seven possible triangles that resulted in various quadrilaterals

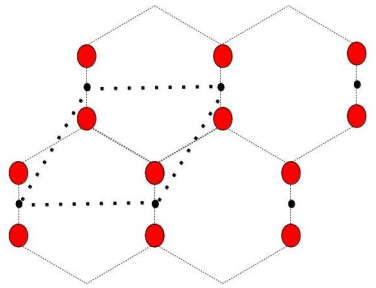


Figure 6: The dotted parallelogram, Rhombus Type-1 shows the underlying Bravais Lattice of Graphene structure with C-C basis(filled circles) at each of its vertices, the lattice points.

were tabulated into two groups based on the criteria whether the given quadrilateral tessellate to cover all space without leaving any gaps. The group of quadrilaterals that satisfy the criteria form the five primitive cells that generate all possible lattices in 2-D. We have shown that primitive cells consisting of Rhombus, Type-1 and Rhombus, Type-2 are equivalent to the non-primitive more conventionally used unit cells of centered rectangular and centered regular hexagonal lattices respectively and thus arrived at the five 2-D Bravais lattices as they are known today.

References

[1] Novoselov et.al., "Electric field effect in atomically thin carbon films", Science, 306(5696), 666-669, 2004.

[2] Wallpaper group. (2015, April 28). In Wikipedia, The Free Encyclopedia. Retrieved 06:36, June 8, 2015, from *http* : *//en.wikipedia.org/w/index.php?title* = *Wallpaper_group&oldid* = 659657015

[3] Novoselov et.al, "Two-dimensional atomic crystals".*PNAS 102*, 10451-10453(2005). Retreieved from *www.pnas.org/content*/102/30/10451 *.full.pdf*

[4] J.N. Coleman, M. Lotya, A. ONeill, et al., Science 331, 568 (2011).

[5] R. J. Beichner, "An Introduction to Physics Education Research", Getting Started in PER, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2009), Reviews in PER Vol. 2, < *http* : *//www.per* − *central.org/items/detail.cfm?ID* = 8806 > .

[6] Thirteen Ed Online (2004). Constructivism as a paradigm for teaching and learning. Retrieved from *http* : *//www.thirteen.org/edonline /concept2class/constructivism/*

[7] Constructionism. (2014, March 28). from EduTech Wiki, A resource kit for educational technology teaching, practice and research. Retrieved 05:24, June 8, 2015 from *http* : *//edutechwiki.unige.ch/mediawiki/ index.php? title* = *Constructionism&oldid* = 52791.

[8] Sharma S., Sastri O.S.K.S., Ahluwalia P. K., "Design of instructional ob-

jectives of undergraduate solid state physics course: a first step to physics education research", International Conference on Physics Education: ICPE2009 (Vol. 1263, No. 1, pp. 171-174), AIP Publishing, 2010.

[9] Kittel C. & McEuen P. "Introduction to Solid State Physics", Vol. 8, Wiley(New york), 1976.

[10] Burke, "Origins of the science of crystals", 1966, p.171.

[11] Chattopadhyay K. K. "Introduction to Nanoscience and Nanotechnology", PHI Learning Pvt. Ltd,2009.

[12] Ashcroft N. W. Mermin N. D., "Solid state physics", Saunders, Philadelphia, 293, 1976.

Satellite in a Circular Orbit About a Rotating Spherical Planet and Calculation of the Velocity Change Along the Ground Track

Jean C. Piquette

72 Botelho Drive, Portsmouth, Rhode Island 02871 USA

jpiquette@verizon.net

(*Submitted* 30-04-2015)

Abstract

An object in frictionless contact with and moving along a rotating spherical surface will experience Coriolis and centrifugal accelerations in the rotating frame. Although these are the only accelerations that appear explicitly in the equations of motion, assuming no other tangential physical forces are present, velocity changes of the object over a finite time interval cannot be correctly computed by integrating the sum of only these two accelerations over that time interval. This was proven in a recent publication [J.C. Piquette, "Velocity Change Calculation for an Object Moving on a Rotating Spherical Surface," *Phys. Educ.* 31(1) (2015), art. num. 3, pp. 1-12]. It was found there that an unexpected additional acceleration, therein termed the "kinematic" acceleration, was also required to be integrated over the finite time interval in order to deduce the correct velocity change. Interestingly, a satellite in circular orbit about a spherical rotating planet satisfies everything required for the results of this previous work to apply. Hence, for example, the change in velocity of such a satellite, as seen in the rotating frame, cannot be determined by integrating over only the sum of the Coriolis and centrifugal accelerations. It was also found in the earlier work that the influence of the kinematic acceleration is dominant for high initial object speeds. The kinematic acceleration dramatically dominates both the Coriolis and centrifugal accelerations in the case of a satellite in circular near-Earth orbit, since such a satellite has a speed of about 18000 miles/hour. These conclusions also apply to the calculation of velocity changes along the ground track. To permit detailed understanding of the satellite's motion along the ground track, the notion of a "shadow satellite" is introduced. The results and examples given here can be used in an undergraduate- or graduate-level classical mechanics course as modern space-age applications of classical mechanics that may be of high interest to students.

1. Introduction

The problem of computing the velocity change over a finite time interval of an object moving on the surface of a rotating sphere was considered in a recent publication [1]. In the problem considered there, the object was taken to be free of any applied tangential physical forces, and was constrained to remain on the sphere's surface. Naturally, from the perspective of the rotating frame, the object experiences both Coriolis and centrifugal accelerations. Indeed, these two accelerations are the only accelerations that appear explicitly in the equations of motion. But despite that fact, if one attempts to compute the change of velocity of the object over a finite time interval by integrating over only the sum of these two accelerations, an incorrect result is obtained. By integrating the two equations of motion over the finite time interval of interest, it was found that a

third acceleration component appears. That third component was termed the "kinematic" acceleration.

Interestingly, a satellite in circular orbit obeys all the requirements considered in the previous work, and hence the same conclusions apply to a satellite in such an orbit. An example of such a satellite that is especially interesting is one that orbits the Earth and passes over the two poles [2]. The ground track of such a satellite carries it over a very large percentage of the Earth's surface.

The influence of the kinematic acceleration was found to become increasingly dominant over the Coriolis and centrifugal accelerations as the initial velocity of the object increases. For an object in near-Earth orbit, which travels at a speed of about 18000 miles/hour, the kinematic acceleration is by far the largest of the three acceleration terms that contribute to changes in velocity.

Here, applications of the results of Ref. 1 to satellites in circular orbit are considered. For simplicity, in example calculations involving the Earth, it is assumed the Earth is a perfect sphere of radius 4000 miles, and completes one rotation in exactly 24 hours. It is also assumed that the orbital velocity of a satellite at the approximate POES [2] altitude of 700 miles above the surface is exactly 18000 miles/hour.

It is hoped that the results and examples given here may be useful in either an undergraduate- or graduate-level classical mechanics course. As space-age applications of classical physics, these examples may be of high interest to students. Those who would like to see additional references related to non-inertial frames and relating the material to classroom teaching are directed to the larger list of references given in Ref. 1.

In Sec. 2, the two coordinate systems of interest are described, and the equations of motion are presented. A brief summary of the velocity-change calculation developed in Ref. 1 is given in Sec. 3. It is shown in Sec. 4 that the solution of an object given an initial velocity at the equator can actually be applied to cases with more generality than initially considered. The concept of the shadow satellite, which is useful for studying the satellite's ground track, is developed in Sec. 5. Numerical examples are given in Sec. 6. A summary and conclusion are given in Sec. 7.

2. Coordinate Systems and Equations of Motion

The approach used here applies the results of Ref. 1 to a satellite in a circular orbit about a spherical rotating planet. The results of that reference apply directly to the satellite problem if the sphere radius is simply replaced by the radius of the circular orbit. Of course, when applied in that way the rotating sphere is actually an imaginary mathematical surface, rotating with the same angular speed as the planet, over which the satellite is assumed to be moving. Also of interest is the ground track of the satellite, and the notion of a "shadow satellite" is introduced for studying the ground track. The shadow satellite is taken to be a physical object located on the rotating planet's surface, with the planet assumed airless and frictionless. The shadow satellite moves along the surface of the rotating planet, and always remains directly underneath the orbiting satellite.

Two coordinate systems are used in the analysis. These are termed the "unprimed" and "primed" coordinate systems. The systems are depicted in Fig. 1.The unprimed coordinate system is an inertial frame at rest with respect to the fixed stars. The Cartesian coordinates of this system are denoted (x, y, z). Not shown is a related unprimed spherical coordinate system (r, θ, ϕ). However, the angle ϕ of this system is depicted,

and expresses the angle between the x and x' axes, and between the y and y' axes. The primed system also consists of Cartesian and spherical coordinates as shown. However, the primed system rotates about the common z, z' axes at constant angular speed $\dot{\phi} = \omega$.

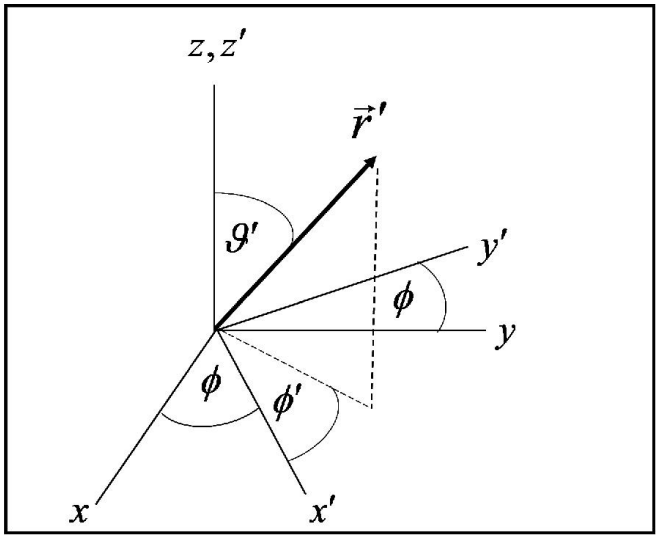


Fig 1 Primed and unprimed coordinate systems

In the primed coordinate system, the well-known equations of motion of an object moving on the surface with no physical tangential forces can be expressed as [3]

$$r'\ddot{\theta}' - r'\dot{\phi}'^2 \sin\theta' \cos\theta' = \left[a_{COR}\right]_{\theta'} + \left[a_{CENT}\right]_{\theta'} \qquad (1)$$

and,

$$r'\ddot{\phi}' \sin\theta' + 2r'\dot{\theta}'\dot{\phi}' \cos\theta' = \left[a_{COR}\right]_{\phi'}. \qquad (2)$$

Here r' is assumed constant, and for the cases of interest either $r' = R$, where R is the radius of the rotating spherical planet, or $r' = R + h$, where h is the height of the satellite above the planet's surface. But when $r' = R + h$, the spherical surface in question is an imaginary, mathematical surface having this radius and rotating at the same rate ω as the planet. The notations $\left[a_{COR}\right]$ and $\left[a_{CENT}\right]$ denote the Coriolis and centrifugal accelerations, respectively. These are considered in more detail in the next section.

3. Summary of the Velocity-Change Calculation

The Coriolis acceleration is $\left[\vec{a}_{COR}\right] = -2\vec{\omega} \times \vec{\mathrm{v}}'$ and the centrifugal acceleration is $\left[\vec{a}_{CENT}\right] = -\vec{\omega} \times \left(\vec{\omega} \times \vec{r}'\right)$, with $\vec{\mathrm{v}}'$ denoting velocity of the object of interest in the primed frame. These are expressed in component forms in Eq. (3) through Eq. (5)

$$\left[a_{COR}\right]_{\theta'} = 2\omega\, r'\dot{\phi}'\sin\theta'\cos\theta', \quad \textbf{(3)}$$

$$\left[a_{COR}\right]_{\phi'} = -2\omega\, r'\dot{\theta}'\cos\theta', \quad \textbf{(4)}$$

and,

$$\left[a_{CENT}\right]_{\theta'} = r'\omega^2\sin\theta'\cos\theta'. \quad \textbf{(5)}$$

It should be noted that $\left[a_{CENT}\right]_{\phi'} = 0$, always.

The components of the kinematic acceleration, as introduced in Ref. 1, are

$$\left[a_{KIN}\right]_{\theta'} = r'\dot{\varphi}'^2\sin\theta'\cos\theta', \quad \textbf{(6)}$$

and

$$\left[a_{KIN}\right]_{\phi'} = -r'\dot{\theta}'\dot{\varphi}'\cos\theta'. \quad \textbf{(7)}$$

(The need to consider more than just the Coriolis and centrifugal accelerations was also discussed in Ref. 4.) In the primed frame, the tangential velocity $\left[\vec{\mathrm{v}}'\right]_{\theta',\phi'}$ is expressed as

$$\left[\vec{\mathrm{v}}'\right]_{\theta',\phi'} = r'\dot{\theta}'\hat{\theta}' + r'\dot{\phi}'\sin\theta'\hat{\phi}', \quad \textbf{(8)}$$

where $\hat{\theta}'$ and $\hat{\phi}'$ are the usual spherical unit vectors, and the components of the velocity changes over the finite time interval $(0,t)$ are expressed as

$$\Delta\left[\vec{\mathrm{v}}'\right]_{\theta'} = \Delta\left[r'\dot{\theta}'\right] = \left[r'\dot{\theta}'\right]_0^t, \quad \textbf{(9)}$$

and,

$$\Delta\left[\vec{\mathrm{v}}'\right]_{\phi'} = \Delta\left[r'\dot{\phi}'\sin\theta'\right] = \left[r'\dot{\phi}'\sin\theta'\right]_0^t. \quad \textbf{(10)}$$

Notice from Eq. (9) and Eq. (10) that the velocity changes of interest here are the velocity changes in a given compass direction. That is, it is the changes in the coefficients of the unit vectors of Eq. (8) that are of interest, not the changes in the unit vectors themselves. The symbol Δ has the usual meaning of "change."

As shown in Ref. 1, the components of the velocity change over a finite time interval $(0, t)$ in the primed frame are computed from the components of the acceleration as

$$\left[\Delta\vec{\mathrm{v}}'\right]_{\theta'} = \int_0^t \left[a_{COR}\right]_{\theta'} dt + \int_0^t \left[a_{CENT}\right]_{\theta'} dt + \int_0^t \left[a_{KIN}\right]_{\theta'} dt\,, \tag{11}$$

and,

$$\left[\Delta\vec{\mathrm{v}}'\right]_{\phi'} = \int_0^t \left[a_{COR}\right]_{\phi'} dt + \int_0^t \left[a_{KIN}\right]_{\phi'} dt\,. \tag{12}$$

There is no loss in generality by taking the time interval to start at $t = 0$.

4. Generalized Solution

The problem of an object initially located at the equator and given an initial tangential velocity was also considered in Ref. 1. The problem is depicted in Fig. 2.

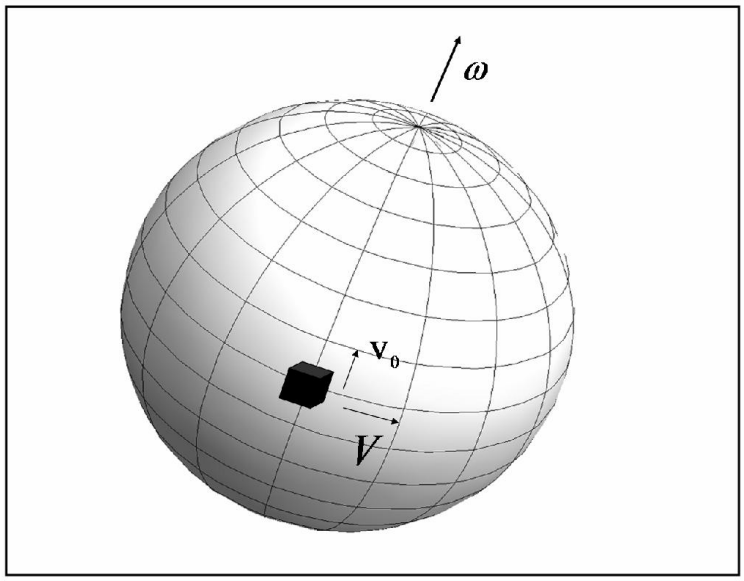


Fig 2 A sphere of radius *R* rotates uniformly at angular speed ω.

Here, V_0 is the initial tangential velocity in the northern direction and V is the initial tangential velocity in the eastern direction as specified in the unprimed, or inertial, frame.

The solutions of Eq. (1) and Eq. (2) for this problem are [1]

$$\theta'(t) = \arccos\left[\frac{\mathrm{v}_0}{\sqrt{\mathrm{v}_0^2 + V^2}}\sin\left(\frac{t\sqrt{\mathrm{v}_0^2 + V^2}}{r'}\right)\right], \quad \textbf{(13)}$$

and

$$\phi'(t) = \arctan\left[\frac{V}{\sqrt{\mathrm{v}_0^2 + V^2}}\tan\left(\frac{t\sqrt{\mathrm{v}_0^2 + V^2}}{r'}\right)\right] - \omega\ t. \quad \textbf{(14)}$$

(This problem has also been solved numerically[5].) In Ref. 1, the case of interest was the object having the same initial tangential speed in the eastern direction as the tangential speed of the sphere at the equator. However, Eq. (13) and Eq. (14) remain valid for arbitrary speeds v_0 and V. Hence the restriction to a specific value V can be removed, and these solutions are valid for this more general case. It should be understood that Eq. (13) and Eq. (14) are the solutions to Eq. (1) and Eq. (2) subject to the initial conditions $\left(\vartheta'(0), \phi'(0)\right) = \left(\pi/2, 0\right)$ and

$$\left(\dot{\vartheta}'(0), \dot{\phi}'(0)\right) = \left(V/r' - \omega, -\mathrm{v}_0/r'\right).$$

Here, v_0, V, and ω again have the same meaning as in Fig. 2, and $r' = R$ for an object on the surface of the rotating planet, and $r' = R + h$ for the orbiting satellite, with R being the sphere radius and h being the height of the orbiting satellite above the sphere's surface.

5. The Shadow Satellite

Although Eq. (13) and Eq. (14) are valid for an orbiting satellite, where $r' = R + h$, the equations assume the satellite would be viewed from an imaginary spherical surface located at the same height as the satellite, and rotating at the angular speed of the planet. Since such a surface is a purely mathematical construction, it is more helpful to transform the solution to the planet's surface.

The path directly under the satellite along the planet's surface is the satellite's ground track. To find the ground track, the idea of the "shadow" satellite is now introduced. The shadow satellite is taken to be an object located on the planet's surface directly underneath the orbiting satellite. The planet is assumed airless and the shadow satellite is assumed to be in frictionless contact with the surface. It is first assumed, and then proven, that it is possible to impart an initial velocity to the shadow satellite such that the shadow satellite remains directly underneath the orbiting satellite at all times.

We now consider the calculation of the initial velocity of the shadow satellite that in fact produces the behavior of always remaining directly underneath the orbiting satellite. For definiteness, we will also assume the planet of interest is the Earth, having the idealized physical properties previously mentioned, although the analysis applies to any spherical planet. To find the required initial velocity of the shadow satellite, we consider first a special case: The geosynchronous satellite. A satellite in geosynchronous orbit appears to an observer on the Earth's surface to remain stationary above a given point on the equator. Clearly, any object located on the equator directly underneath the geosynchronous satellite will serve as its shadow satellite, as defined here. For both the geosynchronous

satellite and for the shadow satellite, the tangential velocities obey the equation

$$\mathrm{v} = \mathrm{r}\,\omega \,. \qquad \textbf{(15)}$$

For the shadow satellite, $r = R$, the Earth's radius, and for the orbiting satellite $r = R + h$, where h is the height of the geosynchronous satellite above the surface. Writing out Eq. (15) for both the shadow and the orbiting satellite, and forming the ratio of these two equations, gives

$$\frac{\mathrm{v}_{SHADOW}}{\mathrm{v}_{ORBIT}} = \frac{R}{R+h}. \qquad \textbf{(16)}$$

Solving Eq.(16) for v_{SHADOW} gives

$$\mathrm{v}_{SHADOW} = \frac{R}{R+h}\ \mathrm{v}_{ORBIT}. \qquad \textbf{(17)}$$

Although there is no *a priori* reason to believe it will work, at this point we use Eq. (17) as a guide, and assume that each of the components of the tangential velocity of the shadow and orbiting satellite will obey an equation of the same form as Eq. (17), or

$$\left(\mathrm{v}_0\right)_{SHADOW} = \frac{R}{R+h}\ \left(\mathrm{v}_0\right)_{ORBIT}, \textbf{(18)}$$

and,

$$\left(\mathrm{V}\right)_{SHADOW} = \frac{R}{R+h}\ \left(\mathrm{V}\right)_{ORBIT}, \qquad \textbf{(19)}$$

where v_0 and V again are the northward and eastward tangential velocity components, respectively, as viewed in the inertial, or unprimed, frame. In this case, it is the shadow satellite that is depicted in Fig. 2, and the orbiting satellite, which is directly above it, is not shown.

In order to prove that Eq. (18) and Eq. (19) are valid for the general case, one starts by writing out Eq. (13) and Eq. (14) for both the shadow satellite and for the orbiting satellite. Naturally, the initial tangential velocities appearing in these equations should be replaced in each case with the appropriate subscript depending upon which of the two satellites the equations are being written for. Also, in the case of the shadow satellite, r' is replaced by R, and in the case of the orbiting satellite r' is replaced by $R + h$.

Next, the tangential velocities $\left(\mathrm{v}_0\right)_{SHADOW}$ and $\left(\mathrm{V}\right)_{SHADOW}$ are replaced in the two equations for the shadow satellite by the expressions for these quantities as given by Eq. (18) and Eq. (19), respectively. If the resulting equations are then simplified algebraically, the resulting pair of equations will be found to be identical to the two equations for the orbiting satellite.

This exercise proves that the solutions $\theta'(t)$ and $\phi'(t)$ are identical for both the shadow and the orbiting satellite. This therefore proves that if the shadow satellite is given the initial tangential velocity components as specified by Eq. (18) and Eq. (19), the shadow satellite will remain perpetually directly underneath the orbiting satellite. It thus also proves that the transformations given by these two equations are correct. And since the shadow satellite always remains directly underneath the orbiting satellite, it follows that the shadow satellite will trace out the ground track of the orbiting satellite, as desired.

6. Examples

Two numerical examples will now be considered. In the first example, the satellite is considered to have been launched from the equator and no attempt has been made to counter the effect of the Earth's rotation. We also assume the satellite orbits at a height of about 700 miles. These conditions approximate those of the POES program [2]. However, rather than considering the orbiting satellite in the example, we will consider instead the shadow satellite, in order to see the properties of the ground track.

The initial eastward and northward speeds of the shadow satellite are calculated by first computing these quantities for the orbiting satellite, and then transforming the results down to the shadow using Eq. (18) and Eq. (19). The initial eastward speed of the orbiting satellite is taken to be $V_{ORBIT} = (R+h)\,\omega$, where $R =$ 4000 miles and $h =$ 700 miles. The quantity ω is calculated assuming the Earth rotates in exactly 24 hours, giving $V_{ORBIT} \approx$ 1230.46 miles/hour. This result is then projected to the surface using Eq.(19). Carrying out these calculations gives $V_{SHADOW} \approx$ 1047.2 miles/hour. Not surprisingly, this is exactly the speed of the (idealized) Earth's rotation at the equator as seen from the unprimed, or inertial, frame.

The initial northward speed V_0 of the orbiting satellite is computed from the assumption that the orbital speed at the height of 700 miles is exactly 18000 miles/hour. Thus, $(\mathrm{V}_0)_{ORBIT} = \sqrt{18000^2 - V_{ORBIT}^2}$, or $(\mathrm{V}_0)_{ORBIT} \approx$ 17957.9 miles/hour. Projecting this result to the surface using Eq. (18) then gives $(\mathrm{V}_0)_{SHADOW} \approx$ 15283.3 miles/hour. Using these speeds together with $r' = R =$ 4000 miles in Eq. (13) and Eq. (14) produces the $(\vartheta'(t), \phi'(t))$ values of the ground track. Taking the complement of $\theta'(t)$ to be the latitude, and $\phi'(t)$ to be the effective longitude, the ground track produced by these calculations is as shown in Fig. 3. The result is clearly similar to the ground track of a POES-type satellite [2].

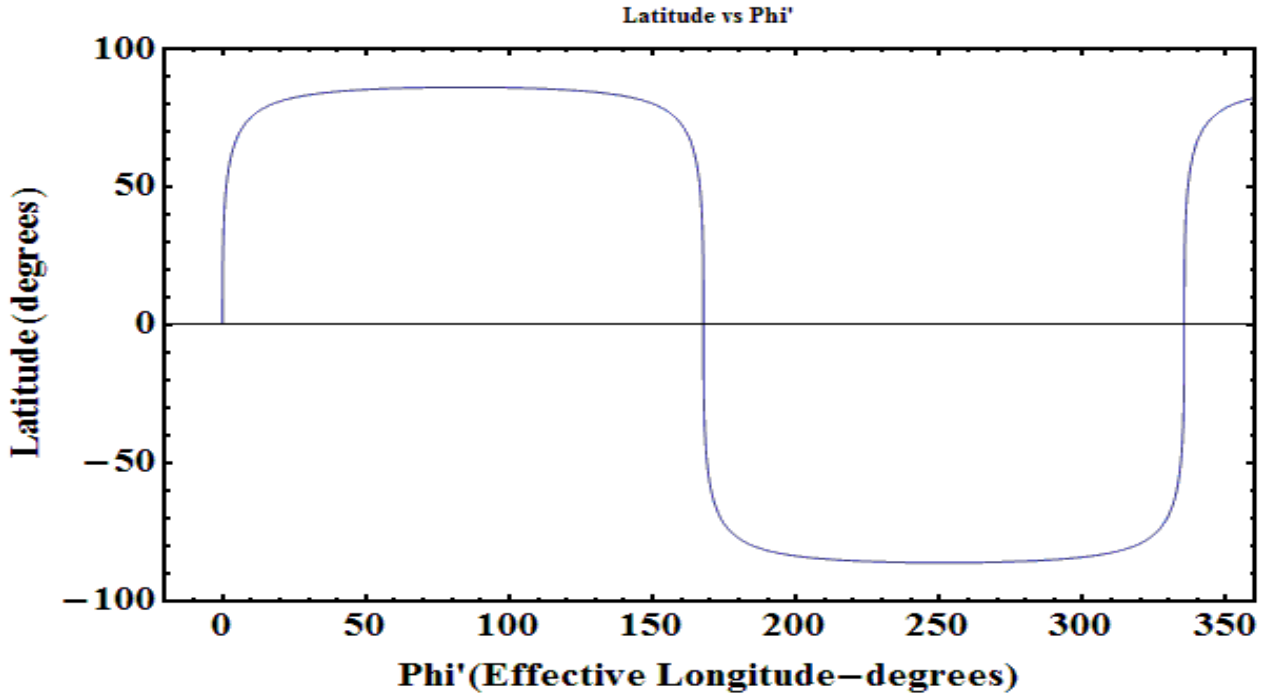


Fig. 3 - Ground track latitude vs. longitude for a POES-type satellite.

Examining Fig. 3, it is evident that that path is not a closed path. This is why a satellite in polar orbit can be used to observe the majority of the Earth's surface, since such a satellite passes over the equator at a different longitude on each orbit.

It is also interesting to consider specific velocity changes that occur for this satellite along the ground track, and to identify how much of the velocity change is attributable to each of the three accelerations involved in the problem.

Again examining Fig.3, it is clear that after starting at the equator $\left(\text{latitude} = 0 \text{ deg}\right)$, the shadow satellite closely approaches the northern pole $\left(\text{latitude} = 90 \text{ deg}\right)$, then rapidly reverses its northern motion and heads south. The distance of closest approach to the northern pole is about 273.6 miles. At the moment that the northern velocity reverses direction, it is clear that the initial northern velocity of 15283.3 miles/hour has been reduced to zero. The individual contributions to the change in northern velocity of the shadow satellite contributed by each of the three accelerations are:

$\left(\Delta \text{v}\right)_{COR} =$ - 410.647 miles/hour,

$\left(\Delta \text{v}\right)_{CENT} =$ - 36.273 miles/hour,

and $\left(\Delta \text{v}\right)_{KIN} =$ - 14836.4 miles/hour. It is evident that the overwhelming contributor to the northern velocity change is the kinematic acceleration. These results were computed by separately evaluating each of the corresponding integrals in Eq. (11).

It is also interesting to consider the velocity change in the eastern direction and the contributions of the individual accelerations to it. The shadow satellite initially has zero velocity in the eastern direction as seen in the rotating frame. At the moment the velocity in the northern direction reduces to zero, that is, at the moment of closest approach to the northern pole, the velocity in the eastern direction reaches its maximum value. (This was proven in Ref. 1.) At that moment, the eastern velocity of the shadow is about 15247.56 miles/hour. The contributions of the individual accelerations to producing this velocity are:

$\left(\Delta \text{v}\right)_{COR} =$ 1951.22miles/hour,

$\left(\Delta \text{v}\right)_{CENT} =$ 0, and $\left(\Delta \text{v}\right)_{KIN} =$ 13296.3 miles/hour.

These results were computed by evaluating the corresponding integrals in Eq. (12). Again we note that the null centrifugal contribution results from the fact that the centrifugal acceleration in the ϕ' direction is always identically zero. Again, it is apparent that the kinematic contribution strongly dominates the easterly velocity change, although not as significantly as in the northern direction.

We next consider a second example, but in this case it is assumed that the satellite has been launched in a way that almost cancels the velocity component due to the Earth's rotation at the equator as seen in the inertial frame. For the case of interest, it is assumed that the satellite that is in orbit at a height of 700 miles above the surface has been launched so that the initial velocity component that is directed toward the east has been reduced to just 10 miles/hour as seen from the unprimed, or inertial, frame.

Again, we are interested in the shadow satellite for the current example. Applying the same procedures as were described for the first example gives for the initial velocity components of the shadow satellite as seen in the unprimed, or inertial, frame $V_{SHADOW} \approx$ 8.51 miles/hour and $\left(\text{v}_0\right)_{SHADOW} \approx$ 15319.15 miles/hour. The initial eastward velocity seen in primed, or rotating, frame is -1038.69 miles/hour. The minus sign signifies that the shadow satellite is actually moving westward as seen in the rotating frame. The initial northward velocity in the rotating frame is the same as that seen in the inertial frame.

Again using the initial shadow speeds for this case together with $r' = R =$ 4000 miles in Eq. (13) and Eq. (14) produces the $\left(\vartheta'(t), \phi'(t)\right)$ values of the ground track. Plotting these in the same way

as was done to produce Fig. 3 results in the ground track as shown in Fig. 4. The fact that the ground satellite has an initially westward velocity can be seen by noting the negative slope of the initial ground-path curve as seen in Fig. 4. Again it is clear that when the shadow satellite reaches the northern pole it reverses its direction of travel and then heads southward. Hence the satellite again reaches a point where its northward velocity becomes zero as seen in the rotating frame. This happens at the point of closest approach to the northern pole, which happens when the shadow satellite is just 2.22 miles distant from the pole.

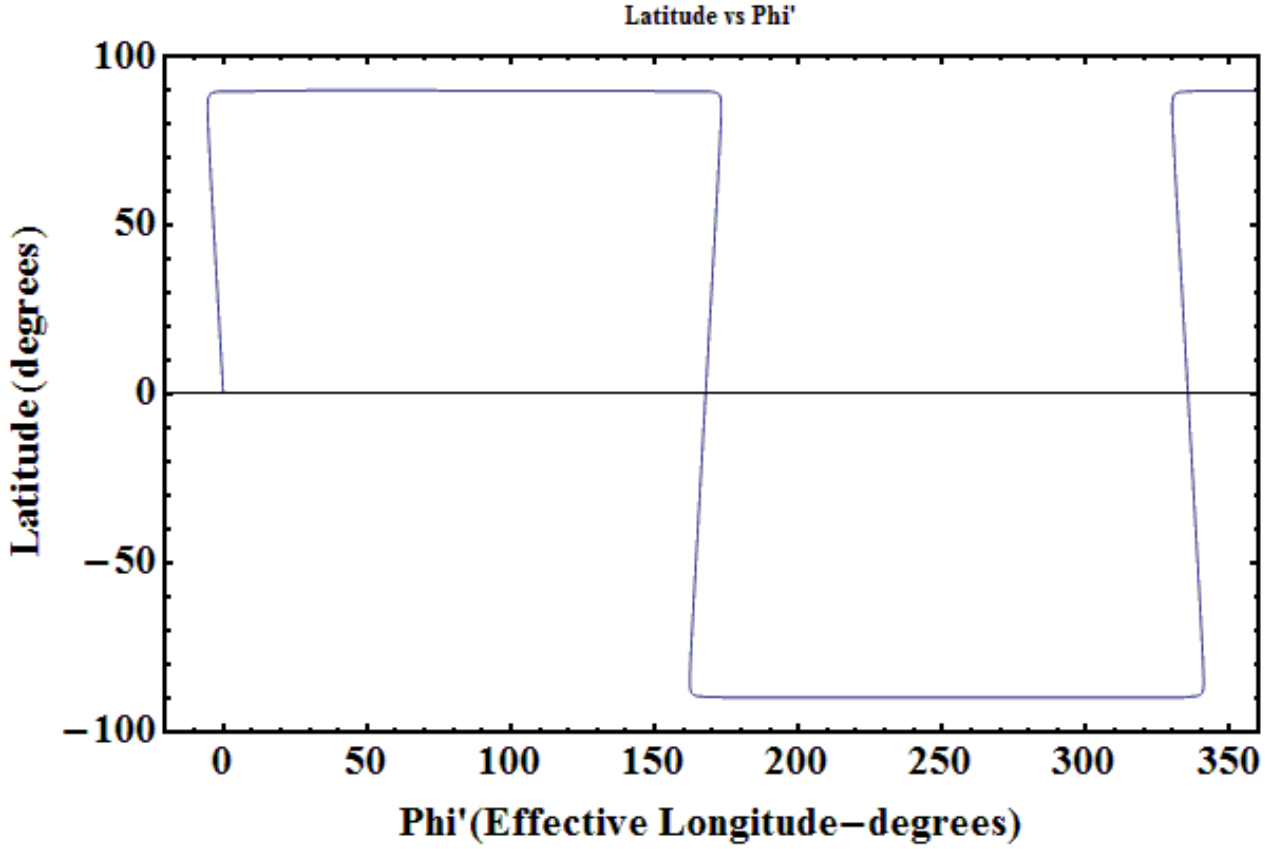


Fig. 4 - Ground track latitude vs. longitude for a satellite launched to almost cancel the Earth's rotational velocity.

As can be seen in Fig 4, as the shadow gets close to the northern pole (latitude 90 degrees), it undergoes an extremely rapid change in longitude. This is a consequence of the very close approach of the shadow satellite to the northern pole, where there is a discontinuity in the value of ϕ' . That is, if the satellite were approaching the northern pole along the zero-degree longitude line, there would be a discontinuous change in longitude as it passed the pole, because that line abruptly changes in longitude from 0 degrees to 180 degrees at the location of the pole. The longitude values will vary in a similar way for the shadow satellite in this example, owing to its very close approach to the northern pole.

Again it is of interest how much each of the acceleration terms contributes to reducing the shadow's initial northward velocity from $\left(\mathrm{V}_0\right)_{SHADOW} \approx 15319.15$ miles/hour to zero. The results are:

$(\Delta v)_{COR}$ = +62.06 miles/hour, $(\Delta v)_{CENT}$ = - 35.79 miles/hour, and $(\Delta v)_{KIN}$ = - 15345.40 miles/hour.

It is again evident that the overwhelming contributor to the northern velocity change is the kinematic acceleration. It is also interesting to notice that the Coriolis contribution is positive, which occurs due to the westward-directed initial velocity component, and hence the Coriolis force actually acts to increase the northward velocity during the trip from the equator to the northern pole. As before, these results were computed by separately evaluating each of the corresponding integrals in Eq. (11).

Again considering the motion in the easterly direction, it is evident from Fig. 4 that the shadow suffers a huge apparent eastern acceleration as it approaches the northern pole. It then just as rapidly decreases its eastward velocity, and shortly past the northern pole the satellite again has a westward-directed velocity component, as can be seen from the fact that the longitude values begin to decrease as the shadow begins to move southward. The maximum eastward velocity of the shadow in the rotating frame, which again occurs at the point of closest approach to the northern pole, is 15318.57 miles/hour. Keeping in mind that the "eastern" velocity was initially -1038.69 miles/hour in the rotating frame (actually westward), the overall *change* in eastern velocity is 16357.3 miles/hour ($\approx$ 15318.57 miles/hour+1038.69 miles/hour, to within rounding error). Breaking down the overall eastward velocity change by acceleration term gives:

$(\Delta v)_{COR}$ = 2093.23 miles/hour,
$(\Delta v)_{CENT}$ = 0,
and $(\Delta v)_{KIN}$ = 14264.00 miles/hour.

These results were again computed by evaluating the corresponding integrals in Eq. (12), and once again the ϕ' component of the centrifugal acceleration is identically zero.

It may be of concern as to why the shadow acquires such a high easterly directed velocity at the moment of closest approach to the northern pole. This again happens due to the fact that the shadow approaches so closely to the pole. If one considers the latitude circle at the point of closest approach to the pole, that circle has a radius of just over 2.2 miles. As with all latitude circles, motion along their length is either purely eastward or purely westward. At the point of closest approach to the northern pole, the shadow is moving tangent to this latitude circle, and thus is moving purely eastward as seen from the surface. Since the northward movement is null at this point, it is clear that the entire surface velocity of the satellite must appear in the eastward direction. To do this, the eastward velocity must suffer a sharp increase in the vicinity of this latitude circle. This very high eastward velocity, however, is only present while the satellite is in close proximity to this latitude circle, an event which is of extremely short duration, and this occurs while the shadow moves over a very small area of the surface.

7. Summary and Conclusion

A satellite in a circular orbit obeys all the requirements for the methods of Ref. 1 to be applicable. Thus velocity changes for such satellites over a finite time interval cannot be computed correctly by integrating only the sum of the Coriolis and centrifugal accelerations over that time interval. It is also necessary to include the contributions from the kinematic acceleration. Interestingly, in the case of a satellite in circular orbit, the kinematic contribution is dramatically dominant over the Coriolis and centrifugal accelerations, despite the fact that those are the only two accelerations that appear in the equations of motion in the rotating frame.

The notion of a shadow satellite was introduced to allow detailed study of the ground track of the orbiting satellite. With suitable transformations applied, all the results that apply to the orbiting satellite also apply to the shadow satellite.

The results and examples presented here should be useful to those who teach graduate- or undergraduate-level classical mechanics courses. These space-age applications of classical mechanics are potentially of high interest to students.

ACKNOWLEDGMENT

The drawing of the sphere shown in Fig. 2 was produced using *Mathematica*, from Wolfram Research.

References:

[1] J.C.Piquette, "Velocity Change Calculation for an Object Moving on a Rotating Spherical Surface," *Phys. Educ.* 31(1) (2015), art. num. 3, pp. 1-12.

[2]http://ww2010.atmos.uiuc.edu/(Gh)/guides/rs/sat/poes/home.rxml .

[3] See, for example, K. R. Symon, *Mechanics*, Addison-Wesley, Reading, Massachusetts, 1964, p. 94, Eq. (3-102) for the components of acceleration in spherical coordinates, and p. 277. Eq. (7-37) for expressions for the Coriolis and centrifugal accelerations. The sum of these two accelerations is set equal to the acceleration expressed in spherical coordinates; the constraint of no radial motion is imposed; and the angular components identified to derive Eq. (1) and Eq. (2).

[4] D. H. McIntyre, "Using great circles to understand motion on a rotating sphere," Am. J. Phys. **68** 1097-1105 (2000).

[5] A. Amengual, "Noninertial trajectories on a fast rotating planet," Am. J. Phys. **68** 1106-1108 (2000).

A Simple Estimation of Young's Modulus and Tensile Strength of Carbon Nanotubes

Debnarayan Jana[1]

[1]Department of Physics
University of Calcutta, 92 A P C Road
Kolkata 700 009, India.
E-mail: djphy@caluniv.ac.in

(*Submitted* 14-11-2014)

Abstract

Within the harmonic approximation, we estimate the Young's modulus and tensile strength of typical carbon nanotubes at zero temperature following four important potentials used in solid state physics. The theoretical results are compared with experimental observations. The results thus obtained also find an important link with the applicability of carbon nanotubes as a space elevator.

1. Introduction

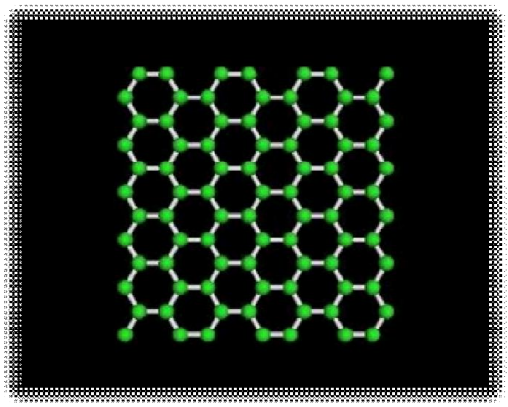
Scientific discovery along with technological revolutions strongly depends on materials. To achieve smaller, faster and reliable smart devices for the continuing need of present age, it is highly essential to understand the properties of the materials in depth. Mechanical, optical properties of nano-structures [1,2] and nano particles coupled with electronic study are indeed one of the exciting fields of research from basic science. This study also forms the basis for future smart devices.

After Iijima's pioneering and illustrious work on carbon nanotubes [3] and single-walled carbon nanotubes (SWCNTs) [4–8] in the early 1990s sparked a general growing interest in fundamental condensed matter physics as well as nano science and nanotechnology. To describe the structures of SWCNT, one needs to know graphene, the basic building block of carbon allotropes.

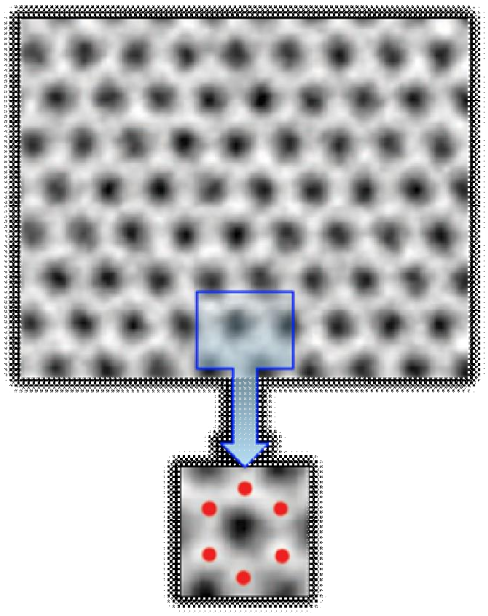
Graphene is a one-atom-thick planar sheet of carbon atoms that is densely packed in a honeycomb crystal lattice. We show schematically in Fig. 1 the computer generated graphene and the direct image obtained by Meyer et al. [9] side by side. It is worthy to mention at this point that the graphene [6,10-12], by itself, can be characterized as either a zero-gap semiconductor or a metal (since the density of states (DOS) is zero at the Fermi energy (E_F)) and naturally, graphene imparts these properties to a nanotube.

Carbon nanotubes (CNTs) are the allotropes of carbon with a cylindrical nanostructure. Conceptually, CNTs are cylindrically shells made by rolling graphene sheets into a seamless cylinder. These sheets are rolled at specific and discrete angles. The typical length of a SWCNT can be between 1–100 μm and diameter around 1-10 nm. The perfect CNTs have crystalline structures formed by the hexagonal rings of benzene molecule with double and single C–C bonding. In Fig. 2, we schematically show the formation of SWCNT from graphene and fullerene. Their unique one dimensional structure with curvature in the sidewall is one of the

paradigms in the low dimensional systems of inter-disciplinary research.



(a)



(b)

Fig. 1. (a) Computer generated graphene visual showing the honeycomb lattice structure. (b) Direct image of a single-layer graphene membrane (Red dots denote carbon atoms)

The typical molecular structure of SWCNTs can be characterized by a chiral circumferential vector $A\vec{B} = n\vec{a} + m\vec{b}$, a linear combination of two unit lattice vectors a and b with m and n being integers. The pair of indices (n,m) for any given nanotube structure determines its diameter, chirality, and the basic electronic character.

For example, if $n = m$, the nanotube is designated as armchair and is metallic in nature (with a zero bandgap, strictly speaking). While for $n \neq m$ and neither n nor m are zero, the CNT exhibits chirality, having important implications in optical properties.

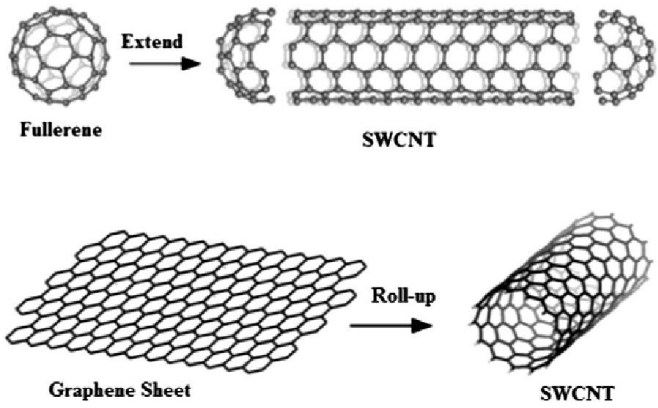


Fig. 2. Schematic illustration of formation of SWCNT from Fullerene and graphene sheet

For $n = 0$ or $m = 0$, the CNT is termed zigzag. If $n - m \neq 3p$, where p is a non-zero integer, the CNT is semimetallic/ quasi-metallic with a band gap of the order of meV. For $n - m = 3p$, where p is a non-zero integer, the CNT is semi-conducting having a band gap of the order of 1 eV. In fig. 3, the various flavors of CNTs are illustrated through the chiral circumferential vector. From the circumferential vector AB, one can easily obtain the diameter and chiral angle, the characteristic features of a typical CNT.

The diameter of a given nanotube can be expressed in terms of (n,m) and the carbon–carbon bond length 0.142 nm. The typical geometry of simple hexagonal unit used in generating the structures of CNT or graphene is shown in Fig. 4.

The distance (a) between two carbon atoms as shown in Fig. 4 is 0.142 nm while the half distance between any two parallel bonds ($b/2 = (\sqrt{3} \times 0.142)/2$ nm=0.123 nm). The diameter of a given nanotube can be expressed [1,4] in terms of (n,m) as $d_{nm} = \frac{\sqrt{3} \times 0.142}{\pi} \sqrt{n^2 + nm + m^2}$.

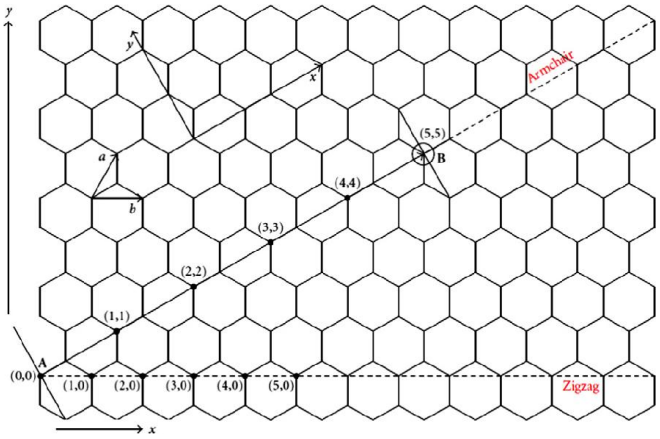


Fig. 3. Schematic illustration of various carbon nanotubes

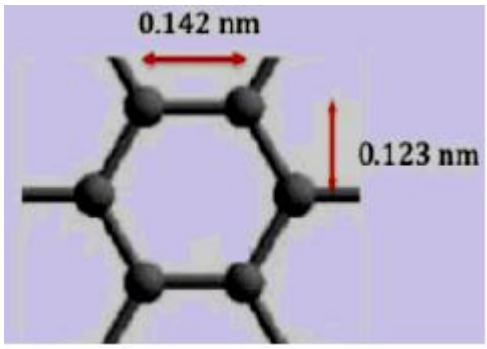


Fig. 4 A Typical hexagonal unit for formation of graphene

C-G bonds are one of the strongest bonds in nature. It has been illustrated in Fig 3 and 4 that the carbon nanotube is composed of perfect arrangement of these bonds. Because of these bonds, carbon nanotubes are the strongest material [1,4] known ever having Young's modulus (Y) 1250 GPa and tensile strength (σ) 11-63 GPa quite comparable to diamond. In comparison, the values of Y and σ in steel are 200 GPa and 2 GPa respectively. In this paper, we would like to present a simple model calculation to estimate the high values of Y and σ. In table 1 and 2, we show the comparison of various related physical parameters with two different forms carbon allotropes, diamond and graphite.

Name of Elements	Li	Be	B	C(Dia)	C(Graph
Atomic Number	3	4	5	6	6
Y(GPa)	11.5	289	440	1140	8.3
Melting Point (^{0}C)	181	1277	2030	3550	3550
Density (10^3 Kg/m^3)	0.531	1.85	2.34	2.25	2.25

Table 1: Comparison of Various physical parameters with diamond and graphite

Intuitively speaking, the materials with strong covalent bonds have a deep potential energy with a sharp curvature. Therefore, strong bonding naturally results in large values for Young's modulus. Similarly, the shallow potential well of the weakly bonded materials is responsible for small values of Y as evident above from table 1. Elements beyond carbon do not form solids with a three dimensional network of covalent bonds. For example, graphite having two dimensional sheet held together by van der Waals bonds possess a very small value of Y (about 8 GPa). Again, the interatomic distance varies with crystal direction in a solid along with corresponding variation of bond strength. This results an elastic anisotropy as a function of crystal direction. More importantly, this effect is prominent for those materials having two types of bonds. As an illustration, Y for

graphite in a direction parallel to the sheets is 950 GPa which is sensibly larger than that averaged over all direction (about 8 GPa). This picture is valid at absolute zero temperature.

Name of Elements	Na	Mg	Al	Si
Atomic Number	11	12	13	14
Y(GPa)	8.9	44	71	103
Melting Point (^{0}C)	98	650	660	1410
Density (10^3 Kg/m^3)	0.97	1.74	2.70	2.33

Table 2: Comparison of Various physical parameters of Alkali metals and Si

2. Computation of Young's Modulus

We begin the section with the definition of Young's modulus. It is defined as the ratio of longitudinal stress to longitudinal strain within elastic limit. Within elastic limit, the elongation (x) is proportional to the applied force (F) according to Hooke's law. This gives us the force constant (k) as the ratio of F to x. We would like to compute the Young's modulus of the material within this elastic limit of the chemical bond so called as harmonic approximation. If l_0 is the stress free length, A is the cross sectional area, we can write the Young's modulus as

$$Y = \frac{kl_0}{A} \qquad (1)$$

It is to be noted that the force constant k is related to the force between pair of atoms as $k = -\left[\frac{d^2U}{dr^2}\right]_{r=R_0}$ with $F = -\frac{dU}{dr}$. Now, in the macroscopic specimen of length l_0 and cross-sectional area A, the specimen has roughly l_0/r_0 number of bonds where R_0 is the typical equilibrium distance of the interaction potential U(r). If one stretches this specimen by an infinitesimal distance dx, then the typical length of these bonds will increase by $\frac{R_0 dx}{l_0}$. Naturally, the tension in the each chain of atoms will be $\frac{kR_0 dx}{l_0}$. But the specimen contains $\frac{A}{R_0^2}$ of these chain of atoms, hence, the total force needed to produce an extension dx will be $\frac{kAdx}{l_0 R_0}$. Thus, the stress developed in the specimen is simply $\frac{kdx}{l_0 R_0}$. Therefore, the Young's modulus can be written in terms of force constant k as

$$Y = \frac{1}{R_0}\left[\frac{d^2U}{dr^2}\right]_{r=r_0} \qquad (2)$$

We model the potential of the chemical bonds by four different functions often taken in solid state physics. The first one in this category is the famous (6-12) Lennard-Jones potential given by

$$U_1(r) = U_0\left[\left(\frac{R_0}{r}\right)^{12} - 2\left(\frac{R_0}{r}\right)^{6}\right] \qquad (3)$$

This potential is also used in statistical mechanics and liquid state theory. The power 6 is due to the fluctuating dipolar interaction energy [13] while the power 12 is not fixed in the sense that it could be any power greater than 6 for the stability reason. In the fluctuating dipolar theory [13], it is assumed that the two atoms are separated by a distance r. If the instantaneous dipole moment of

the first atom is $\vec{p}_1$, then within the dipole approximation, the electric field at a distance r will scale as $\frac{\vec{p}_1}{r^3}$. As a result this field will eventually induce a dipole moment $\vec{p}_2$ in the second atom of the order of $\frac{\vec{p}_1}{r^3}$. Therefore, the typical fluctuating dipolar energy will be of the order $\frac{\vec{p}_1 \cdot \vec{p}_2}{r^3} \cong \frac{\vec{p}_1^2}{r^6}$. Interestingly, although the time average value of $\vec{p}_1$ and $\vec{p}_2$ is zero but the time average of interaction energy is non-zero because of the quadratic dependence on $\vec{p}_1$. It is interesting to note that the potential goes to zero in the asymptotic limit of the distance. Moreover, the force generated from the above interaction is restoring one because above the equilibrium distance R_0, the force is attractive and below R_0, repulsive in nature. We expand the force around the equilibrium distance $(r = R_0 + x_0)$ and restricting to linear elastic limit, we find

$$F = -\frac{72U_0}{R_0^2} x_0 = -kx_0 \qquad (4)$$

This helps to identify the relevant force constant (*k*). Hence, the Young's modulus can be written as

$$Y = \frac{288U_0}{\pi a d^2} \qquad (5)$$

Here, we have used $R_0 = a$ and the cross-sectional area $A = \frac{\pi d^2}{4}$ with d being the relevant diameter. One can also give a simple justification to the above formula from simple dimensional analysis [14]. Given the parameters used in the potential i.e. U_0 and R_0, it is easy to note that $Y \propto \frac{U_0}{R_0^3}$.

The expression obtained in equation (4) can now be used to estimate the typical magnitude of Young's modulus of CNT. Assuming $U_0 = 4.93\,\text{eV}$ as noted in case of Morse potential [15], we find the typical magnitude of force constant as 2790 N/m. For (9,0) SWCNT (the circumference turns out to be 2.2 nm), the equation (4) should be multiplied by 9 for the computation of Y. The estimated value thus turns out as 9363 GPa. This value is around 7.5 times larger than the experimentally observed value. The large value of U_0 taken in the above calculation is the reason for this discrepancy. However, 0.5 eV value of U_0 gives us a reasonable value of 948 GPa. If we take typical value $(1.68 \times 10^{-21}\,\text{J})$ used for Ar atoms [15], we get back a very low value of Y (20 GPa) for (9,0) SWCNT.

As a generalization of the above potential, we consider the following one:

$$U(r) = U_0 \left(\frac{mn}{n-m}\right)\left[\frac{1}{m}\left(\frac{R_0}{r}\right)^m - \frac{1}{n}\left(\frac{R_0}{r}\right)^n\right] \qquad (6)$$

with $m. > n$. The equation (3) follows from the above equation in the limit $m = 12$ and $n = 6$. With this generalized potential, the expression for Y turns out as

$$Y = \frac{4mnU_0}{\pi a d^2} \qquad (7)$$

Another potential energy often used in molecular physics [16] is defined as

$$U_2(r) = B\left[\frac{R_0^7}{8r^8} - \frac{1}{r}\right] \qquad (8)$$

with $B = 2.31 \times 10^{-28}\,\text{Jm}$. The second term in the above equation is the Coulomb term. In this case, the value of Y in terms of B becomes

$$Y = \frac{28B}{\pi a^2 d^2} \qquad (9)$$

When three positive charge particles are maintained in a straight line with the end particles being identical and are held fixed at a distance R_0, a new type potential [16] emerges

$$U_3(r) = B\left[\frac{1}{r} - \frac{1}{r-2R_0}\right] \qquad (10)$$

The above elastic limit calculation yields the Young's modulus as

$$Y = \frac{16B}{\pi a^2 d^2} \qquad (11)$$

Another variant of anharmonic potential

$$U_4(r) = \left[-\frac{A_0}{r} + \frac{B_0}{r^9}\right] \qquad (12)$$

in material science [17] can also be used for rough estimation of Y. With $A_0 = 7.68\times10^{-29}$ Jm , it is easy to notice that the expression of $\frac{d^2U_4}{dr^2}$ at the equilibrium distance (r_0) becomes

$$\frac{d^2U_4(r_0)}{dr^2} = -\frac{2A_0}{r_0^3} + \frac{90B_0}{r_0^{11}} \qquad (13)$$

with $B_0 = \frac{A_0 r_0^8}{9}$. Therefore, the numerical value of Y for (9,0) becomes 825 GPa comparable to experimental observation.

The above picture is valid for absolute zero temperature. A suitable generalization to finite temperature can be done in the following way. At finite higher temperature, with the help of thermal energy, the atoms vibrate about their mean equilibrium positions. As a result, the amplitude of the vibration increases with increase of temperature. With further increase of temperature (still quite far away from their respective melting points), the bonds between the atoms loosened up. As a consequence, there is a decrease of Young's modulus with temperature. With the help of linear expansion coefficient α , the expression (6) can be written at finite non-zero temperature as

$$Y(T) = \frac{4mnU_0}{\pi a_0(1+\alpha T)d_0^2(1+\alpha T)^2} \approx \frac{4mnU_0}{\pi a_0 d_0^2(1+3\alpha T)} \qquad (14)$$

In other words, the zero temperature and finite temperature Young's modulus are related roughly by

$$Y(T) \approx \frac{Y(0)}{1+3\alpha T} \qquad (15)$$

2. Computation of Tensile Strength

It is clear that for the computation of tensile strength, one has to take the single C-C bond in the hexagonal network. However, in all the above potentials, there is no information about the rupturing of the bond itself. However, we use the harmonic approximation to estimate the magnitude of the ultimate tensile strength in the following way. In Fig. 5, we show the variation of the scaled van der Waal interaction with scaled distance and the harmonic approximation adopted in it. The maximal elongation (x_{max}) is computed by equating the maximum potential energy at this distance to the bond energy. This gives us

$$x_{max} = \sqrt{\frac{2U_0}{k}} \qquad (16)$$

for the first potential in equation (3).

Therefore, the final maximum tensile strength in terms of Y can be written as

$$\sigma_{max} = \frac{Yx_{max}}{a} = \frac{Y}{6} \qquad (17)$$

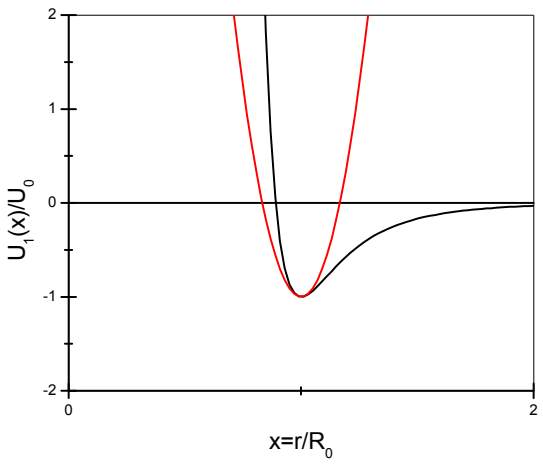


Fig. 5. Sketch of the scaled typical van der Waal interaction (black) and its harmonic approximation (red) with dimensionless distance

For (9,0) SWCNT, this value is simply 1560.5 GPa quite large compared to experimental observed value. Similar analysis for generalized Lennard –Jones potential in equation (6) yields the maximal value of tensile strength as

$$\sigma_{max} = Y\sqrt{\frac{2}{mn}} \qquad (18)$$

The equation (14) can also be used to estimate Y if σ_{max} of the material composing the system is known.

Fig. 6 represents the schematic variation of the interaction used in equation (8) with scaled distance. Please note the long tail of the interaction due to Coulomb interaction in contrast to Fig. 5. For the potential in equation (8), the ultimate tensile strength reduces to

$$\sigma_{max} = \frac{Y}{2} \qquad (19)$$

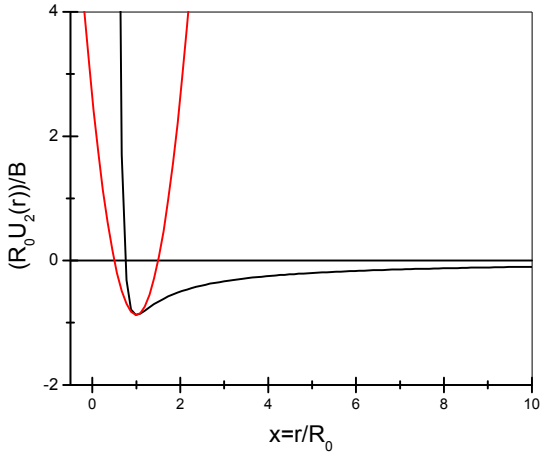


Fig. 6. Sketch of the scaled second interaction potential (black) used in equation (8) and its harmonic approximation (red) with dimensionless distance

Similarly, the expression for the potential given in equation (10) is recast as

$$\sigma_{max} = \frac{Y}{2} \qquad (20)$$

It should be remembered that the discrepancy arises in the large estimated values lies with the validity of harmonic approximation in such a situation. In fact, the harmonic approximation breaks down for rupturing the bonds in SWCNT.

In Table 3, we compare the different values of Y and σ_{max} obtained from all the above interaction potential with other observed values. The large difference between the theoretical predicted values and real experimental values may be due to structural defects inherent in SWCNT [15].

Table 3: Comparison of values of Young's modulus and tensile strength of (9,0) SWCNT

	Expt	Morse	U_1	U_2	U_3	U_4
(9,0) SWCNT Y (GPa)	270-950	1000	9533 948(*)	1872	1071	825
(9,0) SWCNT σ_{max} (GPa)	20-130	500	1560.5 158(*)	936	1071	389

Before we conclude we would like to point the possibility or prospect of carbon nanotubes as a space elevator [15, 18]. Till date CNT has not been produced at the macroscopic scale to produce a very long cable/rope connecting earth and space station/satellite. There might be occurrence of damage from atmosphere by storms and lightning. One has also to be careful to take appropriate precautions from collisions rendered by heavenly bodies moving around earth. Therefore, we still believe that its status as a space elevator remains in the arena of science fiction.

3. Conclusions

Within the limit of harmonic approximation, we have made a rough estimation of the Young's modulus and tensile strength of typical carbon nanotubes at zero temperature following four important potentials used in solid state physics. The theoretical results are compared with experimental observations. We have also generalized the zero temperature result to finite non-zero temperature with the help of linear expansion parameter. The results thus obtained also find an important connection with the applicability of carbon nanotubes as a space elevator.

Acknowledgements

The author gratefully acknowledges all those postgraduate students who showed keen interest in learning these low dimensional systems. Special thanks to Miss Namrata Dhar for careful reading of the manuscript.

References :

[1] C. Binns, *Introduction to Nanoscience and Nanotechnology* (Wiley, NJ, 2010).
[2] D. Das, K. Nandi and D. Jana, Physics Education, **19**, No. 3, 155 (2002).
[3] S. Iijima, *Nature*, **354**, 56 (1991).
[4] S.Reich, C. Thomsen, J. Maulzsch, *Carbon nanotubes: basic concepts and physical properties* (Weinheim: Wiley-VCH; 2004).
[5] J. M. Schnorr, T. M. Swager, Chem Mater, **23**, 646 (2011).
[6] M. I. Katsnelson, Mater Today, **10**, 20 (2007).
[7] D. Jana, C.-L. Sun, L.-C. Chen, K.H. Chen, *Prog. Mat .Science*, **58**, 565 (2013).
[8] D. Jana, L.-C. Chen, C.W. Chen, S. Chattopadhyay, K.H. Chen, *Carbon*, **45**, 1482 (2007).
[9] J. C. Meyer et al., *Nano Lett*., **8**, 3582 (2008).
[10] A. K. Geim and K. S. Novoselov, *Nature Materials*, **6**, 183 (2007).
[11] K. S. Novoselov et al., *Nature*, **438**, 201 (2005).
[12] K. S. Novoselov et al., *Science*, **306**, 666 (2004
[13] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Thomson, 2007).
[14] D. Jana**,** *Dimensional Analysis: Modern Perspectives* (Published by Lambert Academic Publishing, Germany, 2011).
[15] Z. Bochnicek, *Physics Teacher,* **51**, 462 (2013).

[16] H. D. Young and R. A. Fredman, *University Physics* (Pearson Education, 2004, India).

[17] M. S. Rogalski and S. B. Palmer,*Solid State Physics* (Gordon and Berach Science Publications, 2001).

[18] P. K. Aravind, *Am. J. Phys*., **75**, 125 (2007).

A Low-cost Refractometer in an Undergraduate laboratory

Nutan Chandra*, Puja K Singh**,Pallavi Bera**, Shilpi Adhikari**

*Associate Professor,at present in P.G. Dept. Kolhan Univ. Chaibasa, ** students of Department of Physics,
Jamshedpur Women's College,
Jamshedpur 831001, India.
nutan.chand@gmail.com

(*Submitted* 11-08-2015)

Abstract

The Undergraduate syllabus of Physics Honours course in Jamshedpur Women's College, Jamshedpur expects the students to carry out a project (30 marks) in the Fifth semester. A laser refractometer built out of a spectrometer (which has seen better days), a hollow glass prism(fabricated) and some other apparatus mandatorily available in the Undergraduate laboratory unfolds some interesting ideas to carry out projects on a shoestring budget.

1. Introduction

One of the most important optical properties of a medium is its refractive index. As it is a fundamental property of the substance it is often used to identify a particular substance, confirm its purity or measure its concentration.

The absolute refractive index of a medium is the ratio of the speed of electromagnetic radiation in free space to the speed of the radiation in the medium. The relative refractive index is the ratio of the speed of light in one medium to that in the adjacent medium. Refraction occurs with all types of electromagnetic waves but the refractive index of a medium differs with the frequency of the wave. For a given frequency the refractive index of a medium depends on the density of the medium, which is again a function of temperature.

Beside other methods, the refractive index of a substance can be measured using a refractometer. i.e. how much a light beam bends on passing through a medium.

Petrol is at times suspected to be adulterated with cheaper fuels like kerosene oil .An effort was made to determine the "purity" of petrol by calculating the refractive indices of petrol and that of a mixture of petrol and kerosene oil (in various proportions) and the values plotted to give a calibration curve to determine the degree of adulteration in any given sample. The usual method of determining the refractive index of a liquid with a hollow prism, a sodium lamp source and a spectrometer was found unsuitable as the image obtained was blurred making it difficult to take measurements.[1].

A make-shift refractometer was, therefore, made from a discarded spectrometer and a laser source replaced the conventional sodium lamp source.[2]

2.Fabrication

(a)Construction of a hollow glass prism

A transparent glass sheet 3mm thick was used from which a square 5cm x 5cm and three rectangular strips 2cm x 4 cm were cut out. Using a water-proof adhesive, the pieces were glued together and left to dry for 24 hours.(see Fig. 1)

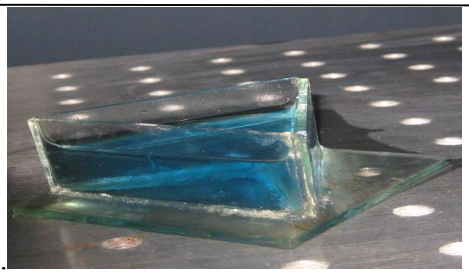


Fig.1 Hollow glass prism filled with kerosene

(b) Setting up a make-shift Refractometer.

A spectrometer was used to make a refactometer. The telescope tube was made open ended on both ends by removing the eye-piece and the lens and simply used as case for placing the pencil laser source.

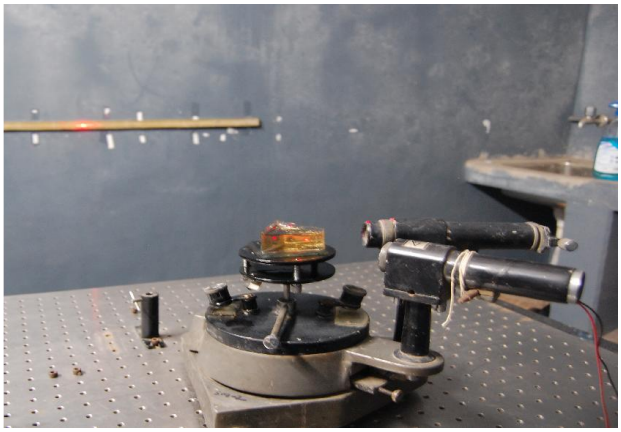


Fig.2

The prism table was used to keep the hollow prism filled with the desired liquid and also used to rotate it about a vertical axis passing through it. The collimator had no role to play and hence was rotated out of the way of the laser beam. A metre-rule was fixed on the wall to receive the incident laser beam after refraction through the liquid-filled prism as shown in Fig.2.

3.Theory and Method

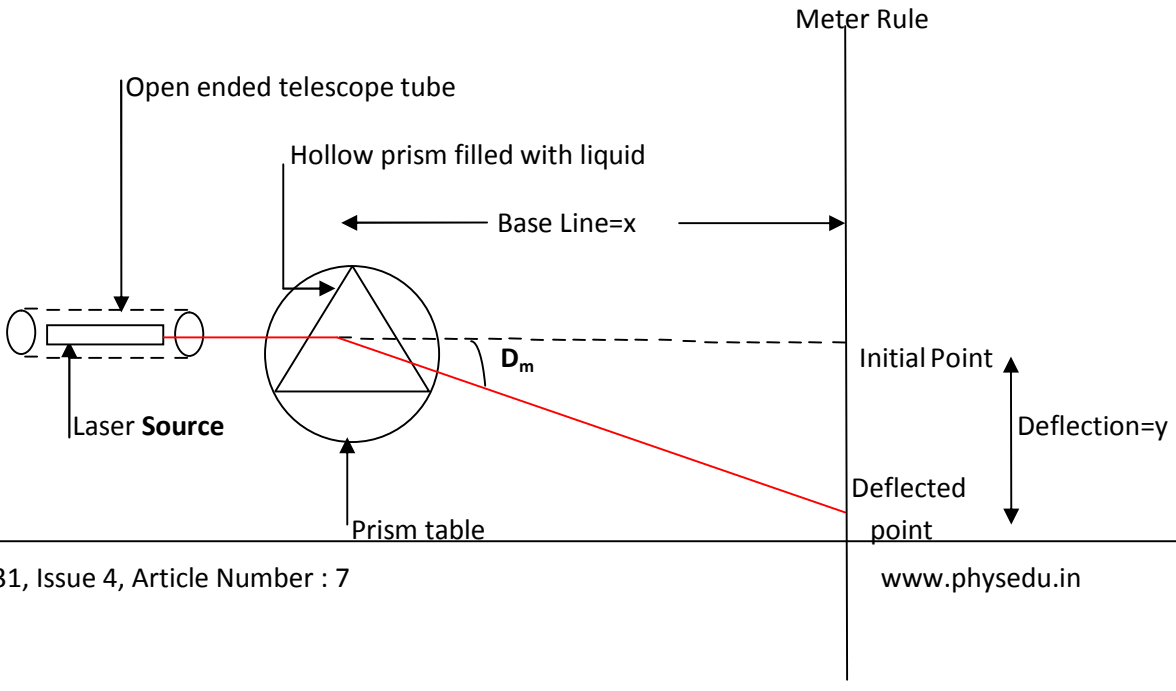


Fig.3

The above figure 3 shows the layout of the make-shift refractometer and the path of a laser beam(red) undergoing minimum deviation. D_m refers to the angle of minimum deviation. The point on the ruler where the laser light falls in the absence of the prism is called the "Initial point" and the distance between the centre of the prism and the Initial point is called the base line..

The refracting angle A and the angle of minimum deviation D_m give the refractive index μ of the medium by the following well known formula,

$$\mu = \sin\{ (A + D_m)/2\}/\sin(A/2) \text{ ----------(1)}$$

From fig.3 we have

$$D_m = \tan^{-1}(y/x)=\tan^{-1}(\text{deflection/baseline})\text{----(2)}$$

Using equations (1) and (2) we get

$$\mu \sin(A/2) =\sin\{(A + \tan^{-1} (y/x))/2\}\text{-------(3)}$$

Which finally gives an expression for the angle of the prism as

$$A= [2 \tan^{-1} \{\sin (D_m /2)\}] / \{ \mu - \cos(D_m /2)\} \text{ ------(4)}$$

Calibration of prism In theory, each angle of an equilateral prism is 60˚ but for a self-fabricated prism, this may not be true. Hence the prism needs to be be calibrated which in effect means that the angle between a pair of refracting sides of the prism must be determined.

To calibrate, first a pencil laser source was placed in the hollow open-ended telescope tube so that the laser beam travelled to the mounted ruler. The prism should not be on the prism table at this point. The reading on the ruler where the laser spot fell denoted the "Initial point". Next a hollow prism, filled with water, was placed on the prism table. The laser spot on the ruler shifted. The prism table was rotated about the vertical axis passing through it so that the shift or deflection was a minimum. The reading on the meter-rule now represents the "deflected point".

The actual deflection is given by

Deflection = (Deflected point) – (Initial point)

= y--------- (5)

The distance between the "initial point" and the centre of the prism was measured and denoted as baseline (= x). Assuming μ= 1.33 (water) and using Equations (2) and (4), the angle of the prism "A" was calculated. This value was written on a sticker and stuck on the side of the prism opposite to the angle measured .This would keep the value of "A" handy and also ensure that in the rest of the experiment too, the laser beam passed through the calibrated sides.

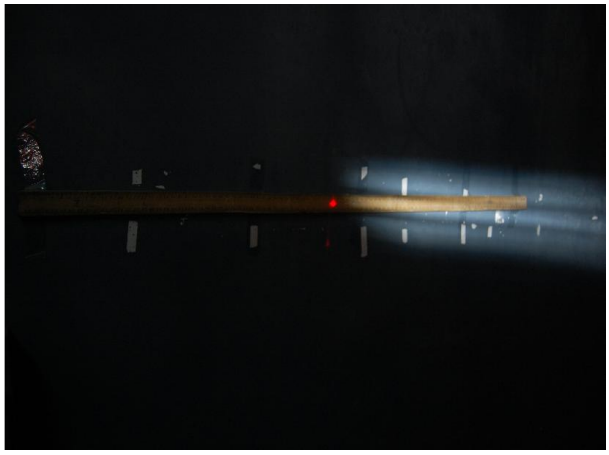


Fig.4 Laser spot on mounted ruler

Measurement of refractive index of the mixture

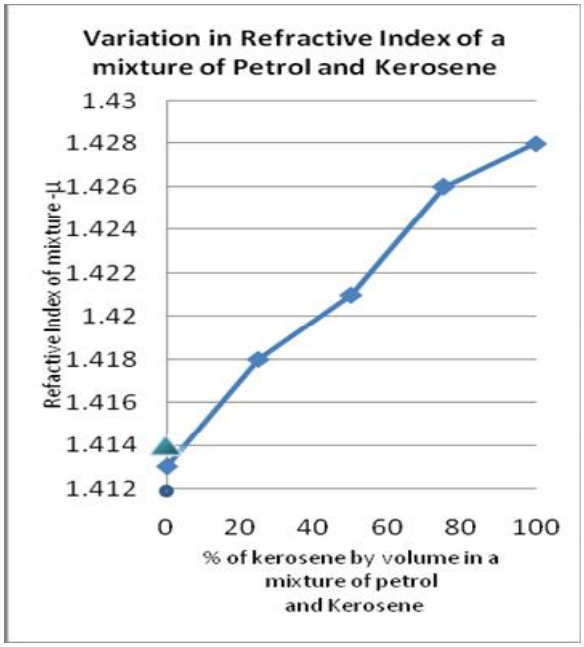
Without disturbing the laser source, the hollow prism now filled with petrol was placed on the prism table. Care was taken to align the prism so that the laser beam passed through the calibrated sides. As before, the prism was rotated about the vertical axis to obtain the minimum deflection. The deflected point was noted and the deflection and the angle of minimum deflection calculated using (5) and (2). Equation (1) was now used to calculate the refractive index of the petrol in the hollow prism. The value for "A" obtained in the previous section was used in Equation (1).

Next Kerosene was mixed with petrol such that the ratio by volume was 25: 75 and the above process was repeated. Several readings with increasing proportions of kerosene in the mixture were taken ending with 100% kerosene (no petrol). The refractive index of the mixture was calculated each time.

It was ascertained that the laser source did not undergo any motion (either rotational or translational) throughout the experiment. Its alignment with respect to the centre of the prism was kept constant and the prism was only rotated about the vertical axis through its centre. All these conditions were automatically satisfied in the above method by virtue of the construction of the spectrometer.

4.Results and Discussion

A graph was plotted between the refractive index of the mixture and the percentage of kerosene in it. It was observed that with an increase in the percentage of kerosene, the refractive index of the mixture increased. (Fig. 5).



The above triangle ▲ and ● dot show the refractive indices of petrol from two different petrol pump stations.

Fig.5

As petrol is made up of multi-components which can be broadly divided into alcohols, alkanes and aromatic compounds, the

refractive index of petrol/gasoline varies from country to country and also from company to company. Moreover it is also dependent on the method of extraction, area of extraction and even the time of the year. As such comparison of our results with any standard result was not possible. The studies done in Sweden on gasoline gives its refractive index as 1.429 at 27˚C using laser light (blue-green $\Delta\lambda$ = 488-514 nm) [3]. Our study was done with a source of He-Ne laser (λ= 668 nm) and for a particular brand of petrol, the value was 1.413. As it was diluted with kerosene (μ = 1.44)[4] its refractive index increased depending on the dilution.

5.Limitations of the method

Our result for refractive index of kerosene turned out to be 1.429, a little less than the standard value. Our set up could not be expected to yield very accurate results as a self-fabricated hollow glass prism was used. Moreover, in our deductions we have treated the thickness of glass as negligible which is not really true.

6. Learning

However, on performing the project, the student understands that refractive index can well be an index of adulteration in many liquids. The student also learns how to use an old instrument in a new and innovative way besides using the geometry of the set up and the concepts of Physics to deduce an expression for the refractive index. In a nutshell. the student is introduced to the flavor of research, something which she might savour and gravitate towards, later in her career.

7. Way forward

- A thin-walled and commercially manufactured hollow glass prism may give more accurate results.
- A calibration graph of y(= deflection) against % of dilution can be used as a quick way of indication of adulteration in petrol.
- The above set-up may be used to measure the refractive indicies of many transparent liquids and also to study the variation of the refractive index with change in temperature and also change in concentration in case of solutions.
- A fibre optic sensor for determining the refractive index of liquds is likely to give better results as was done by Roy in 1999 [5]

8. Expenses incurred on

- Waterproof adhesive
- Petrol and Kerosene
- Labour charge for cutting glass for making the prism.

.

5. References

[1] Subedi D.P.,Adhikari .R.,Joshi U.M.,Poudel H.N,Niraula B. Study of temperature and concentration dependence of refractive index of liquids using a novel technique. Kathmandu University Journal of Science, Engineering and Technology. Vol.II,February 2006.

[2]www.sciencebuddies.org/science-fair-projects/project..../Phys_p028.shtml

[3] Analytical Methods for Determining Automative Fuel Composition Jonas Gruber,Renata Lippi,Rosamaria W.C.Li andAdriano R.V. Benvenho.

[4]www.industrialmindworks.com/inwit/writings/refractiveindicies.html

[5] Sukhdev Roy Fibre optic sensor for determining adulteration of petrol and diesel by kerosene. Sensors and Actuators B Chemical

Visualizing Electromagnetic Fields Using Gnuplot

Somnath Datta
Professor of Physics (Retired), National Council of Educational Research and Training, New Delhi
Res: 656, "Snehalata", 13th Main, 4th Stage, T K Layout, Mysore 570009, India
datta.som@gmail.com; http://sites.google.com/site/physicsforpleasure

(Submitted 30 - 08 -2015)

Abstract

Gnuplot provides an excellent tool for plotting electromagnetic fields from various sources of electric charges and currents, stationary as well as time-varying. Its extraordinary power can be particularly exploited in plotting the propagating electromagnetic field from a localized source, e.g., an *oscillating electric or magnetic dipole*. In this article we have written the explicit commands in Gnuplot which will draw the **E** field lines of the electromagnetic field due to a harmonically oscillating electric dipole aligned with the Z axis. We have plotted the **E** field lines on the XZ plane over a limited region of radius 2.2 wavelengths around the source. We have shown, side by side, the **B** field lines from the same source. However, the **B** lines are coaxial circles around the Z-axis, and hence do not warrant plotting. The plotting of the **E** field involves plotting a 3D "relief map" with *contours* embedded on it, for a certain function $\psi(x, y, t)$, with t held constant. The contour levels are selected by applying certain criteria. The plotted contours are converted to *directed* contours, by drawing arrowheads on them, indicating the direction of the **E** field, so that they qualify as field lines. We have demonstrated two alternative methods of adding this qualification, namely (a) *planting* the **E** field vectors at selected points on the XZ plane, and extrapolating them to the contour lines; (b) plotting the E_θ component along the X axis, its positive value implying **E** pointing in the negative Z direction and vice versa, and then following this direction around the entire contour. We have worked out method (a) only for $t = 0$; and method (b) for one full cycle of oscillation corresponding to eight values of t spaced at equal intervals. Looking at these eight plots sequentially one sees how the electromagnetic field is propagating across space. We have plotted E_θ along the X axis for the same eight values of t, all of them on the same graph, to get a clear view of how the field is oscillating and moving, and its amplitude is falling as the inverse of the distance. At the end we have used Gnuplot to plot the **E** and **B** fields for a linearly polarized plane electromagnetic wave.

1 Introduction

Distributions of stationary charges and steady currents create *time-independent* fields $\mathbf{E}$ (electric field) and $\mathbf{B}$ (magnetic field), respectively. On the other hand, moving charges, or localized distributions of *time-varying* electric charges and electric currents create both $\mathbf{E}$ and $\mathbf{B}$. We shall represent these fields as $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$, where t stands for time, and $\mathbf{r}$ for (x, y, z) in the Cartesian coordinate system, (r, θ, ϕ) in the spherical coordinate system, and represents the coordinates of an arbitrary *field point* (rather, the radius vector to the field point.) We usually call these combined fields $(\mathbf{E}, \mathbf{B})$ the *electromagnetic field*.

There are standard formulas for the electromagnetic field for simple cases: (i) A *plane wave* $(\mathbf{E}, \mathbf{B})$ propagating through a limited region of space located far away from the source, and (ii) the Electromagnetic field $(\mathbf{E}, \mathbf{B})$ originating from a *localized source*, namely, a *harmonically oscillating electric dipole, or magnetic dipole*, the mathematical expressions for $(\mathbf{E}, \mathbf{B})$ valid for all space, with the origin of the coordinate system located at the source point (i.e., where the point-like dipole is located.) In this article our interest is limited to these two examples.

The mathematical formulas for these two fields are simple, in fact deceptively simple for the case (ii). We have seen them umpteen number of times in text books, but failed to visualize how the field *looks* like, how exactly it is spread over space, how its angular distribution changes as we go from the near region to the far region. One has to see plots of these fields, to comprehend them, and get a picture. There are articles, and some modern books where one can find plots in which the field lines have been displayed beautifully. Even then, the examples seen by us are deficient in one respect, absence of 'arrowheads' indicating whether the field is pointing 'up or down' along the field lines.

Partly to remove this deficiency, partly to equip the reader with his own tool for drawing the fields, and partly to continue our efforts to clarify concepts in Classical Electrodynamics with illustrated examples, graphics and plots[1, 2, 3], that we are writing this article.

All our plotting work is based on freely available software: (i) Gnuplot (Version 4.6), and (ii) Xfig (version 3.2.5c). One can get them and their Manuals[4] from the internet for licence free use. The operating system is Linux. We have used Debian distribution (version 5.0 Lenny) on our desktop computer, and Mint distribution on our laptop computer.

The author had used Gnuplot extensively in his latest book[5] *Mechanics*, in which he had also written an *Appendix* to introduce the reader to Gnuplot and its applications, namely drawing orbits of planets and space vehicles, trajectories of various objects, plotting coordinate vs time of particles in motion. The reader can benefit from these examples, can use them as a starting point for learning and practising Gnuplot. However, these experiences proved to be inadequate for drawing the field lines of the $(\mathbf{E}, \mathbf{B})$ fields discussed in this article. It required two months

of self training, working through a large variety of exercises using two resource books, namely the Gnuplot Manual[4] and Gnuplot Cookbook[6], before the author gained confidence in writing the commands to create the field lines and 'plant' the **E** vectors along the field lines.

The objective of this article is twofold: (1) To present a graphical illustration of how electromagnetic field propagates in space (i) from a central localized source, viz., an oscillating dipole, and (ii) as a linearly polarized plane wave; (2) How to use Gnuplot for replicating the same graphs on the reader's own computer.

We have placed greater emphasis on objective (2). For this purpose we have copied the actual commands from the Console, presented them in 14 *Exercises*, starting each one with a heading preceded by a serial number in bold italics, e.g., ***Ex.1, Ex.2***, etc. These exercises can be used to replicate *all* the graphs shown in this article. We have provided explanation of some of the commands, up to Ex.8, in two ways: (a) its meaning at the end of the command line, separated by the symbol #, (b) general explanations/instructions at the end of the Exercise, emphasizing them with the "bullet" symbol "**•** ". For understanding the unexplained commands, the reader should look them up in the references just cited, in particular the Gnuplot Manual.

It is hoped that experience gained by doing the exercises shown in this article will be found useful by students and teachers for application in a variety of other problems and assignments in physics.

We shall begin with the Electromagnetic field $(\mathbf{E}, \mathbf{B})$ originating from an oscillating Electric dipole, which is more interesting. This work will cover most of this article, and is spread over Sections 2-10 (Pages 4 - 33).

2 The $(\mathbf{E}, \mathbf{B})$ Field of an Oscillating Electric Dipole

Fig.1(a) gives a schematic picture of an oscillating electric dipole. This figure also explains the spherical coordinates used, and the unit vectors associated with them. The observation point P (we shall call it *field point*) is located at the radius vector $\mathbf{r}$, has polar coordinates (r, θ, ϕ). It should be remembered that, unlike Cartesian unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$, the spherical unit vectors $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi$ should be drawn at P (its components in the directions of the XYZ axes are functions of the angular coordinates θ, ϕ of P.) They are pointing in the directions in which the respective coordinates are increasing. In particular the unit vector $\mathbf{e}_\phi$ lies in a plane parallel to the XY plane and is tangent to a circle with centre on the Z axis. We have brought it down from P to its projection N on the XY plane, and shown it separately in Fig (c), in a reduced scale.

The dipole consists of two metallic domes A and B, each spherical in shape, mounted on a metallic pole of length ℓ. The system is neutral as a whole, but has equal and opposite charges on the opposite domes. If at some instant of time t the sphere A has charge $q_a(t)$, then at the same instant of time the other

sphere has charge $q_b(t) = -q_a(t)$. A current $I(t)$ flowing through the pole, driven by an oscillating voltage source (Fig.(b)), will make the charge move back and forth between the domes. It is assumed that this oscillation is taking place harmonically at the angular frequency ω. As a consequence the system will develop a harmonically *oscillating electric dipole moment* $\mathbf{p}(t) = p(t)\mathbf{e}_z = q_a(t)\ell\,\mathbf{e}_z$. Here we have used the symbol $\mathbf{e}_z$ to mean a unit vector in the Z-direction. The *wave vector* $\mathbf{k}$ is defined as

$$\mathbf{k} = k\mathbf{e}_r; \quad \text{where} \quad k = \frac{\omega}{c} = \frac{2\pi}{\lambda} \qquad (1)$$

to be called the *wave number*, and λ is the wavelength of the ensuing radiation.

We shall write the "source quantities" in a proper form. We shall assume that the current through the metallic pole is uniform[1] (but time varying). Let q_0 be the maximum charge collecting on each dome. Then

$$\begin{aligned} &q_a(t) = q_0\,\cos\omega t; \\ &q_b(t) = -q_a(t) = -q_0\,\cos\omega t; \\ &\mathbf{p}(t) = q_0\ell\,\cos\omega t\,\mathbf{e}_z = p_o\,\cos\omega t\,\mathbf{e}_z; \\ &\text{where } p_o = q_0\,\ell \end{aligned} \qquad (2)$$

is the scalar amplitude of this electric dipole moment. The $(\mathbf{E}, \mathbf{B})$ field from this oscillating dipole, at some point P far away from the dipole, and located at spherical coordinates (r, θ, ϕ) is given by the following formulas [7, 8, 9], assuming that $r \gg \ell$.

$$\begin{aligned} \mathbf{E}(\mathbf{r},t) &= \frac{p_o}{4\pi\epsilon_0 r^3}\left[\{\cos(kr-\omega t) + kr\sin(kr-\omega t)\}\,2\cos\theta\,\mathbf{e}_r\right. \\ &\quad \left. + \left\{(1-k^2r^2)\cos(kr-\omega t) + kr\sin(kr-\omega t)\right\}\sin\theta\,\mathbf{e}_\theta\right]. \quad (a) \\ c\mathbf{B}(\mathbf{r},t) &= \frac{p_o}{4\pi\epsilon_0 r^3}\left[\left\{kr\sin(kr-\omega t) - k^2r^2\cos(kr-\omega t)\right\}\sin\theta\,\mathbf{e}_\phi\right]. \quad (b) \end{aligned} \qquad (3)$$

Here $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$ are unit vectors associated with the polar coordinates (r, θ, ϕ), i.e., pointing in the directions of the increments of the respective coordinates.

Note that we have multiplied the magnetic field $\mathbf{B}$ with c, the speed of light in vacuum, to get the modified field $c\mathbf{B}$, which has the same unit as $\mathbf{E}$ in the SI units (volt/m) (and which becomes necessary to express the EM field in a relativistically covariant manner.) Moreover $E = cB$ in the radiation zone, i.e., far away from the source, i.e., regions for which

$$kr \gg 1, \quad \text{or,}\ r \gg \lambda. \qquad (4)$$

In this radiation zone approximation, or *far*

[1] This assumption will be valid for the special case of our investigation in which the wavelength λ of the resulting em wave is much much larger than the length of the rod, i.e., $\lambda \gg \ell$.

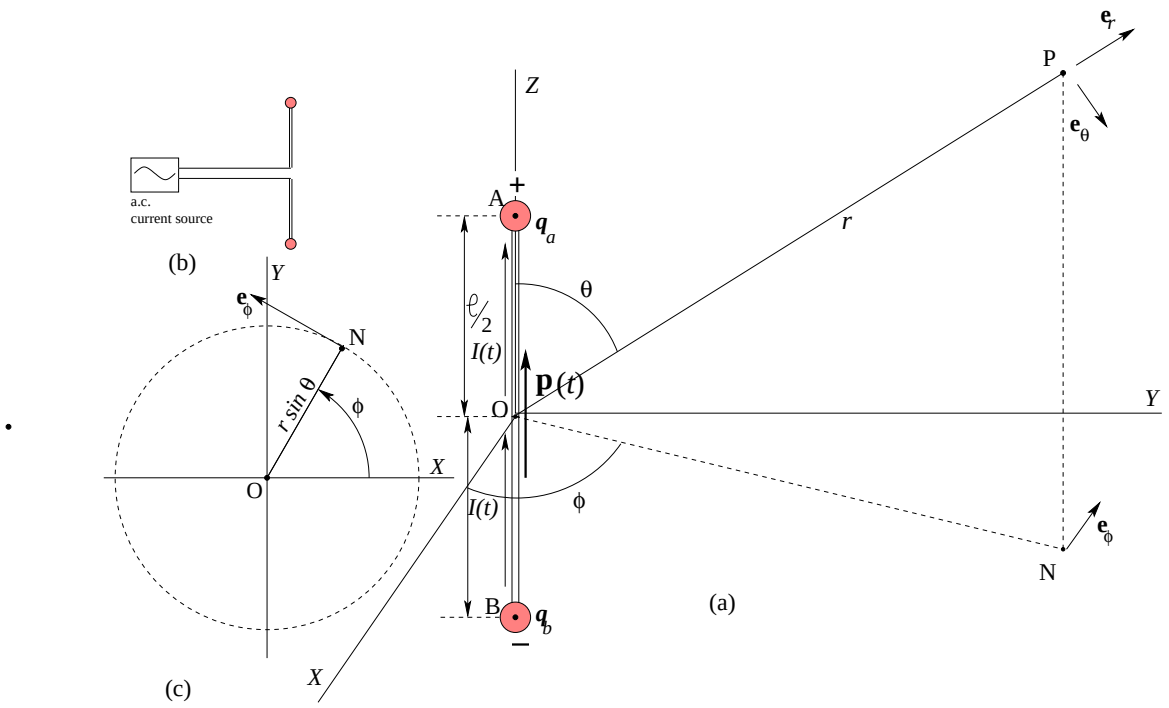


Figure 1: *(a) Oscillating Electric Dipole at the origin, and the field point P at spherical coordinates* (r, θ, ϕ); (b) The oscillator driving the dipole; (c) Explaining the unit vector $\mathbf{e}_\phi$, by projecting it on the XY plane.

zone approximation (4), the EM field takes the simple form[**?**]

$$\mathbf{E}(\mathbf{r},t) = -\frac{k^2 p_o}{4\pi\epsilon_0}\frac{\cos(kr-\omega t)}{r}\sin\theta\,\mathbf{e}_\theta.$$
$$c\mathbf{B}(\mathbf{r},t) = -\frac{k^2 p_o}{4\pi\epsilon_0}\frac{\cos(kr-\omega t)}{r}\sin\theta\,\mathbf{e}_\phi. \tag{5}$$

This is the radiation field repesenting an electromagnetic wave emanating from a point like dipole source located at the origin and propagating along the direction of the radius vector $\mathbf{r}$ drawn from the origin. We prefer to use the symbol $\mathbf{n}$ to indicate the direction of propagation. Which means that $\mathbf{n} = \frac{\mathbf{k}}{r} = \mathbf{e}_r$. The radiation field is marked by three important characteristics:

$$\begin{array}{lrcl} \text{(I)} & c\mathbf{B} & = & \mathbf{n}\times\mathbf{E}; \\ \text{(II)} & cB & = & E; \\ \text{(III)} & E, cB & \propto & \frac{1}{r}. \end{array} \tag{6}$$

The opposite of the approximation (4) is the *near zone* approximation:

$$kr \ll 1; \quad \text{or, } r \ll \lambda. \tag{7}$$

In this approximation we consider only the zeroth power of kr. Ignoring the first and

second power of kr in Eqs.(3) we get

$$\mathbf{E}\left(\mathbf{r},t\right) \approx \left(\frac{p_o}{4\pi\epsilon_0}\frac{2\cos\theta\,\mathbf{e}_r + \sin\theta\,\mathbf{e}_\theta}{r^3}\right)\cos\omega t.$$
$$c\mathbf{B}\left(\mathbf{r},t\right) \approx \mathbf{0}. \tag{8}$$

3 Angular Distribution of the Power Radiated by the Oscillating Dipole

The Poynting's vector, defined as

$$\mathbf{S} = \epsilon_0 c\left[\mathbf{E}(\mathbf{r},t)\times c\mathbf{B}(\mathbf{r},t)\right] \tag{9}$$

gives the flux density of the radiated electromagnetic energy. Consider a point P located far away from the origin, at the spherical coordinates (r,θ,ϕ). The EM field at this point is given by Eq. (5). Therefore, the instantaneous electromagnetic energy flux density at this point is given as

$$\mathbf{S}(\mathbf{r},t) = \epsilon_0 c\left\{\frac{k^2 p_o}{4\pi\epsilon_0}\frac{\cos(kr-\omega t)}{r}\sin\theta\right\}^2(\mathbf{e}_\theta\times\mathbf{e}_\phi)$$
$$= \frac{ck^4p_o^2}{16\pi^2\epsilon_0}\frac{\cos^2(kr-\omega t)}{r^2}\sin^2\theta\,\mathbf{e}_r. \tag{10}$$

The above equation gives the instantaneous value of $\mathbf{S}$. Since the dipole will be oscillating very fast, at frequencies of the order of kHz, what is more relevant is the *time-averaged value* of $\mathbf{S}$, to be written as $<\mathbf{S}>$. This is easily obtained by noting that $<\cos^2(kr-\omega t)>=\frac{1}{2}$. Then,

$$<\mathbf{S}(r,\theta,\phi)>=\frac{ck^4p_o^2}{32\pi^2\epsilon_0}\left(\frac{\sin^2\theta}{r^2}\right)\mathbf{e}_r. \tag{11}$$

The power radiated per unit solid angle is

$$\frac{dP}{d\Omega} = r^2 <\mathbf{S}(r,\theta,\phi)>=\frac{ck^4p_o^2}{32\pi^2\epsilon_0}\sin^2\theta. \tag{12}$$

Plotting the $\sin^2\theta$-angular distribution of the radiated power, as given in Fig. 2, is a trivial application of Gnuplot. The dipole is oriented along the Z axis and is labelled $\tilde{\mathbf{p}}$ (to indicate that it is alternating harmonically). The polar angle θ is measured from the positive Z axis.

We have indicated the strength of the Poynting's vector $\mathbf{S}$ over a sphere of radius r, by the length of the arrow representing this vector. The shaded double-lobe about the Z axis is the XZ plane cross section of an axially symmetrical doughnut like 3-dimensional plot of the radiated power, varying as $\sin^2\theta$, which is characteristic not only of dipole radiations (from both electric and magnetic dipoles), but also of radiation from an accelerating charge (in non-relativistic motion.) In the latter case, the angle θ is measured from the direction of the instantaneous acceleration vector.

It is especially notable that there is *no radiation along the axis of the dipole*, and *maximum radiation along the plane perpendicular to it.*

4 Plotting with Gnuplot

The commands of Gnuplot are to be written on Console. We shall show a few examples of how to use Gnuplot. Let us observe a few conventions. Every command in Gnuplot is preceded by the following prompt

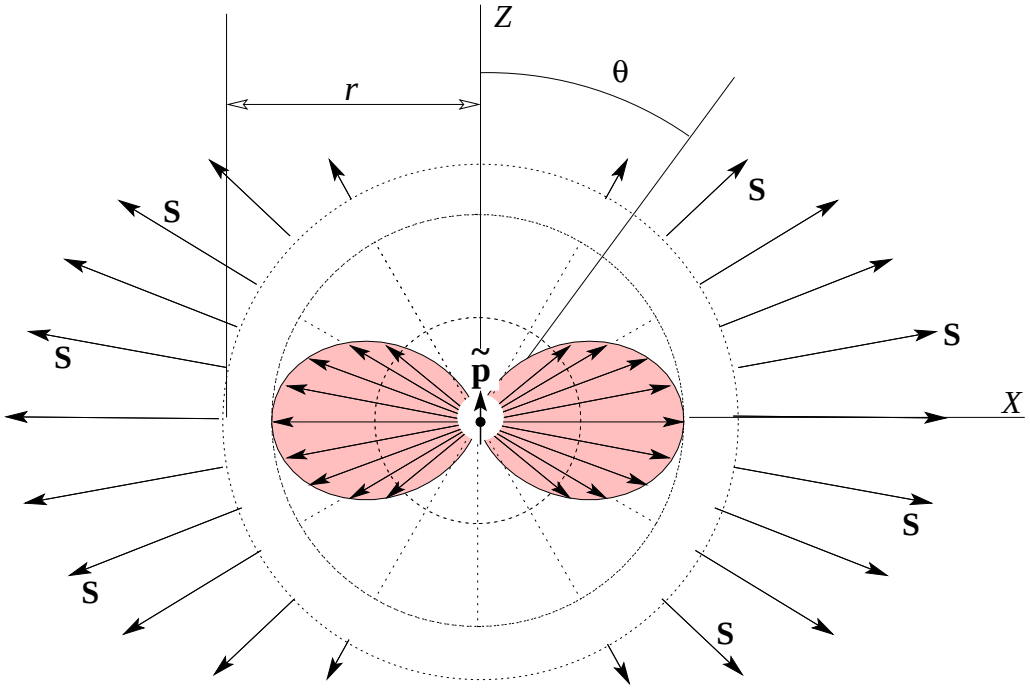


Figure 2: *Energy flow from an oscillating Electric Dipole*

gnuplot>
by which Gnuplot asks us to write a command line. On the other hand when Gnuplot executes a command and gives its answer or response, there is no such prompt. For economy of space we shall write
>
instead of the full prompt "gnuplot>". We shall also use the "comment" symbol '#' for explaining a particular command to the reader. This symbol '#' is analogous to the "comment" symbol '%' used in LaTeX. Gnoplot ignores anything written after # on the same line.

To illustrate these points we shall ask Gnuplot to give the value of a certain function $f(x)$ which is either already there in its own "library", or which we have just defined, for a few specific values of the argument.

We have copied below the "script" from the Console in "footnote size typewriter font", to mark them out from the main text, and to adjust them within the limited space of a column.

Ex.0

```
UrComputer:~ UrDir$ gnuplot # First line
G N U P L O T
    Version 4.6 patchlevel 4 ....
    Build System: Linux i686
    Copyright (C) ....
    Terminal type set to 'wxt'
> f(x)=3*x**2*cos(x)
  # The function f(x) = 3x^2 cos(x) is defined
> print f(1), sin(pi/2), cos(pi/2)
  #  f(x) at x = 1, others at x = π/2.
```

```
1.62090691760442 1.0 6.12323399573677e-17
  # values are 1.62090691760442, 1.0,
  6.12323399573677 × 10^-17 = 0
```

- The *first line* is the same for all examples to follow. It shows the name of your computer, working directory, and what you have typed just after the $ sign to start Gnuplot.

We shall begin with our first real example.

Ex.1. To plot the angular distribution of the *radiated power* as given in Eq. (12). Here let us note that in the 3-dimensional spherical coordinate system the the polar angle θ is measured from the Z axis, as shown in Fig. 1. In the plane polar coordinate system, on the other hand, the polar angle is measured from the X axis. Since we are going to plot a 2-dimensional curve, the latter coordinate system has to be used and the same function $\sin^2\theta$ written in Eq. (12) has to be written as $\sin^2(\pi/2-\theta) = \cos^2\theta$. We now write the commands.

```
> set polar #  polar coordinates
    dummy variable is t for curves
> set grid polar # draw grid lines
> set trange [0:2*pi] # range of θ
> set rrange [0:1.2] # range of r
> set size square # shape of the plot area
> set xrange [-1.2:1.2]; set yrange [-1.2:1.2]
  # range of x,y coordinates
> plot ((cos(t))**2) # the plot is drawn instantly
> set term fig color portrait size 15, 15
metric pointsmax 1000 solid font "Times-Roman,12"
  # Terminal specification
Terminal type set to 'fig'
Options are 'color small pointsmax 1000 portrait
metric solid textnormal font "Times Roman,12"
linewidth 1 depth 10 version 3.2 size 15 15'
> set title "S4dpol-150623.fig"
  # a label to appear on the plot
> set out "S4dpol-150623.fig"
  # to save as, file name
> replot # replots and saves
```

- The default coordinate system is Cartesian. To plot in the polar system, the command "set polar" is necessary. The function has to be written in the form $r = f(\theta)$. The letter t stands for θ, when plotting "polar".

- Immediately after the command "plot ((cos(t))**2)" the plot appears on the screen, but we are unable to save it in the computer. For this purpose we have to specify a "terminal". The saved plot will look identical with the one now on screen, if we choose "set term png". However, we are more comfortable with the terminal "fig" for editing using Xfig. The command "set term fig ..." not only sets the terminal, but also makes further specifications, e.g, solid and coloured lines for plots as well as grids, font name and size for labelling, size of the screen 15 cm × 15 cm, etc. The plot is now stored in the working directory with extension ".fig".

- It has been our practice to add today's date (yymmdd) as part of the file name (which the reader need not follow.) We now start Xfig, look for the file name, bring the plot on the screen, for which we shall use the term "canvas". We work on this canvas and do some editing. For example, we give extra labels, draw

X, Z coordinate axes, change the grid lines from "solid" to "dashed", change the colours of the curve, as we feel necessary. See Ex.2 for a better example.

- Note that the "first line" pointed out in Ex.0, and "the terminal specification" in Ex.1 will be common for all Exercises to follow, so that we shall not repeat them any more.

We shall refer to the entire set of commands given in Ex.1 as "script" (term borrowed from Gnuplot Cookbook.) The script written in Ex.1 results in the plot shown in Fig. 3.

5 Plotting the Electric Field from the Oscillating Dipole in the Near Zone

The term "plotting the field" here means plotting the field lines, i.e., imaginary curves in space such that the **E** field at any point in space is tangential to such a curve passing through that point. But why are we restricting ourselves to the **E** field only, ignoring the **B** field? The answer comes from Eqs. (3). The **B** field has only ϕ-component, i.e., it is pointing in the direction of the $\mathbf{e}_\phi$ vector, so that the field lines are coaxial circles around the Z axis (see Figs. 1 (c), 15(b).) In contrast the **E** field has both r and θ components (i.e., having components along $\mathbf{e}_r$ as well as $\mathbf{e}_\theta$ vectors.) They create interesting and beautiful patters in space, which we need to see, appreciate and admire. Also, **B** accompanies **E** everywhere, being comparatively weaker in the near zone, but equallly strong in the radiation zone. Hence the "graph" of the strength of **E** field is also a graph of the strength of **B** field, especially in the radiation zone. See Fig. 13.

We shall first set up the general differential equation in the *spherical polar co-ordinate system* for field lines of **E** and **B** that are symmetrical about the Z-axis (azimuthal symmetry) so that one complete field line, from beginning to end, is confined to an azimuthal plane (ϕ = constant). For convenience of drawing we have taken this plane to be the XZ plane in Fig. 4(a), which shows a part such a field line. The points P and Q are infinitesimally close points on this curve at radius vectors $\mathbf{r}$ and $\mathbf{r} + d\mathbf{r}$ respectively, so that $d\mathbf{r} = dr\,\mathbf{e}_r + r\,d\theta\,\mathbf{e}_\theta$. The lines of **E** or **B** being tangential to the field line at every point, $E_r/E_\theta = dr/r\,d\theta$ for electric field lines; $B_r/B_\theta = dr/r\,d\theta$ for magnetic field lines. Confining ourselves to electric field, let the field lines be represented by the family of curves $C : r = f(\theta, k)$ where k is a constant. It follows that

$$\frac{dr}{r\,d\theta} = \frac{f'(\theta)\,d\theta}{f(\theta)d\theta} = \frac{E_r}{E_\theta}; \quad \text{or,} \quad \boxed{\frac{f'(\theta)}{f(\theta)} = \frac{E_r}{E_\theta}}. \tag{13}$$

Eq. (13) gives the differential equation[8] for azimuthally symmetrical field lines at any point in space where $E_\theta \neq 0$. We shall now use this equation to plot the **E** field from the oscillating dipole for the near zone. We shall plot the field at $t = 0$, so that $\cos\omega t = 1$

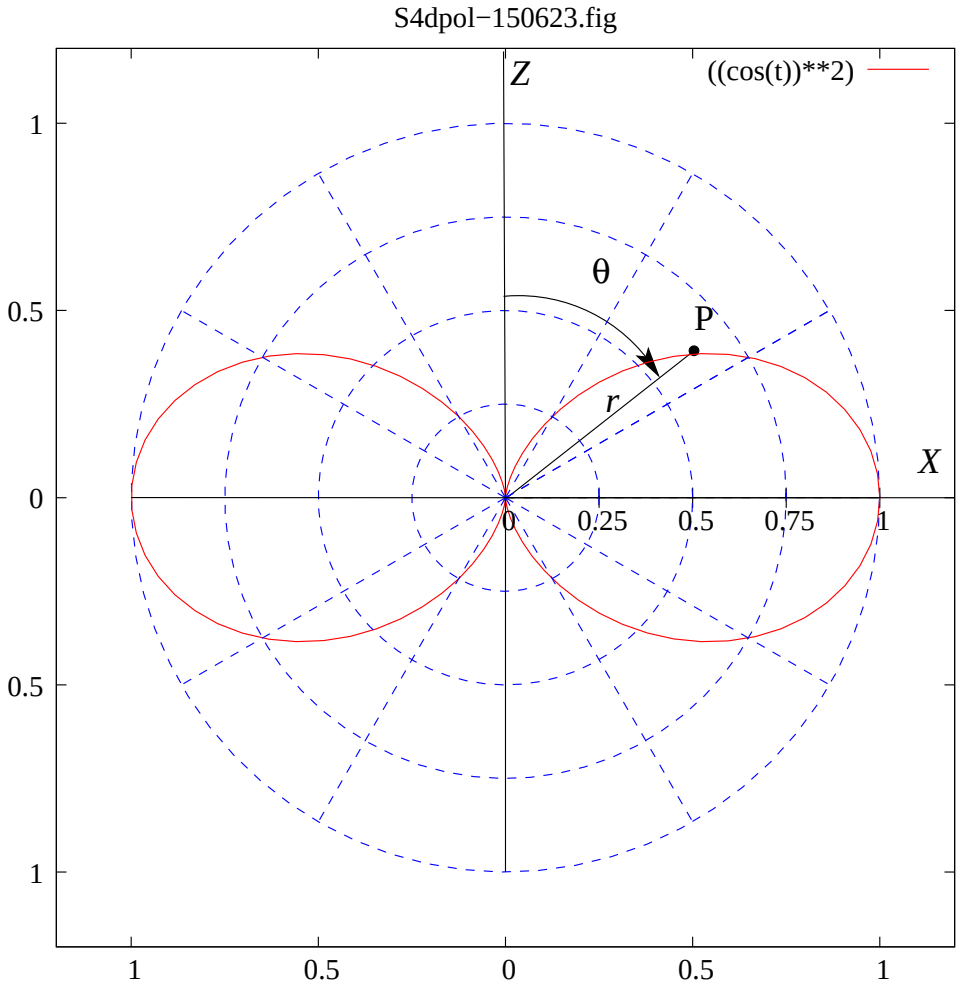


Figure 3: *Angular distribution of the energy flow from an oscillating Electric Dipole*

in Eq. (8a). The oscillating dipole is at its peak value and is pointing up i.e., towards the positive Z axis. Also, $E_r/E_\theta = \frac{2\cos\theta}{\sin\theta} = 2\cot\theta$. Hence,

$$\frac{df}{d\theta} = 2\cot\theta\, f. \qquad \text{Or,} \quad \frac{df}{f} = 2\cot\theta\, d\theta. \quad (a)$$

$$\text{By integration} \quad r = f(\theta, k) = k\sin^2\theta. \quad (b) \qquad (14)$$

Eq. (14b) gives the family of dipole field lines, different members of this family corresponding to different, positive values of the constant k. In Fig.4(b) we have shown three such curves on each side of the Z-axis. We created them with Gnuplot, added subsequently "arrowheads" manually to indicate the direction of the field at some selected points.

Note that for each value of k there is one

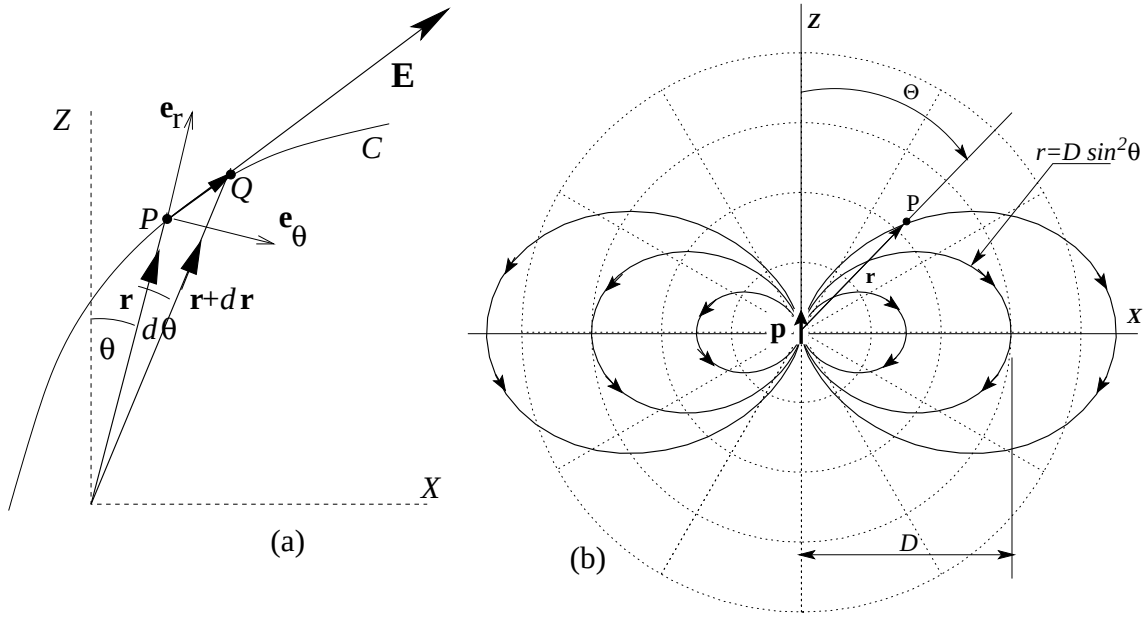


Figure 4: *(a) Plotting field lines; (b) Quasistatic* **E** *lines from an oscillating electric dipole*

pair of symmetrical lobes about the Z axis on the XZ plane (looking somewhat like the wings of a butterfly), representing a dipole field. These contour curves exist symmetrically around the Z axis, covering the vicinity of the oscillating dipole. To evaluate k for a particular field line Γ we have to find the distance D of a point P on the XY plane through which Γ passes. Since $\theta = \pi/2$ for such a point, it follows that $k = D$. We shall now demonstrate the actual plotting of the field lines.

Ex.2. To plot the field lines from the oscillating dipole at $t = 0$. This example also illustrates the use of "do" command and "line type".

```
> set title "Ed-150623.fig"
> set out "Ed-150623.fig"
> Ed(n,t)=0.2*(1+n)*((cos(t))**2)
> do for [n=0:4] {plot Ed(n,t) lt n}
```

- The script written above has been continued from Ex.1.
- The defined function Ed(n,t) has 2 arguments, t for θ and n for the "iteration" number = 0,1,2,3,4 as specified in the next "do" command. It creates five plots, intersecting the X axis at $\pm(0.2, 0.4, 0.6, 0.8, 1.0)$
- The sub-command"lt n" specifies "line type" for each value of n. In term fig, lt 0= black, 1= red, 2= light green, 3= dark blue, 4= magenta. However, we changed the colour of lt 2 from light green to dark green for better visibility,

the grid lines from solid to dashed, their colours from black to blue.

The actual plot is shown in Fig.5.

6 Fieldlines from Oscillating Dipole

Plotting the **E** field as given by Eq. (3) is much more difficult, because in this case the differential equation (13) is not so easy to integrate. However we shall follow the path given by Orfanidis[7] and achieve a wonderful result.

Let us set $P = \frac{k^3 p_o}{4\pi\epsilon_0}$; $\rho = kr$, $\tau = \omega t$. The EM field $(\mathbf{E}, \mathbf{B})$ has only three non-zero components, as seen from Eq. (3):

$$\begin{aligned} E_r &= \frac{P}{\rho^2}\left[\frac{\cos(\rho-\tau)}{\rho} + \sin(\rho-\tau)\right] 2\cos\theta. \qquad (a) \\ E_\theta &= \frac{P}{\rho^3}\left[(1-\rho^2)\cos(\rho-\tau) + \rho\sin(\rho-\tau)\right]\sin\theta. \quad (b) \\ cB_\phi &= \frac{P}{\rho^3}\left[-\rho^2\cos(\rho-\tau) + \rho\sin(\rho-\tau)\right]\sin\theta. \qquad (c) \end{aligned} \tag{15}$$

We shall plot only the **E** field, ignoring the **B** field, for the reason cited at the beginning of Sec. 5. The **B** field lines on the XY plane are coaxial circles around the Z-axis, as shown later, in Fig.15.

Plotting implies a "picture" of the field with time frozen, i.e., with t held constant. We shall set

$$\begin{aligned} &\Psi(\rho) \equiv \frac{\cos(\rho-\tau)}{\rho} + \sin(\rho-\tau). \\ &\text{so that } \tfrac{d\Psi}{d\rho} = \\ &-\tfrac{1}{\rho^2}\left[(1-\rho^2)\cos(\rho-\tau) + \rho\sin(\rho-\tau)\right], \end{aligned} \tag{16}$$

with τ held constant. It follows that

$$\begin{aligned} E_r &= \tfrac{P}{\rho^2}\Psi(\rho)\, 2\cos\theta; \\ E_\theta &= -\tfrac{P}{\rho}\tfrac{d\Psi}{d\rho}\,\sin\theta. \end{aligned} \tag{17}$$

Then from (13)

$$\begin{aligned} &\tfrac{dr}{d\theta} = r\tfrac{E_r}{E_\theta}, \Rightarrow \tfrac{d\rho}{d\theta} = \rho\tfrac{E_r}{E_\theta} = -\tfrac{\Psi}{\frac{d\Psi}{d\rho}} 2\cot\theta. \\ &\Rightarrow\ \tfrac{d\Psi}{d\theta} = -2\Psi\cot\theta. \end{aligned} \tag{18}$$

This differential equation is similar to the one in Eq. (14a), except for a negative sign on the right side. Solving it we get

$$\begin{aligned} &\Psi(\rho)\sin^2\theta = C \\ &\Rightarrow \left[\tfrac{\cos(\rho-\tau)}{\rho} + \sin(\rho-\tau)\right]\sin^2\theta = C' \end{aligned} \tag{19}$$

where C' is a constant.

We now return to the original variables and write the above equations of the field lines as

$$\left[\frac{\cos(kr-\omega t)}{kr} + \sin(kr-\omega t)\right]\sin^2\theta = C \tag{20}$$

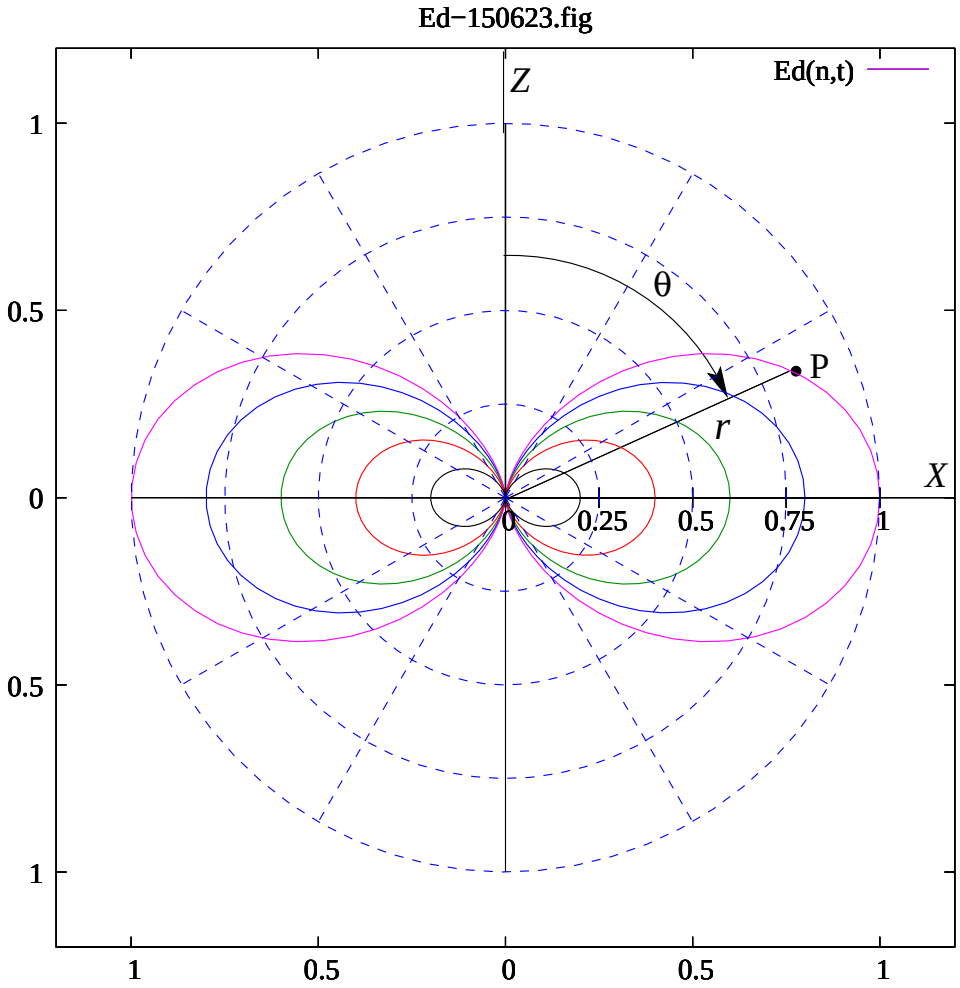


Figure 5: **E** *field lines intersecting the* X*-axis at* $x = \pm(.2, .4, .6, .8, 1.0)$

Before we start plotting let us be careful about what this plotting operation will involve. We are going to plot the **E** field lines on a predecided fixed plane Σ which passes through the line of the oscillating dipole, the same as the Z axis. This plane is to be identified as the XZ plane. The field lines will be shown as contour plots of the function that appears on the left side of Eq. (20), for a fixed value of t. The above function will now be considered to be a function of the *Cartesian* coordinates (x, z), with y set to zero. This is because Gnuplot can make contour plots of functions only of Cartesian coordinates.

When Gnuplot plots either a surface, or a series of contours on that susrface, it expects the equation of the surface to be written in Cartesian coordinates as $z = f(x, y)$. For this

purpose we shall *change the variable* $z \to y$, denote the left side of Eq. (20) as $\psi(x, y, t)$, and prepare to plot contours on the *surface* defined as:

$$z = \psi(x, y, t) \stackrel{\text{def}}{=} \left[\frac{\cos(kr - \omega t)}{kr} + \sin(kr - \omega t)\right] \sin^2\theta. \tag{21}$$

In this change of variables the surface is raised above the XY plane (instead of the XZ plane). The Z axis now represents the function $\psi(x, y, t)$. *Plotting the field lines* for a given value of t is the same as plotting contours on this surface for selected levels, i.e., for selected values of z, with t held constant. The constant C appearing on the right side of Eq. (20) represents one of those "selected evels". After the plotting operation is done, at some convenient point we shall restore the y-coordinate to the status of the z-coordinate.

The contours on this surface are now given by the equation

$$z = \psi(x, y, t) = \left[\frac{\cos(kr - \omega t)}{kr} + \sin(kr - \omega t)\right] \sin^2\theta = C. \tag{22}$$

For every time $t > 0$ and every real value of and C lying within $z_{\max}$ and $z_{\min}$ there is a contour .

Now suppose we take the unit of distance as the wavelength λ, and the unit of time as the time period the harmonic oscillation T, i.e.,

$$\rho = kr = \tfrac{2\pi r}{\lambda} \to 2\pi r;\ \tau = \omega t = \tfrac{2\pi t}{T} \to 2\pi t,$$

Hence,

$$\left[\frac{\cos 2\pi(r - t)}{2\pi r} + \sin 2\pi(r - t)\right] \sin^2\theta = C \tag{23}$$

becomes the equation of the field lines in the polar coordinate sytem.

We shall illustrate this contour plot for the field at $t = 0$, so that

$$z = \psi(x, y) \equiv \psi(x, y, t = 0) = \left[\tfrac{\cos 2\pi r}{2\pi r} + \sin 2\pi r\right] \sin^2\theta = C. \tag{24}$$

Alternatively,

$$\left[\tfrac{\cos\rho}{\rho} + \sin\rho\right] \sin^2\theta = C.$$

Gnuplot will not only plot the surface $z = \psi(x, y)$, but show selected contours on this surface, and project them on the XY plane, as we shall demonstrate in the next section.

7 Plotting the E Field at $t = 0$

We shall first draw the field lines at $t = 0$. Note that this instant $t = 0$ is not the begnning of time. The dipole has been oscillating forever, from $t = -\infty$ to $t = +\infty$. However the zero time is taken to be one of those instants when the dipole achieves its peak value, pointing upwards (i.e., in the +Z direction).

For a better view of the details of the field lines limit ourselves to a small region around the origin, within a radius of 2 wavelengths.

It should be remembered that we have chosen the length scale in the unit of the wavelength λ (just before Eq. 23). Therefore our viewing region has a radius of $r = 2$ arond the Z-axis. Before plotting the field lines it will be necessary to decide the values of ψ at which the contours will be drawn. For example we may like the contours to pass through some selected points on the X axis, say, $x = 0.8, 0.9, 1.0, 1.1$. However the contour lines need to be more dense. Hence we shall add more lines. The selection process and the subsequent plot is done in three steps.
Step 1: Selection of the contour levels.

To make the selection we have plotted the $\psi(x, y)$ function along the x-axis, i.e., we have plotted the function

$$\psi(x) \equiv \psi(x, 0) \tag{25}$$

in the range [0.1,2.0], avoiding the origin where the function goes to infinity.

Ex.3. To plot the function $\psi(x)$ along the X-axis.

```
> ro(x)=2*pi*x # defines ρ(x) as in Eq.(23).
> psi(x) = cos(ro(x))/ro(x) + sin(ro(x))
   # defines ψ(x) as in Eq.(24)
> set grid # sets grid lines
> set xtics 0.1; set ytics 0.5; set mytics 5
   # sets tic marks on the X and Y axes
> set xrange [0.1:2]; set yrange [-1.1:2]
   # sets the ranges of x,y values
> set title ‘‘psi(x)-150609.fig"
> set out ‘‘psi(x)-150609.fig"
> plot psi (x)  #     2-D plot
```

The plotted function is shown in Fig.6. It comes with the title "psi(x)-150609.fig", as per the command given in the 3rd line from the bottom of the command chain. On that plot we have indicated

(1) the values $\psi = -0.89, -0.44, 0.16, 0.70$ corresponding to the following intercepts of the $\psi(x, y)$ function on the x-axis: $x = 0.8, 0.9, 1.0, 1.1$. The *exact* values of ψ can be obtained by writing the command:

```
> print psi(0.8),psi(0.9),psi(1.0),psi(1.1)
-0.889579538602437 -0.444719637070117
0.159154943091895 0.704838937474402
```

(2) zeros of ψ at $x \approx 0.44, 0.98, 1.49, 1.99, ...$
(3) maxima of ψ at $x \approx 1.22, ...$
(4) minima of ψ at $x \approx 0.7, 1.73, ...$
We have shown the above points on the plot"psi(x)-150609.fig", below the x axis..

We selected the preliminary contour levels at $z = -0.89, -0.44, 0.16, 0.70$ from consideration of (1), and extra levels at $z = 1, 1.3, -0.6, 0.5, 0.85$ from consideration of (2),(3),(4). We have indicated the above values on the left side and on the right side of the plot.

Step 2: Plotting $z = \psi(x, y)$, as a 3D surface, and show the contours. We have achieved the objective through the following commands:

Ex.4. To plot the surface $\psi(x, y)$ and contours on it.

```
> r(x,y) = sqrt(x*x+y*y) # defines r =√(x²+y²)
> st(x,y)=x/r(x,y)
   # defines sin θ
> sr(x,y) = sin(2*pi*r(x,y))
   # defines sin ρ = sin 2πr
> cr(x,y) = cos(2*pi*r(x,y))
   # defines cos ρ = cos 2πr
> psi(x,y) =( cr(x,y)/(2*pi*r(x,y))
  + sr(x,y) )*(st(x,y))**2
   # defines ψ(x,y) as in Eq.(24)
> set size square
   # shape of the plot area
> set surface # surface plot
> set contour both
   # contour on the surface as well as
```

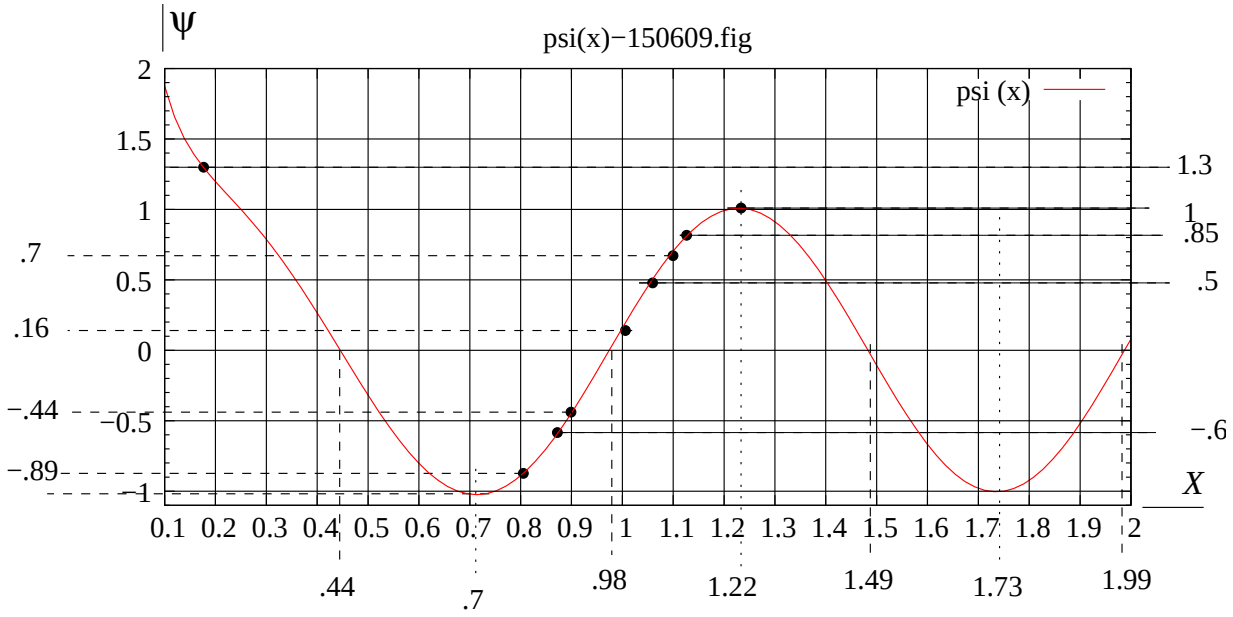


Figure 6: *The function $\psi(x, y)$ drawn along the X-axis, for selection of contour levels*

```
   on the base
> set hidden3d # treat surface as opaque
> set cntrparam levels discrete - 0.89,
  -0.44, 0.16, 0.7, 1, 1.3, -0.6, 0.5, 0.85
   # contour parameters, discrete levels
> set xrange [-2:2]; set yrange [-2:2]
   # domains of x and y
> set xlabel "x-axis"; set ylabel "y-axis"
   # labels on the x and y axes
> set zlabel "psi(x,y)"
   # label on the z axis
> set isosamples 50,50; set samples 10,10
> splot psi(x,y)
   # 3-D surface plot. It is interactive
> set term fig color portrait size 15, 15,
  metric pointsmax 1000 solid font
  "Times-Roman,12"
   # Terminal specification
> set title "psi(x,y)-150610.fig"
   # label the plot
> set out "psi(x,y)-150610.fig"
   # save as, the file name
> replot
```

- The "splot" command, before specification of the terminal, results in a 3-D plot which is *interactive*. You can change the view angle, choose the best perspective, then specify the terminal. Finally when you give the "replot" command, the plot will be saved and seen as you last saw it, before setting the terminal.

The resulting plot is shown in Fig.7

Step 3: Plotting the Field lines of **E**. This is achieved through the following commands, starting with "unset surface" to make sure that the surface will not be drawn again.

Ex.5 Plotting field lines on the XZ plane.

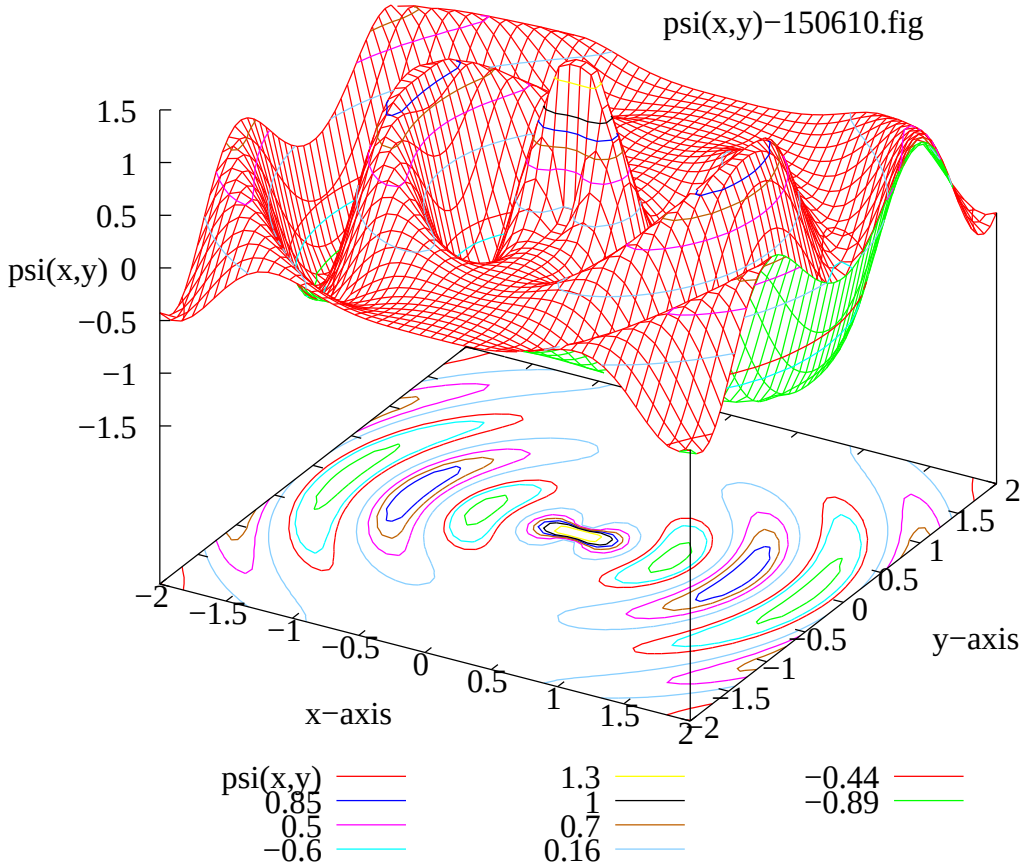


Figure 7: *The isometric plot of the surface function $\psi(x, y)$, the z axis representing the height of the surface above the XY plane. The contours at the selected levels are shown on the surface as well as on the XY plane.*

```
> unset sur
> set view map
> set key bmargin
> set title "psi(x,y)cont-150610.fig"
> set out "psi(x,y)cont-150610.fig"
> splot psi(x,y)
```

- The script written above has been continued from Ex.4.

The resulting plot is shown in Fig.8

8 Planting the E vectors

The **E** field lines drawn in the last section may look impressive, but is deficient on one count. There is no indication of the direction of the field along the lines. One can remedy this defect in one of following two ways.
(a) Actual planting of the **E** vector at selected points of the region under scrutiny and, by superimposing the plot of Fig.8 on the same

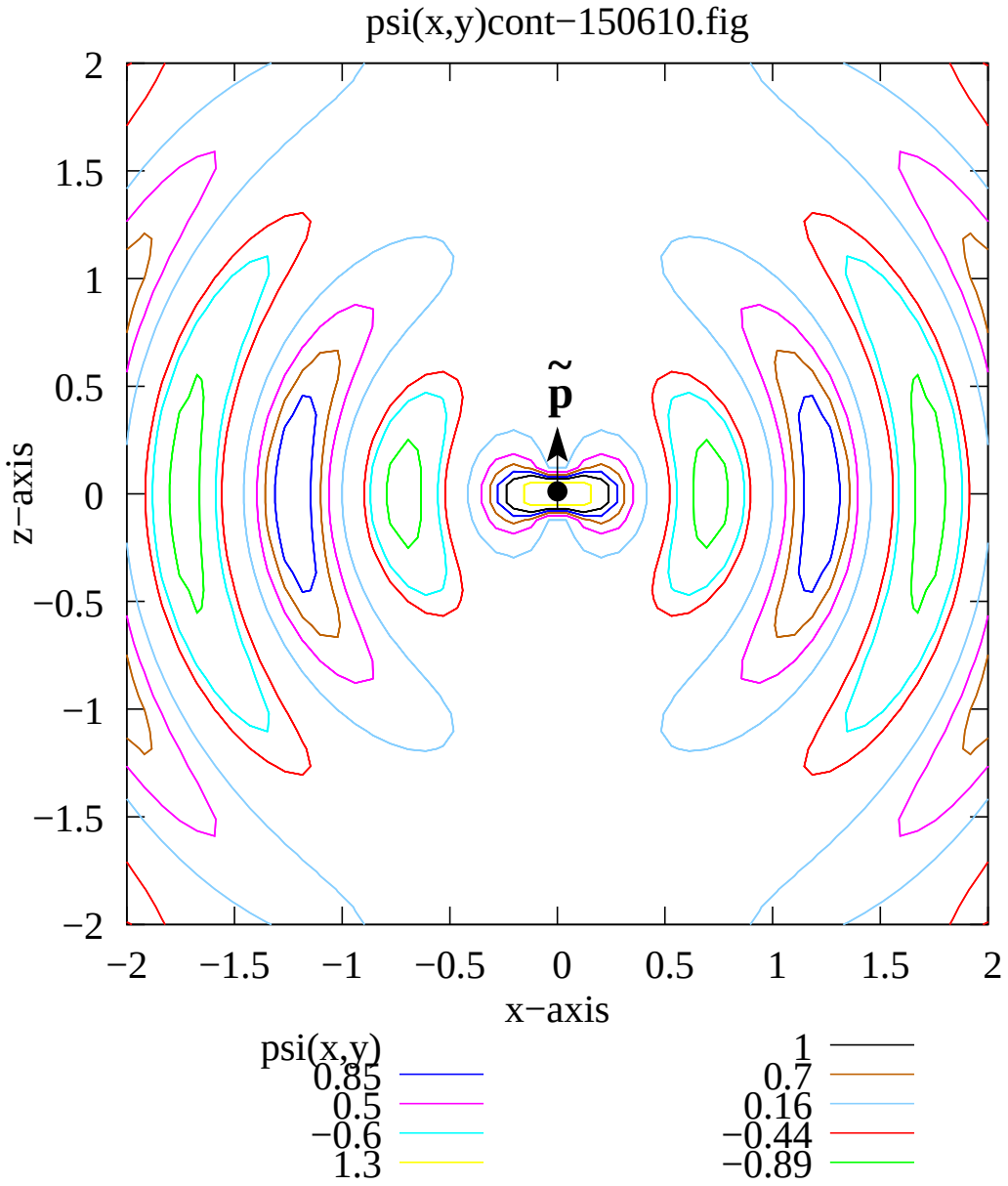


Figure 8: *The field lines of* **E** *on the XZ plane at time* $t = 0$*, due to an oscillating electric dipole* $\tilde{\mathbf{p}}$ *oriented along the Z axis, and placed at the origin. We have added a small circular blob, with an arrow piercing through it (not part of the plot), to indicate the location and the direction of the dipole at* $t = 0$*.*

canvas, get a clear indication of the direction of the field along the field lines;
(b) Make a plot of the value of the transverse component of the field, namely $E_\theta(x)$, along the x axis, which gives the exact value of the field along this axis, since there is no E_r component on the XY-plane. Since $\mathbf{e}_\theta = -\mathbf{k}$ on the XY plane, *positive value of $E_\theta(x)$ implies direction of the $-Z$ axis and vice versa.* Follow these directions along entire field lines.

Each option has its advantage and disadvantage. In option (a) the plotter gets a clear indication of the **E** planted all over the space, with its both magnitude and direction in display. However, the procedure is laborious, because it involves creating a "data table" of (x, y, E_x, E_y) for selected points. In comparison, option (b) is relatively quicker and easier to implement.

In this section we shall take up the first option, i.e., option (a).

Gnuplot can "plot vectors". We are using the term "planting vectors" for the same operation. Page 56 of Gnuplot Manual tells us how to do this.

> The 2D **vectors** style draws a vector from (x,y) to (x+xdelta, y+ydelta). The 3D **vector** style is similar, but requires 6 columns of basic data. A small arrowhead is drawn at the end of each vector.
> 4 columns: x y xdelta ydelta
> 6 columns: x y z xdelta ydelta zdelta
> The keyword "with vectors" may be followed by an in-line arrow style specification.

.... plot ... with vectors filled heads

Therefore the operation "planting vectors" begins with preparation of a 4 column data table, in which each row will have four entries, namely the (x,y) componets of the field point, followed by the (E_x, E_y) components of the **E** field at that point.

At this point let us note that the formulas given in Eqs. (15) give the (r, θ) components of the **E** field. They are to be converted to the (x, y) components of the field by the following formulas.

$$\begin{aligned} E_x &= E_r \sin\theta + E_\theta \cos\theta \\ E_y &= E_r \cos\theta - E_\theta \sin\theta \end{aligned} \qquad (26)$$

We shall now carry out this operation through the following steps.

Step 1: Calculating the values of (x, y, E_x, E_y) for selected distances d measured from the origin, as a prelude to the creation of the Data Table.

We have selected d in the range of 0.1 to 1.6, at a fixed interval of 0.1, and their locations at six values of the polar angle, measured from the Z axis, equal to $\theta = n\pi/12$; $n = 1, ..., 6$. The region under observation is the first quadrant of the XZ plane, i.e, $0 \le \theta \le \pi/2$; $\phi = 0$. In other words $z > 0$; $x > 0$; $y = 0$; $d = f\lambda = f$, since $\lambda = 1$, and $f = 0.1, 0.2, 0.3, \cdots, 1.6$.

We shall illustrate the procedure here only for $d = 1$ and $d = 0.5$.

The reader should interpret all y coordinates written in the commands below as the z coordinate, following the comments made on page 15.

Step 1A

Ex.6. Calculating the coordinates (x, y) of the selected points.

```
> x(n)=d*sin(n*pi/12); y(n)=d* cos(n*pi/12)
   # x(n), y(n) defined
> d=1   #    d=1
> do for [n=1:6] {print x(n)}
   # values of x(n) for d = 1; n = 1, ..., 6
0.258819045102521
0.5
0.707106781186547
0.866025403784439
0.965925826289068
1.0
> do for [n=1:6] {print y(n)}
   # values of y(n) for d = 1; n = 1, ..., 6
0.965925826289068
0.866025403784439
0.707106781186548
0.5
0.258819045102521
6.12323399573677e-17
> d=.5  #    d=.5
> do for [n=1:6] {print x(n)}
0.12940952255126
0.25
0.353553390593274
0.433012701892219
0.482962913144534
0.5
> do for [n=1:6] {print y(n)}
0.482962913144534
0.433012701892219
0.353553390593274
0.25
0.12940952255126
3.06161699786838e-17
```

Step 1B

Ex.7. Calculating E_x, E_y at the selected points at $t = 0$.

```
> r(x,y)=sqrt(x*x + y*y)
> ro(x,y) = 2*pi*r(x,y);
  ro2(x,y)= (ro(x,y))**2
> ro3(x,y) =(ro(x,y))**3
   # ρ, ρ^2, ρ^3 defined
> sr(x,y)=sin(ro(x,y)); cr(x,y)=cos(ro(x,y))
   # sin ρ, cos ρ defined
> st(x,y)=x/r(x,y); ct(x,y)=y/r(x,y)
   # sin θ, cos θ defined
> Er(x,y) = (cr(x,y)/ro3(x,y)
  + sr(x,y)/ro2(x,y)) * 2 * ct(x,y)
> Et(x,y)=(cr(x,y)/ro3(x,y)
   # (E_r, E_θ) as defined in Eqs.(15)
  + sr(x,y)/ro2(x,y)-cr(x,y)/ro(x,y))*st(x,y)
> Ex(x,y)=Er(x,y)*st(x,y)+Et(x,y)*ct(x,y)
> Ey(x,y)=Er(x,y)*ct(x,y)- Et(x,y)*st(x,y)
   # (E_x, E_y) from Eqs.(26)
> d=1.0 \# d=1
> do for [n=1:6] {print Ex(d*sin(n*pi/12),
  d*cos(n*pi/12))}
   # values of E_x(x, y) for d = 1; n = 1, ..., 6
-0.0367651544198614
-0.0636791154033154
-0.0735303088397228
-0.0636791154033154
-0.0367651544198614
-9.00486573608828e-18
> do for [n=1:6] {print Ey(d*sin(n*pi/12),
  d*cos(n*pi/12))}
  # values of E_y(x, y) for d = 1; n = 1, ..., 6
0.0179140770447072
0.0448280380281612
0.0815931924480226
0.118358346867884
0.145272307851338
0.155123501287745
> d=0.5    #            d=.5
> do for [n=1:6] {print Ex(d*sin(n*pi/12),
  d*cos(n*pi/12))}
0.0553888207210481
0.095936251660179
0.110777641442096
0.0959362516601791
0.055388820721048
1.35663484009156e-17
```

Step 2; Constructing the Data Table

This was done in the following way (1) Start a new text document by invoking Libre Office Writer. (2) Save it as a text document with extension .txt. In this case the name of this file is 'E1vecdataC-150423.txt'. (3) Choose any one of the selected points. Copy-paste the values of x,y from Step 1A, and the values of Ex,Ey from Step 1B (after rounding them off to 2 or 3 decimal places.). The 4 numbers are now displayed not in a row, but in a column, as in

a

b

c

d

(4) take cursor just after the top number in this row (i.e., a), go further by one space, and click del. The next lower number (i.e., b) now comes next to the first number with one blank space in between. In this way bring all the numbers in one row. As a result the copied numbers are now rearranged as
a b c d
Continue this procedure to bring the (x,y,Ex,Ey) values of all selected points in as many rows. The composition of this file is now complete.

The (partial) Data Table below shows the data for the 12 selected points corresponding to $d = 1.0, 0.5$.

```
# x y  Ex  Ey
.26 .97 -.037 .018 # d=1
.5 .87 -.064 .045
.71 .71 -.074 .082
.87 .5 -.064 .118
.97 .26 -.037 .145
1.00 0 0 .155
.13 .48 .06 -0.08 # d=0.5
.25 .43 .1 -.12
.35 .35 .11 -.18
.43 .25 .1 -.23
.48 .13 .06 -.27
.5 0 0 -.29
```

The full data table, saved as **'E1vecdataC-150423.txt'** in the working directory, is shown in the Appendix.

Step 3: Planting the **E** *vectors on the first quadrant of the ZX Plane.*

Ex.8.

```
> set  size square
> set  pointsize 0.5
> set xrange [-0.1:1.7];
  set yrange [-0.1:1.7]
  # sets x-range and z-range
> set xtics 0.5; set mxtics 5
  # sets tic marks along the x-axis
> set ytics 0.5; set mytics 5
  # sets tic marks along the z-axis
> set title ‘‘E1vecB-150505.fig"
  # title that appears at the top
  of the plot
> set out ‘‘E1vecA-150505.fig"
  # filename of the plot, stored in UrDir
> set key bmargin
> plot 'E1vecdataC-150423.txt' using
  1:2 with points pt 7,
  'E1vecdataC-150423.txt' using
  1:2:3:4 with vectors size .02, 15
  filled lt 3
```

- The last command plants (a) the field points from col 1,2 of the data file 'E1vecdataC-150423.txt' with 'point type' 7 (circular blob), (b) the vectors from col 1, 2, 3, 4 of the same data file, as straight lines, starting at these field points, and terminating at an arrowhead. The size of the arrowhead is indicated in the subcommand 'with vectors size .02, 15 filled lt 3' (length = .02,

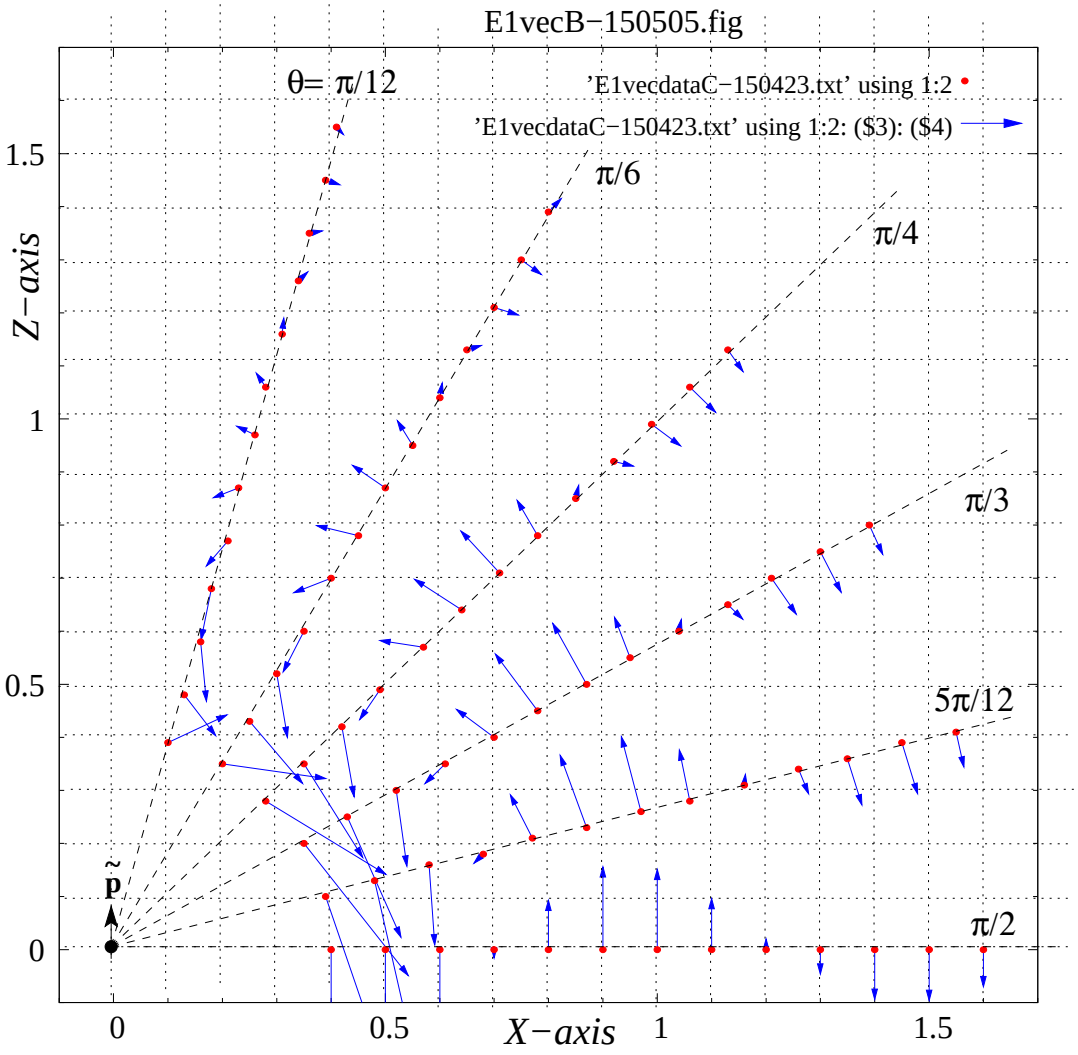


Figure 9: *The* **E** *vectors on the* XZ *plane at time* $t = 0$*, shown on the first quadrant. They are planted at selected points marked with red dots. These points are distributed along radial lines at equal length intervals of 0.1 from* $d = 0.4$ *to* $d = 1.6$*, and at equal angular intervals of* $\pi/12$ *from* $\theta = \pi/12$ *to* $\theta = \pi/2$*. The oscillating dipole is shown at the origin as* $\tilde{\mathbf{p}}$

sloping angle of the arrows = 15^0, line type 3.)

- If we want to make the vectors look longer, by increasing their size by, say 50%, the *second part* of the last command should be modified to:

```
... 'E1vecdataC-150423.txt' using
1:2:(1.5*$3):(1.5*$4) with vectors
size .02, 15 filled lt 3.
```

See pp.72,169 of Gnuplot Manual, Example on p.28 of Gnuplot Cookbook.

We have shown the planted **E** vectors in Fig. 9.

Step 4: Plotting the Field lines

Ex.9

```
> r(x,y) = sqrt(x*x+y*y)
> st(x,y)=x/r(x,y)
> sr(x,y) = sin(2*pi*r(x,y));
  cr(x,y) = cos(2*pi*r(x,y))
> psi(x,y)=( cr(x,y)/(2*pi*r(x,y))
  +sr(x,y) )*(st(x,y))**2
> set contour base
> unset surface
> set view map
> set cntrparam levels discrete -0.89,-0.44,
  0.16,0.70,1,1.3,-0.6,0.5,0.85
> set isosamples 80,80
> set title "E1lineD-150505.fig"
> set out "E1lineD-150505.fig"
> splot psi(x,y)
```

- The commands are a continuation of those in Ex.8.

- The contur plots, even though they look 2-dimensional, are to be treated as surface plots (3-dimensional). Hence the "splot" command.

We have shown the field lines in Fig. 10.

Step 5: Superimposing the Planted Vectors on the Field lines This operation does not involve Gnuplot. It is mostly a copy-paste operation done in xfig, and shown in Fig. 11. The canvas is divided into four partitions: (a), (b), (c), (d). We have copied Figs. 10 and 9, scaled down to about half their dimensions, and pasted them in partitions (a) and (b). These two figures are superimposed in partition (c), in which we find the field vectors embedded in the neighbourhood of the field lines. We extrapolate their directions into the field lines, in partition (d), by drawing short tangents along the curve with arrowheads, and get directed field lines.

9 Plotting the E Field at equal time intervals of $T/8$ for one full period T.

We shall now take up the option (b) mentioned at the beginning of Sec.8, not just for $t = 0$, but for eight instants of time, taken at equal intervals spread over one full period of oscillation of the dipole, namely, $t = 0, T/8, 2T/8, 3T/8, 4T/8, 5T/8, 6T/8, 7T/8$. We shall first make two series of plots for each one of the above instants, namely, (i) plot the field lines on the XZ plane over a square region: $[-2.5\,\lambda \le x \le 2.5\,\lambda, -2.5\,\lambda \le z \le 2.5\,\lambda]$; (ii) plot E vs x along the X-axis for the same x-range $-2.5\,\lambda \le x \le 2.5\,\lambda$. The

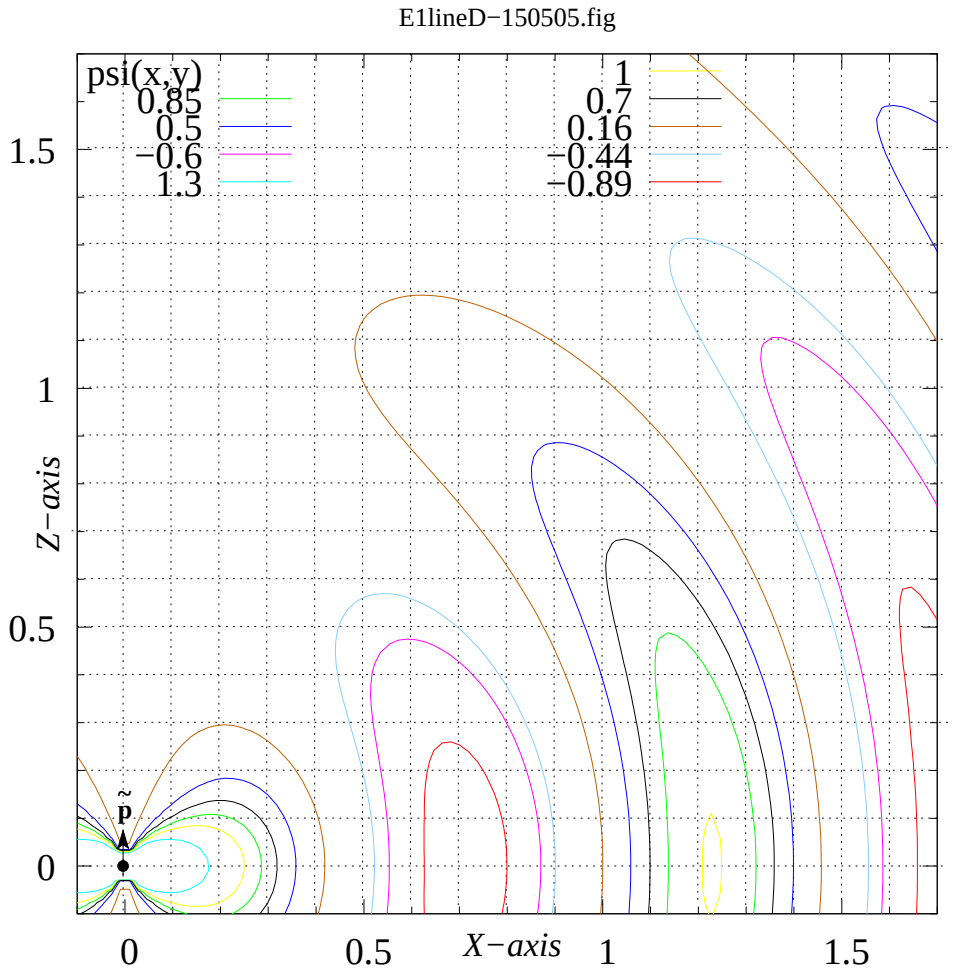


Figure 10: *The* **E** *field lines on the XZ plane at time* $t = 0$*, shown on the first quadrant, drawn in the x-range [-0.1:1.7] and z-range [-0.1:1.7]. The oscillating dipole placed at the origin is shown as* $\widetilde{\mathbf{p}}$.

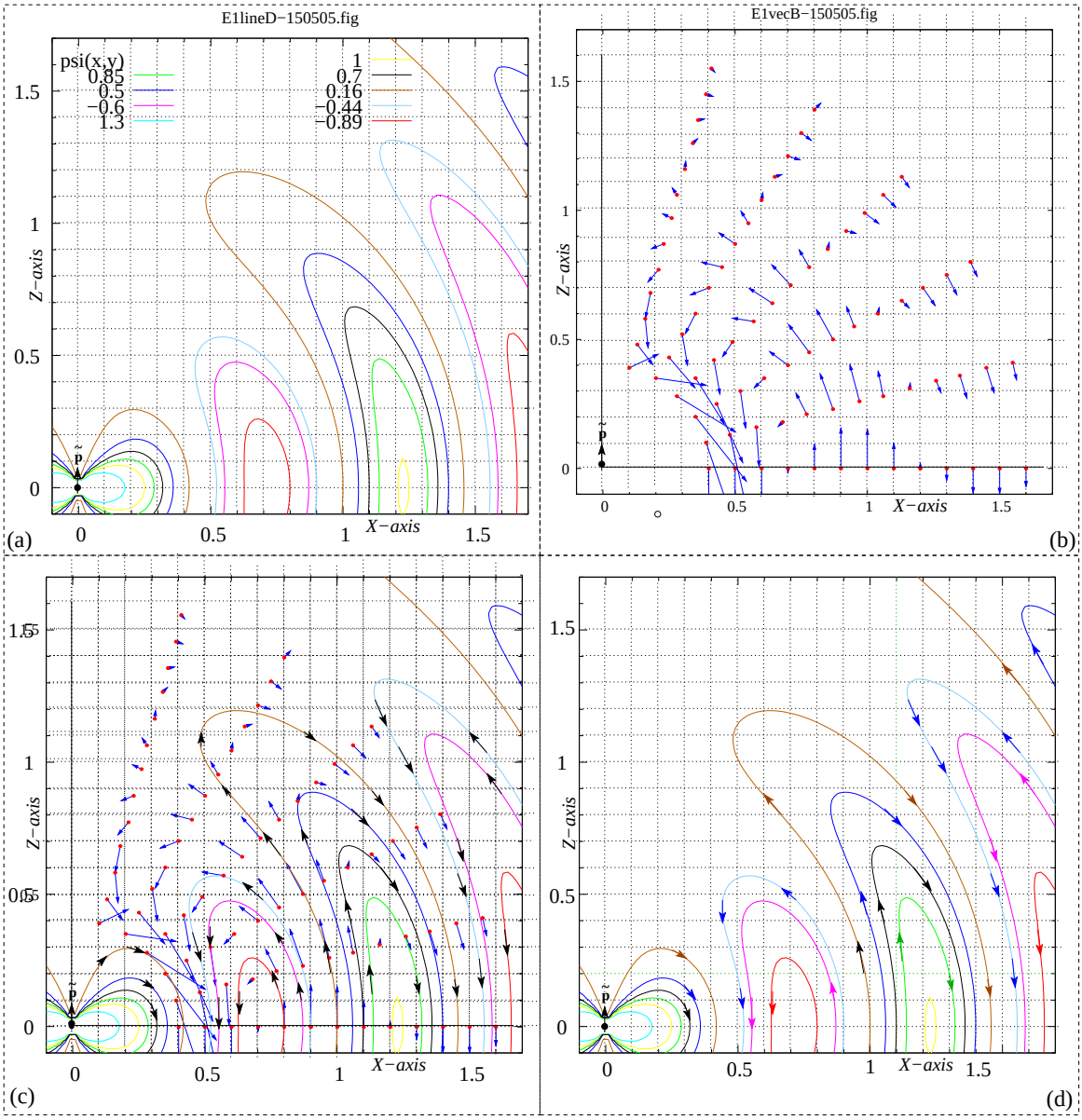


Figure 11: *Drawing directed* **E** *field lines (lines of force) on the* XZ *plane at time* $t = 0$*, on the first quadrant.*

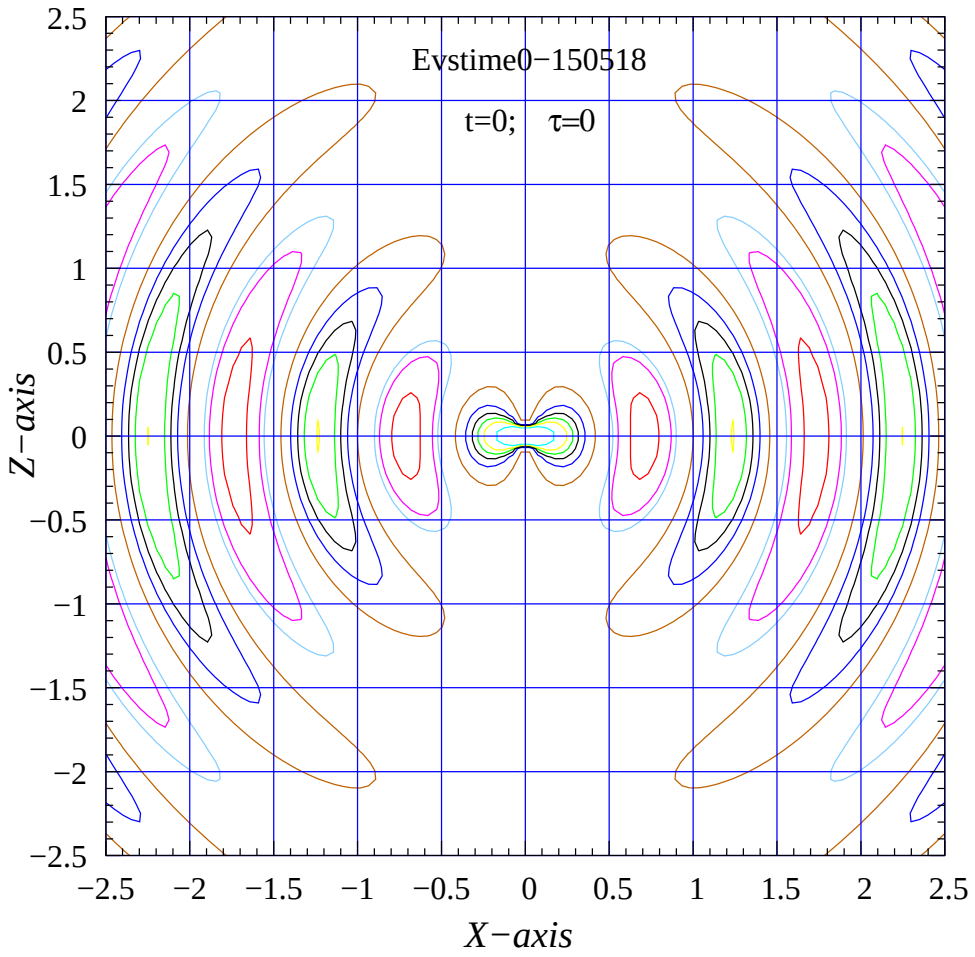


Figure 12: **E** *field lines at t=0.*

directions of the field obtained in (ii) will be extrapolated in (i) along the field lines.

Step 1: Plotting the field lines over one full period at 8 equally spaced instants: $t = nT/4$; $\tau = n\pi/4$; $n = 0, 1, ..., 7$

Ex.10

```
> set size square
> set xrange [-2.5:2.5]; set yrange [-2.5:2.5]
> set xtics 0.5; set mxtics 5
> set ytics 0.5; set mytics 5
> set grid xtics ytics back linetype 3
> unset key
> r(x,y) = sqrt(x*x+y*y)
> st(x,y)=x/r(x,y)
> sr(x,y) = sin(2*pi*r(x,y));
  cr(x,y)=cos(2*pi*r(x,y)) # n = 0
> psi(x,y)=(cr(x,y)/(2*pi*r(x,y))
  +sr(x,y))*((st(x,y))**2)
> set contour base
> unset surface
> set view map
> set cntrparam levels discrete -0.89, -0.44,
  0.16,0.70,1,1.3,-0.6,0.5,0.85
> set isosamples 100,100
> set out "Evstime0-150518.fig"
> splot psi(x,y)
> sr(x,y)=sin(2*pi*r(x,y)-pi/4);
  cr(x,y)=cos(2*pi*r(x,y)-pi/4) # n = 1
> psi(x,y) =( cr(x,y)/(2*pi*r(x,y))
  +sr(x,y))*((st(x,y))**2)
> set out "Evstime1-150518.fig"
> splot psi(x,y)
.....
> sr(x,y) = sin(2*pi*r(x,y)-7*pi/4);
  cr(x,y)=cos(2*pi*r(x,y)-7*pi/4) # n = 7
> psi(x,y)=(cr(x,y)/(2*pi*r(x,y))
  +sr(x,y))*((st(x,y))**2)
> set out "Evstime7-150518.fig"
> splot psi(x,y)
```

- The field lines corresponding to $t = 0$ are shown in Fig.12
- The command lines corresponding to n=2,...,6 are not shown.
- The field lines corresponding to $t = nT/4$; $n = 1, ..., 7$ are not shown. However, they have been saved in the working direcory under seven file names as specified above, namely, Evstime1-150518.fig, ... ,Evstime7-150518.fig. They are to be used in the Step 5 below.

Step 2: Plotting E_θ vs x along the X-axis. We shall however go further and plot E_θ, B_ϕ on the same plot only for $t = 0$.

Ex.11 To plot $E_\theta(x)$ and $cB_\phi(x)$ at $t = 0$ along the X-axis.

```
> set xtics 0.5; set mxtics 5
> set ytics 0.15; set mytics 3
> set grid xtics ytics back linetype 3
> set xrange [0.2:2.5];set yrange [-0.3:0.3]
> ro(x) = 2*pi*x; ro2(x)=(ro(x))**2;
  ro3(x) =(ro(x))**3
> sr(x) = sin(ro(x));cr(x) = cos(ro(x))
> Et(x)=cr(x)/ro3(x)+sr(x)/ro2(x)
  -cr(x)/ro(x) # n=0
> Bf(x) = sr(x)/ro2(x)-cr(x)/ro(x)
> set xrange [0.2:2.5];
  set yrange [-0.3:0.3]
> set title "EBvsx0-150621.fig"
> set out "EBvsx0-150621.fig"
> plot [0.2:2.5] [-0.3:0.3] Et(x),Bf(x)
```

We have shown this combined plot in Fig. 13. Note from Eq. (15) that $\mathbf{E}$ has only θ component on the XY plane, and $c\mathbf{B}$ has only ϕ component everywhere. Therefore along the X axis, $E = E_\theta, B = B_\phi$. We have shown in Fig. 13 E and cB vs x on the same graph at $t = 0$. It is seen that the two fields are almost equal for $x > 0.6$.

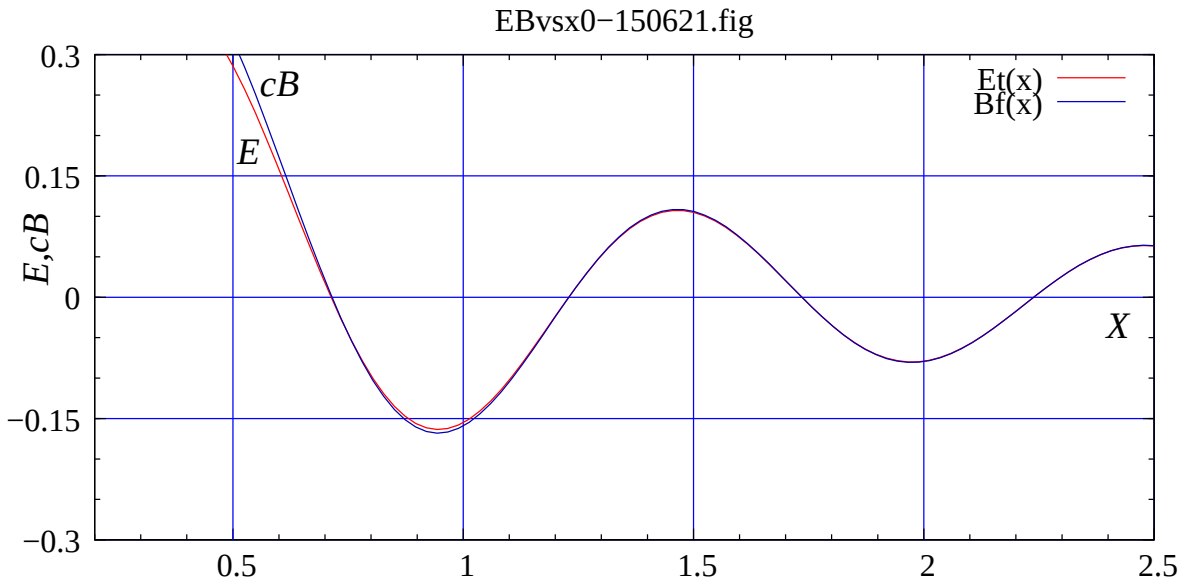


Figure 13: *E and B vs x along the X-axis, at* $t = 0$.

Ex.12 To plot $E_\theta(x)$ along the X-axis, for $t = nT/8;\ n = 1, ...7$.

```
> sr(x)=sin(ro(x)-pi/4);
  cr(x)=cos(ro(x)-pi/4) # n=1
> Et(x)=cr(x)/ro3(x)+sr(x)/ro2(x)-cr(x)/ro(x)
> set title "Evsxn1-150521.fig"
> set out "Evsxn1-150521.fig"
> plot Et(x)
.....................................
> sr(x) = sin(ro(x)-7*pi/4);
  cr(x)=cos(ro(x)-7*pi/4) # n=7
> Et(x)=cr(x)/ro3(x)+sr(x)/ro2(x)-cr(x)/ro(x)
> set title "Evsxn7-150521.fig"
> set out "Evsxn7-150521.fig"
> plot Et(x)
```

- The commands are a continuation of those in Ex.11.
- The command lines corresponding to n=2,...,6 are not shown.
- The seven graphs plotted above are not shown. However, they have been saved in the working direcory under seven file names as specified above, namely, Evsxn1-150521.fig, ... , Evsxn7-150521.fig. These plots are to be used in the Step 5 below.

Step 3: Insert arrows alongside **E** *field lines by finding their directions from the E vs x plots, at* $t = 0$.

We have illustrated in Fig. 14 how this operation has been carried out. We copied two figures, namely (a) Fig. 12, and (b) Fig. 13, on a single canvas in xfig. We removed the plot of B_ϕ, so that (b) has only plot of E_θ. Now we placed (b) beneath (a), scaled down (b) such that its boundary lines (corresponding to $x = 0.2, 2.5$) were aligned with the corresponding tic marks in (a). This was done

because the graph (a) has xrange [-2.5:2.5] and the graph (b) has xrange [0.2:2.5]. Next, we drew vertical lines from the following points on the graph in (b) to the X-axis of (a): (i) the maxima, (ii) the minima, and the (iii) zeros.

The unit vector $\mathbf{e}_\theta$ coincides with $-\mathbf{k}$ on the XY plane, as mentioned earlier. Therefore, the maxima correspond to peak values of $\mathbf{E}$ in the *negative* Z direction, the minima to the peak values *positive* Z direction, and the zeros to the centres of the loops formed by the contours. We followed the directions obtained from (i) and (ii) through entire loops of the contours.

Step 4: Obtaining directed field lines of $\mathbf{E}$ *and* $\mathbf{B}$*, at* $t = 0$

Now that we have drawn directed field lines of $\mathbf{E}$ at $t = 0$, we need to see it side by side with the $\mathbf{B}$ field at the same instant of time. For this purpose we have copied directed field lines of Fig. 14 as part (a) of Fig. 15, and drawn the field lines of $\mathbf{B}$ as concentric circles in part (b) of the same figure, as suggested by Fig. 1(c), and Eq. (15c).

It is seen from Fig. 13 that the positive and negative peak values of $\mathbf{E}$ and $c\mathbf{B}$ occur at the same set of points on the X axis. Using this as a guide, we have obtained the directions of $\mathbf{B}$ along the field lines, and have indicated them with arrowheads.

Motion of Field lines

The field lines as drawn in Fig. 15 look static. Actually they are moving lines, expanding outward into space with the speed of light. This dynamic character of the field lines is clearly seen especially in the radiation zone, from Eqs. (5). In this zone we can assign a phase, defined as $\varphi = (kr - \omega t)$, to every field line. Any particular phase φ associated with, say the crest (positive maximum) or trough (negative maximum), is given as $\varphi = a =$ constant. Hence, as t changes, the value of r associated with that phase changes accordingly, satisfying $r = a + \frac{\omega}{k}t = a + ct$.

For a better and analytical understanding of the "motion of the field lines" we need to go back to Eq. (22), in an attempt to giving a meaning to the term. Field lines are contours on the surface $z = \psi(x, y, t)$. Imagine two field lines $\Gamma(t)$ and its time evolution $\Gamma(t + dt)$ in a small time-interval dt. They correspond to the same value of the constant C, and are drawn at times t and $t + dt$ respectively. An imaginary point $\mathrm{P}(t)$ on $\Gamma(t)$ at the coordinates (r, θ, ϕ) moves radially to the point $\mathrm{P}(t + dt)$ on $\Gamma(t + dt)$ at the coordinates $(r + dr, \theta, \phi)$. We may then refer to $\dot{r} = \frac{dr}{dt}$ as the radial velocity of the field line at P. We can then obtain $\dot{r}$ by differentiating r with respect to t in the implicit equation Eq. (22), and get

$$\left[\cos(kr - \omega t) - \frac{\sin(kr-\omega t)}{kr} - \frac{\cos(kr-\omega t)}{(kr)^2}\right]\dot{r} = \left[\cos(kr - \omega t) - \frac{\sin(kr-\omega t)}{kr}\right]c. \tag{27}$$

A few wavelenghts away from the source the third term within the square brackets on the left side vanishes, making $\dot{r} \approx c$.

At this point let us pause for a while, and introspect whether these field lines are in conformity with Maxwell's equations in *free*

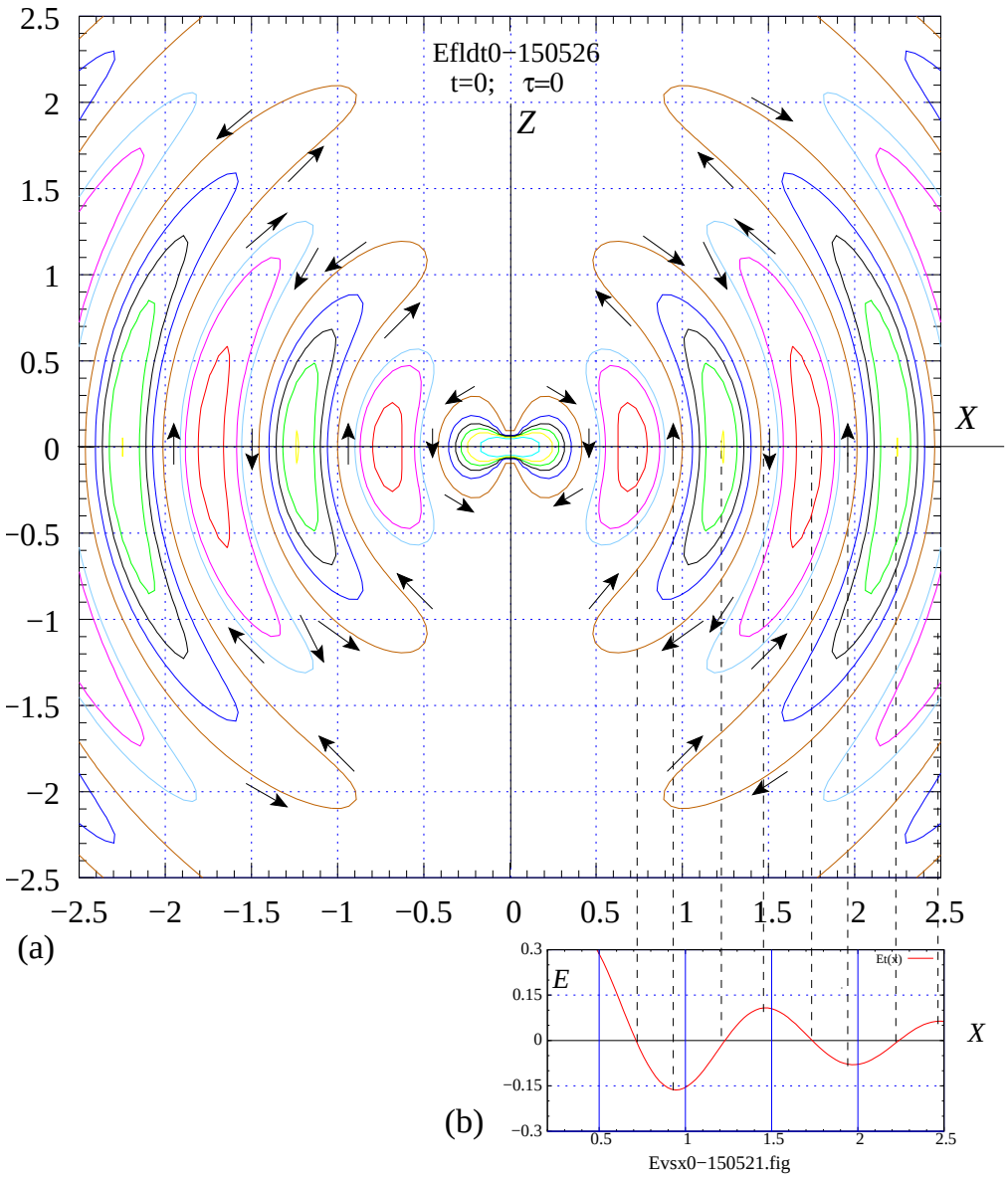


Figure 14: *Extrapolating the direction of the* **E** *field on the* XZ *plane from the* $E - x$ *graph plotted along the* x*-axis. Time* $t = 0$.

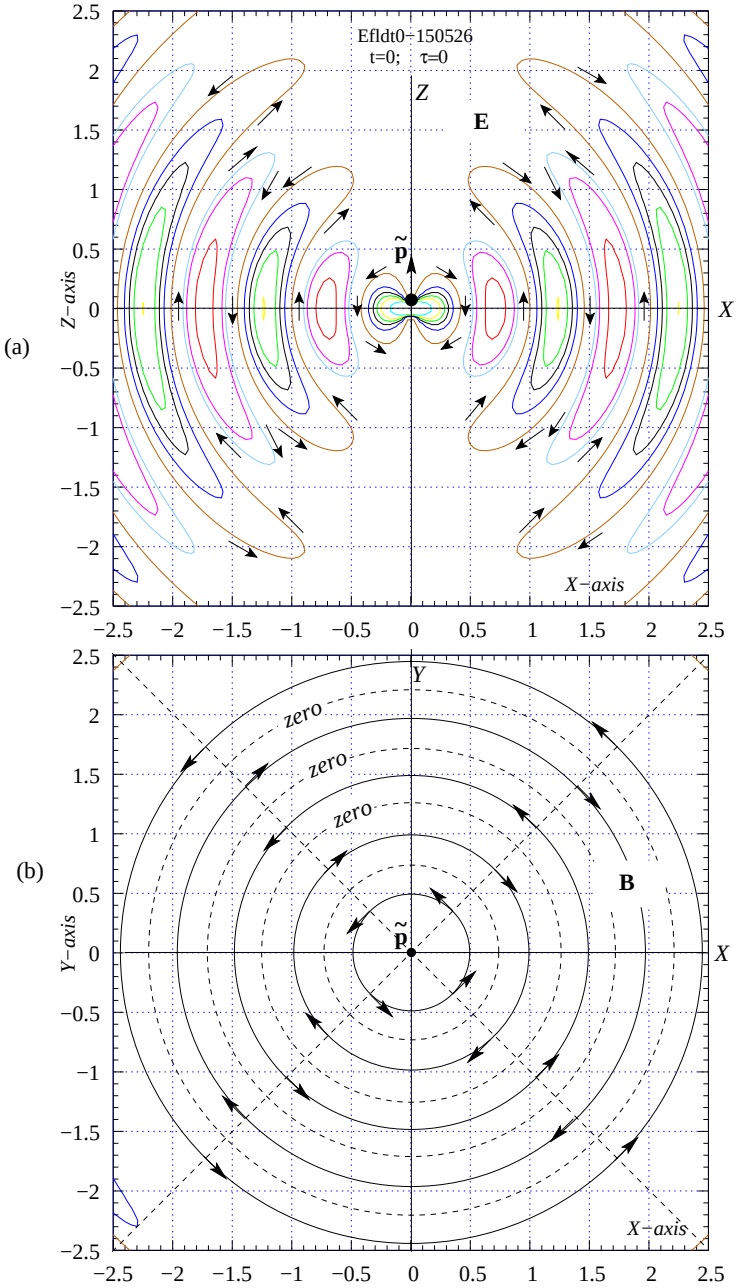


Figure 15: *The* **E** *field on the XZ (Fig.(a)), and the* **B** *field on the XY plane (Fig (b)), corresponding to the same instant of time:* $t = 0$. *The* **B** *field lines are concentric circles. The solid circles correspond to the maximum values of* **B** *in anticlockwise* (B_ϕ *positive) and clockwise* (B_ϕ *negative) directions, the dashed circles to zero value of the field.*

space, written below.

$$\boldsymbol{\nabla}\cdot\mathbf{E}=0. \quad (a) \quad \boldsymbol{\nabla}\times\mathbf{E}=-\frac{\partial c\mathbf{B}}{\partial(ct)}. \quad (b)$$
$$\boldsymbol{\nabla}\cdot c\mathbf{B}=0. \quad (c) \quad \boldsymbol{\nabla}\times c\mathbf{B}=\frac{\partial\mathbf{E}}{\partial(ct)}. \quad (d) \tag{28}$$

The form of the equations is a reminder of the relativistic form (i.e., covariant form) of Maxwell's equations. Eqs. (a),(c) suggest that the field lines of both **E** and **B** should form closed loops, which is obviously the case for both of them. Eq. (c) (representing Fraday's law) requires that the line integral of E around any such closed loop Γ must equal the negative of the area integral of $\frac{\partial cB}{\partial(ct)}$ over the area enclosed by Γ, remembering that the $c\mathbf{B}$ field, lying on the XY plane, is penetrating the XZ plane perpendicularly. This fact cannot be verified by just by looking at the plot.

Step 5: Draw directed field lines of **E** *for one full period T*

We have replicated the operation mentioned in Step 3 (**E** field at $t = 0$) for the remaining seven values of t, covering one full perod, namely, $t = nT/8;\ n = 1, ...7$. For each one of these instants, we have taken the field lines from Step 1 (Ex.10), placed below it the corresponding plot of E_θ vs x plot from Step 2 (Ex.12). After completing this work leading to the composite figure for each value of t, we scaled down each one of them to about half of their size (i.e., linear dimension) so that the figures corresponding to $t = 0, T/8, 2T/8, 3T/8$ were arranged in Fig. 16 in four quadrants, and the figures corresponding to $t = 4T/8, 5T/8, 6T/8, 7T/8$ were arranged in Fig. 17 in four quadrants.

Looking at successive pictures, corresponding to successive values of t covering on full period, the reader should get a reasonably good idea of how the field is evolving in time, resulting in a propagating wave in all directions along the XZ plane. The same picture holds for all planes passing through the Z axis.

Step 6. Finally, we have plotted E *vx* x, along the X axis, for one full period at eight equal time intervals on a single graph, as shown in Fig. 18, using the following commands.

Ex.13

```
> set xtics 0.5; set mxtics 5
> set ytics 0.05
> set grid xtics ytics back linetype 3
> ro(x) = 2*pi*x;ro2(x)=(ro(x))**2
> ro3(x) =(ro(x))**3
> sr(x) = sin(ro(x)-n*pi/4);
  cr(x)=cos(ro(x)-n*pi/4)
> Er0(x)=(cr(x)/ro3(x) + sr(x)/ro2(x))
> Et0(x) = Er0(x) - cr(x)/ro(x)
> set title "EvsxD-150520.fig"
> set out "EvsxD-150520.fig"
> do for [n=0:7]
  {plot [0.2:2.5] [-0.3:0.3] Et0(x) lt n}
```

The plots give a clear picture of the EM wave along the X axis. It should be remembered that as we go far away from the orign, the amplitudes of the **E** and $c\mathbf{B}$ fields fall off as $1/r$ (compared to $1/r^2$ for the Coulomb field), and they become equal in magnitude, but remain perpendiculr to each other. The pattern remains the same in all directions around the Z axis, except for the fact that their amplitude varies as $\sin\theta$, being zero along the Z axis, and maximum along the XY plane.

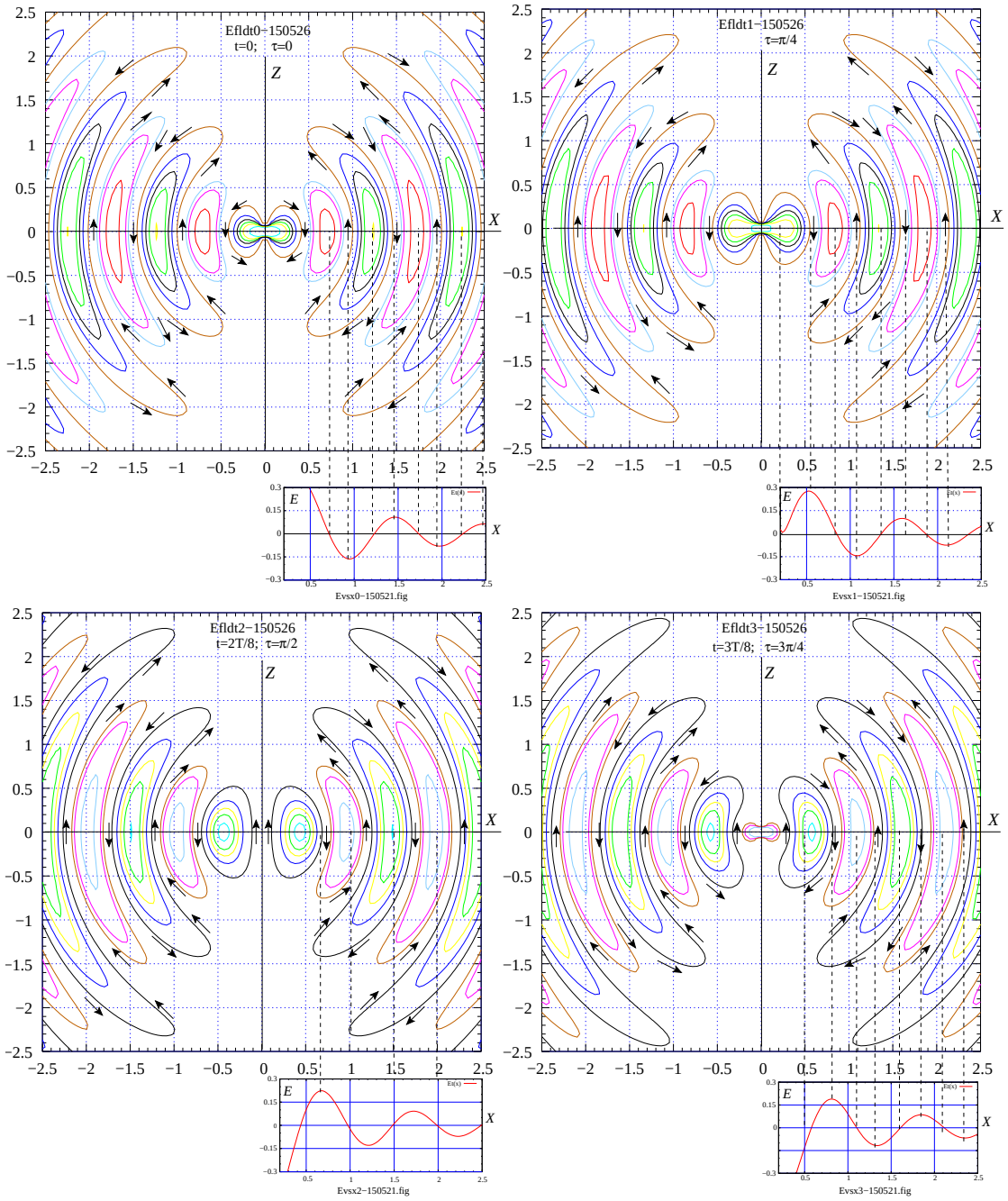


Figure 16: *Extrapolating the direction of the* **E** *field on the* XZ *plane from the* $E-x$ *graph potted along the* x*-axis. Times* $t = 0, T/8, 2T/8, 3T/8$.

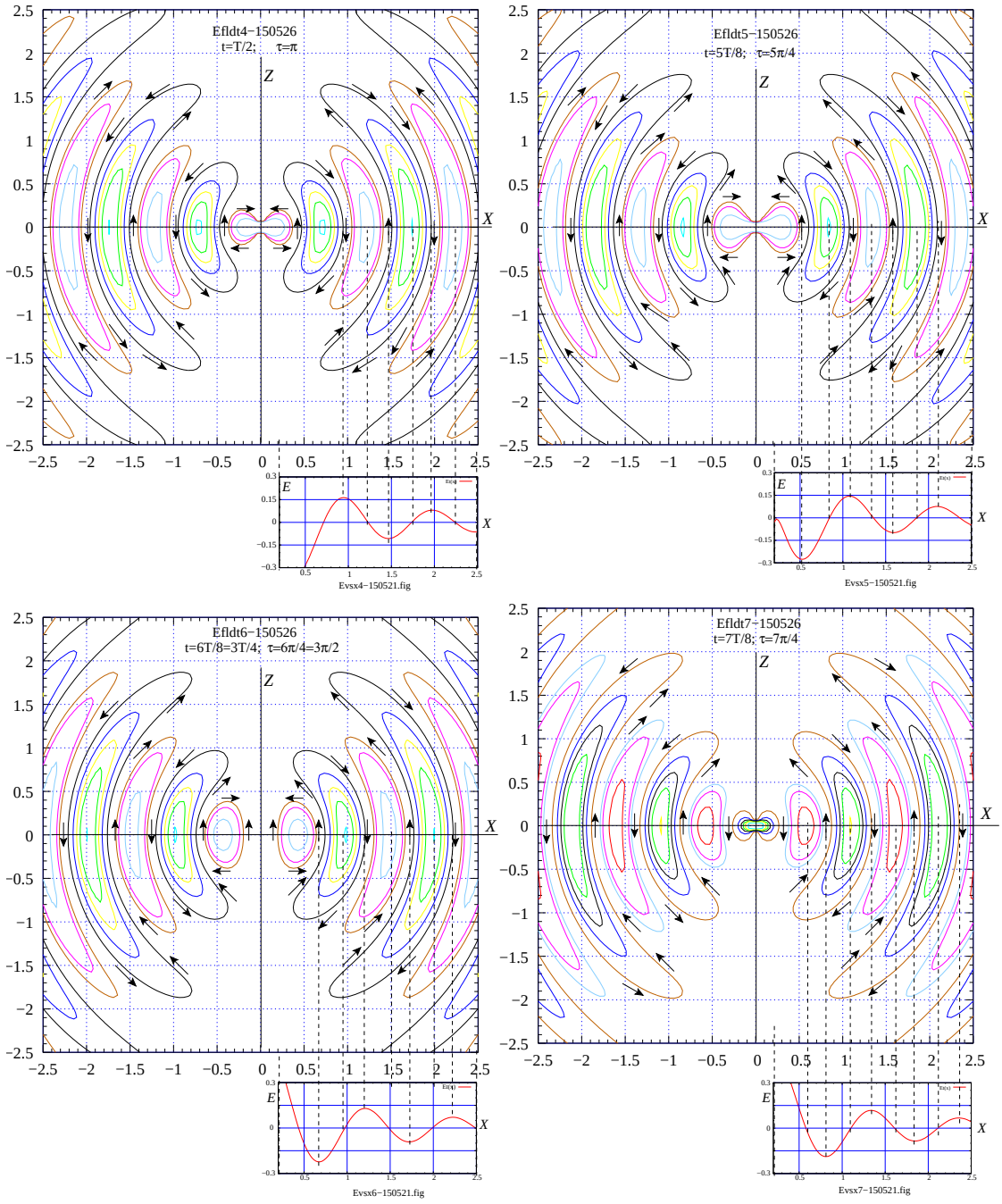


Figure 17: *Extrapolating the direction of the* **E** *field on the* XZ *plane from the* $E - x$ *graph plotted along the* x*-axis. Times* $t = 4T/8, 5T/8, 6T/8, 7T/8$.

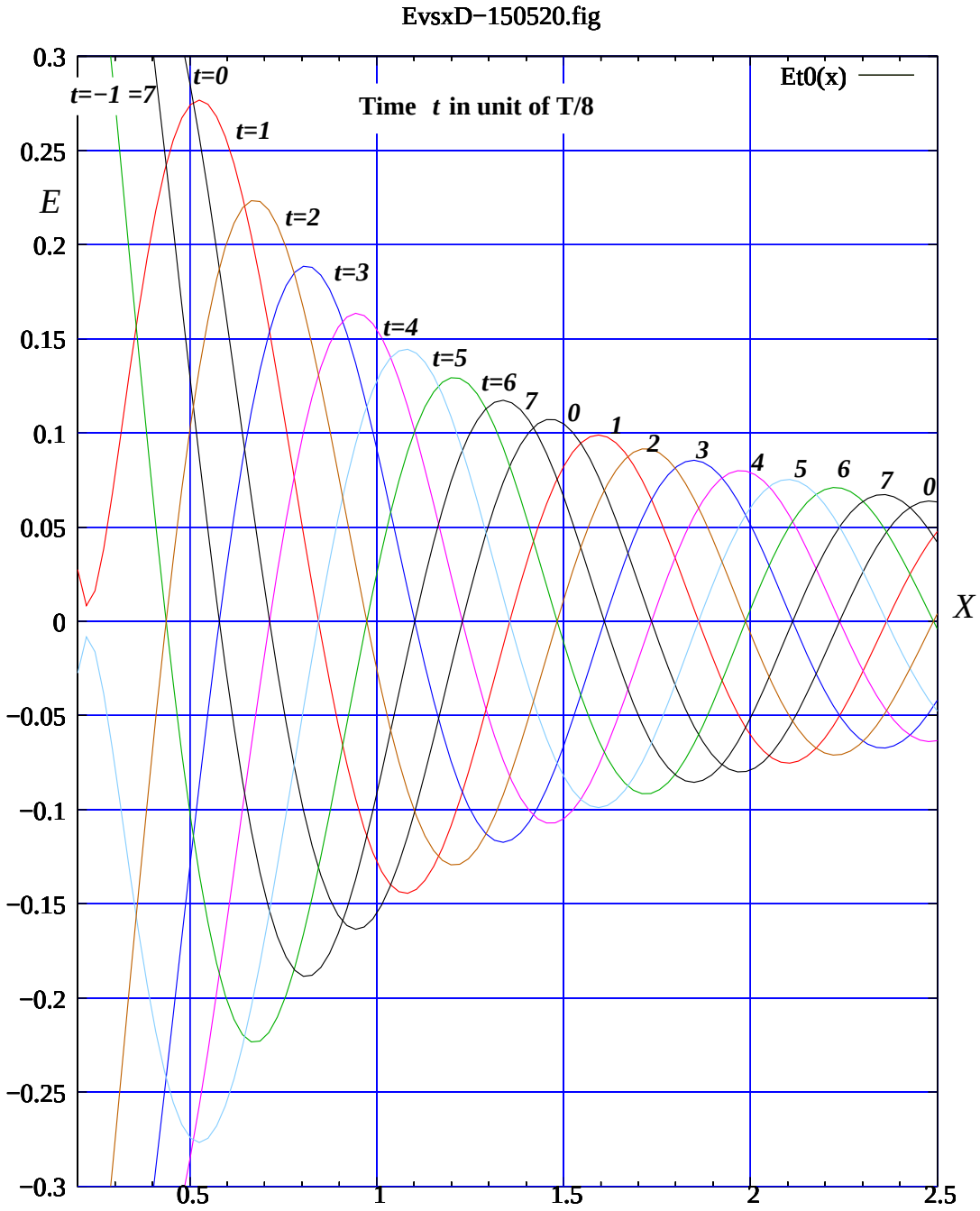


Figure 18: *E vs x at $t = 0, T/8, 2T/8, 3T/8, 4T/8, 5T/8, 6T/8, 7T/8$, covering one full period T, giving an indication of how the EM wave propagates in all directions, its amplitude falling off as $1/r$ as we go far away from the source.*

10 Plotting a linearly polarized plane EM Wave

We now come to the part (i) of our work, as mentioned at the end of the Introduction, namely the simplest example of a plane EM wave, originating from a source which is far away from the region where it is detected.

For the simplest example we shall take the direction of propagation to be the Z direction, and the direction of the **E** and **B** fields to be in the X and Y directions respectively. Let the wave be a harmonically varying field. Then[**?**]

$$\begin{aligned} \mathbf{E} &= E_0 \cos k(z - ct)\mathbf{e}_x \quad (a) \\ c\mathbf{B} &= E_0 \cos k(z - ct)\mathbf{e}_y \quad (b) \end{aligned} \tag{29}$$

The direction of the **E** vector is called the direction of *polarization*. This can be any direction perpendicular to the direction of propagation. Since in the present example, the propagation direction is the Z direction, the **E** and **B** fields must be on the XY plane.

The example shown in Eq.(29) assumes a *constant direction of polarization*. When this is the situation, the EM wave is said to be *linearly polarized*. There is another special case in which the magnitude of the **E** field is constant, but rotates uniformly, perpendicular to the direction of propagation. Such a propagating field is said to be *circularly polarized*. The present example illustrates a linearly polarized wave for which the diection of polarization is the X direction. We may even call it an X-polarized plane EM wave. We have depicted this wave in Fig. 19(c).

Note that the direction of the Poynting's vector **S** is the same as the Z direction, and the the waveform shown in the figure is moving in bulk with the speed c in the Z direction. Also we have shown the $c\mathbf{B}$ field, as the companion of the **E** field, so as to convey to the reader the equality $E = cB$ which is intended to be portrayed by equal lengths of the two vectors **E** and $c\mathbf{B}$.

To create the picture of the propagating field we first created two primitive plots, shown in Figs. 19(a) and (b), each having the same cosine function plotted on the XZ plane and the YZ plane, and covering time ranges $[-\pi/2 : \pi/2], [\pi/2 : 3\pi/2]$ respectively. Together they covered one full period of the wave. The final picture of the propagating wave shown in Fig. 19(c) was created by the editing operation : copying-pasting, joining, adding arrows, filling with colors, and then extending further over two more periods. The plots in Figs (a) and (b) were created using Gnuplot through the following command.

Ex.14

```
> set  parametric
> set urange [-pi/2:pi/2]
> splot u,0,cos(u),u,-cos(u),0,u,0,0
> set term fig color size 27 18 metric
  pointsmax 1000 solid font "Times-Roman,12"
  depth 50
> set title "EMplaneA-150504.fig"
> set output "EMplaneA-150504.fig"
> set key bmargin
> replot
> set urange [pi/2:3*pi/2]
> set title "EMplaneB-150504.fig"
> set key bmargin
```

```
> splot u, 0, cos( u ), u, -cos( u ), 0, u, 0,
> set output "EMplaneB-150504.fig"
```

Appendix

The content of the file E1vecdataC-150423.txt
The data below lists x,y,Ex,Ey for d= 1, .8, .6, .4, .5, .7, .9, 1.1, 1.3, 1.5, 1.4, 1.6, 1.3

```
# x y  Ex  Ey
.26 .97 -.037 .018
.5 .87 -.064 .045
.71 .71 -.074 .082
.87 .5 -.064 .118
.97 .26 -.037 .145
1.00 0 0 .155 #   d=1 on x axis
.21 .77 -.042 -.05
.4 .70 -.072 -.029
.57 .57 -.084 .013
.70 .4 -.072 .055
.77 .21 -.042 .085
.8 0 0 .097 # d=.8 on x axis
.16 .58 .011 -.116
.3 .52 .020 -.124
.42 .42 .023 -.136
.52 .3 .020 -.147
.58 .16 .011 -.155
.6 0 0 -.158 #  d=.6 on x axis
.10 .39 .112 .054
.2 .35 .194 -.028
.28 .28 .224 -.140
.35 .2 .194 -.252
.39 .10 .112 -.334
.4 0 0 -.364 #  d=.4 on x axis
.13 .48 .06 -0.08
.25 .43 .1 -.12
.35 .35 .11 -.18
.43 .25 .1 -.23
.48 .13 .06 -.27
.5 0 0 -.29 # d=.5 on x axis
.18 .68 -.02 -.1
.35 .6 -.04 -.08
.49 .49 -.04 -.06
.61 .35 -.04 -.04
.68 .18 -.02 -.02
.7 0 0 -.02 # d=.7 on x axis
.23 .87 -.05 -.02
.45 .78 -.08 .02
.64 .64 -.09 .06
.78 .45 -.08 .11
.87 .23 -.05 .14
.9 0 0 .16 # d=.9 on x axis
.28 1.06 -.02 .03
.55 .95 -.03 .05
.78 .78 -.04 .07
.95 .55 -.03 .08
1.06 .28 -.02 .1
1.1 0 0 .1 # d=1.1 on x axis
.34 1.26 .02 .02
.65 1.13 .03 .01
.92 .92 .04 -.01
1.13 .65 .03 -.03
1.26 .34 .02 -.05
1.3 0 0 -.05 #  d=1.3 on x axis
.39 1.45 .03 -.01
.75 1.3 .04 -.03
1.06 1.06 .05 -.05
1.3 .75 .04 -.08
1.45 .39 .03 -.1
1.5 0 0 -.1 # d=1.5 on x axis
.36 1.35 .028 .005
.7 1.21 .048 -.015
.99 .99 .056 -.043
1.21 .7 .048 -.071
1.35 .36 .028 -.091
1.4 0 0 -.098 # d=1.4 on x axis
.41 1.55 .015 -.017
.8 1.39 .026 .028
1.13 1.13 .03 -.044
1.39 .8 .026 -.059
1.55 .41 .015 -.07
1.6 0 0 -.074 # d=1.6 on x axis
.31 1.16 .003 .034
.6 1.04 .005 .032
.85 .85 .006 .029
1.04 .6 .005 .026
1.16 .31 .003 .024
```

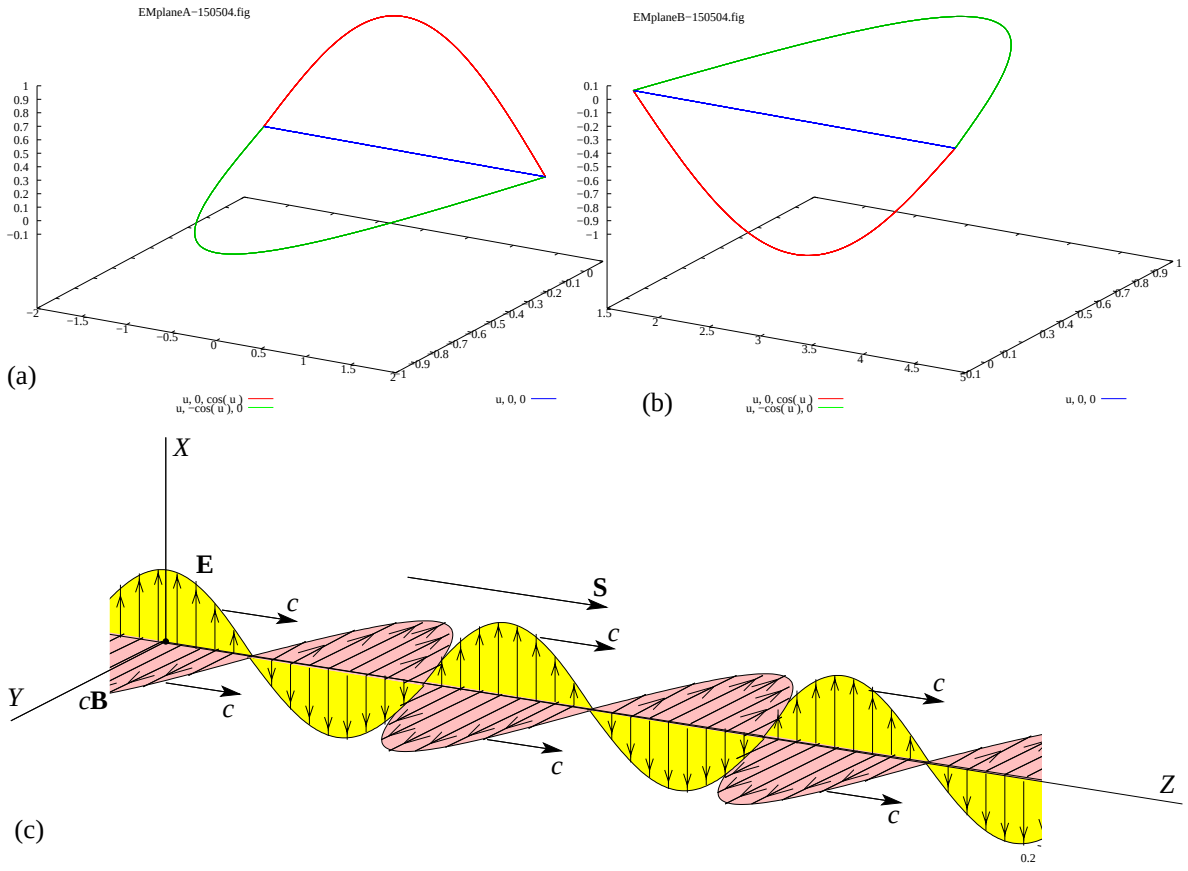


Figure 19: *A linearly polarized plane electromagnetic field propagating in the Z direction*

```
1.2 0 0 .024 # d=1.2 on x axis
```

Acknowledgement

The author received the basic commands shown in Ex.14 (Plane EM Wave) from his former student Bharat M, and is grateful to Prof. Sophocles J.Orfanidis (Rutgers University) for some helpful suggestions.

References

[1] Somnath Datta, *Minkowski's Space Time*, Physics Education, www.physedu.in, Indian Association of Physics Teachers , Vol 29, No. 2 (Apr-Jun 2013). Article No. 2, 50 pages.

[2] Somnath Datta, *Maxwell's Stress Tensor and Conservation of Momentum in Electromagnetic Field*, Physics Education, www.physedu.in, Indian Association of Physics Teachers , Vol 30, No. 3 (Jul-Sep 2014). Article No. 1, 42 pages.

[3] Somnath Datta, *Solution of Maxwell's Equations Illustrated with Examples and Plots*, under preparation.

[4] Thomas Williams & Colin Kelley, *Gnuplot 4.6*, Manual originally prepared by Dick Crawford, 2012, Version 4.6, http://sourceforge.net/projects/gnuplot (2012)

[5] Somnath Datta, *Mechanics*, Pearson Education, New Delhi (2013). 630 pages

[6] Lee Phillips, *Gnuplot Cookbook*, Packt Publishing Ltd., Birmingham B3 2PB, UK..ISBN 978-1-84951-724-9 www.packtpub.com (2012)

[7] Sophocles J. Orfanidis, *Electromagnetic Waves and Antennas*, ECE Department, Rutgers University, NJ 08854-8058 (2014). Web page: www.ece.rutgers.edu/∼orfanidi/ewa. See pp. 721-722.

[8] Paul Lorrain & Dale Corson, *Electromagnetic Fields and Waves*, 2nd Ed, W.H.Freeman and Co., New York - CBS Publishers & Distributors, Delhi (1986). pp. 63, 595-603.

[9] Wolfgang K.H. Panofsky & Melba Phillips, *Classical Electricity and Magnetism,* 2nd Ed, Addison Wesley, Reading, Massachusetts (1964). pp. 257-259.

[10] David J. Griffith, *Introduction to Electrodynamics*, 3rd Ed, Pearson Education, New Delhi (2006). See Sections 11.1, 9.2.