Classical Electrodynamics

Part II

by

Robert G. Brown

Duke University Physics Department
Durham, NC 27708-0305
rgb@phy.duke.edu

Acknowledgements

I'd like to dedicate these notes to the memory of Larry C. Biedenharn. Larry was my Ph.D. advisor at Duke and he generously loaned me his (mostly handwritten or crudely typed) lecture notes when in the natural course of events I came to teach Electrodynamics for the first time. Most of the notes have been completely rewritten, typeset with latex, changed to emphasize the things that I think are important, but there are still important fragments that are more or less pure Biedenharn, in particular the lovely exposition of vector spherical harmonics and Hansen solutions (which a student will very likely be unable to find anywhere else).

I'd also like to acknowledge and thank my many colleagues at Duke and elsewhere who have contributed ideas, criticisms, or encouragement to me over the years, in particular Mikael Ciftan (my "other advisor" for my Ph.D. and beyond), Richard Palmer and Ronen Plesser.

Notice

This set of "lecture notes" is designed to support my personal teaching activities at Duke University, in particular teaching its Physics 318/319 series (graduate level Classical Electrodynamics) using J. D. Jackson's *Classical Electrodynamics* as a primary text. However, the notes may be useful to students studying from other texts or even as a standalone text in its own right.

It is freely available in its entirety online at

http://www.phy.duke.edu/∼rgb/Class/phy319.php

as well as through Lulu's "book previewer" at

http://www.lulu.com/content/1144184

(where one can also purchase an inexpensive clean download of the book PDF in Crown Quarto size – 7.444×9.681 inch pages – that can be read using any PDF browser or locally printed).

In this way the text can be used by students all over the world, where each student can pay (or not) according to their means. Nevertheless, I am hoping that students who truly find this work useful will purchase either the PDF download or the current paper snapshot, if only to help subsidize me while I continue to write more inexpensive textbooks in physics or other subjects.

These are real lecture notes, and they therefore have errors great and small, missing figures (that I usually draw from memory in class), and they cover and omit topics according to my own view of what is or isn't important to cover in a one-semester course. Expect them to change without warning as I add content or correct errors. Purchasers of a paper version should be aware of its imperfection and be prepared to either live with it or mark up their *own* copies with corrections or additions as need be in the lecture note spirit, as I do mine. The text has generous margins, is widely spaced, and contains scattered blank pages for students' or instructors' own use to facilitate this.

I cherish good-hearted communication from students or other instructors pointing out errors or suggesting new content (and have in the past done my best to implement many such corrections or suggestions).

Contents

Preface

Classical Electrodynamics is one of the most beautiful things in the world. Four simple vector equations (or one tensor equation and an asssociated dual) describe the unified electromagnetic field and more or less directly imply the theory of relativity. The discovery and proof that light is an electromagnetic wave stands to this day as one of the greatest moments in the history of science.

These four equations even contain within them the seeds of their own destruction as a classical theory. Once Maxwell's equations were known, the *inconsistency* of the classical physics one could then easily derive from them with countless experimental results associated with electromagnetism forced the classicists of the day, many of them metaphorically kicking or screaming, to invent *quantum* mechanics and *quantum* electrodynamics to explain them. Indeed, once the single fact that an accelerated charged particle necessarily radiates electromagnetic energy was known, it became virtually impossible to conceptually explain the persistence of structure at the microscopic level (since the forces associated with binding objects together out of discrete charged parts inevitably produce an oscillation of charge due to small perturbations of position, with an associated acceleration)., The few hypotheses that were advanced to account for it "without" an overtly oscillatory model were rapidly and decisively shot down by (now famous) experiments by Rutherford, Millikan, and others.

Even though the Universe proves to be quantum mechanical at the microscopic level, classical electrodynamics is nevertheless extremely relevant and useful in the real world today at the *macroscopic* level. It describes extremely precisely nearly all the mundane aspects of ordinary electrical engineering and electromagnetic radiation from the static limit through optical frequencies. Even at the molecular level or photonic level where it breaks down and a quantum theory must be used it is *first* necessary to *understand* the classical theory before exploring the quantum theory, as the quantum theory is built on top of the entire relativistic electrodynamic conceptual framework already established.

This set of lecture notes is designed to be used to teach graduate students (and possibly advanced and motivated undergraduates) classical electrodynamics. In particular, it supports the second (more difficult) semester of a two semester course in electrodynamics that covers pretty much “all” of the theory itself (omitting, of course, many topics or specific areas where it can be applied) out to the points where the theory itself breaks down as noted above. At that point, to make further progress a student needs to learn about more fields, quantum (field) theory, advanced (general) relativity – topics generally beyond the scope of these notes.

The requirements for this course include a thorough understanding of electricity and magnetism at the level of at least one, ideally two, undergraduate courses. At Duke, for example, physics majors are first exposed first to an introductory course that covers the integral formulation of Maxwell's equations and light that uses no multivariate differential calculus, then a second course that develops the vector differential formulation of Maxwell's equations and their consequences) as does this course) but with considerably less mathematical rigor and completeness of the treatment as students taking it have likely still not had a course in e.g. contour integration. Students using these notes will find it useful to be at least somewhat comfortable with vector differential and integral calculus, to have had exposure to the theory and solution methodology of ordinary and partial differential equations, to be familiar with the mathematics of complex variables and analytic functions, contour integration, and it would be simply lovely if they at least knew what a “tensor” was.

However, even more so than is the case for most physics texts, this book will endeavor to provide internal support for students that are weak in one or more of these required mathematical skills. This support will come in one of several forms. At the very least, considerable effort has been made to hunt down on behalf of the student and explicitly recommend useful textbooks and online resources on various mathematical and physical topics that may be of use to them. Many of these resources are freely available on the web. Some mathematical methods are completely developed in the context of the discussion, either because it makes sense to do so or because there simply *are* no references a student is likely to be able to find. Finally, selected topics will be covered in e.g. appendices or as insertions in the text where they are short enough to be coverable in this way and important enough that students are likely to be highly confused without this sort of support.

A very brief review of the *electrodynamics* topics covered includes: plane waves, dispersion, penetration of waves at a boundary (skin depth), wave guides and cavities and the various (TE, TM, TEM) modes associated with them, radiation in the more general case beginning with sources. This latter exhibition goes considerably beyond Jackson, treating multipolar radiation in detail. It includes a fairly thorough exposition of the underlying PDEs, the properties of the Green's functions used to generate multipoles both approximate and exact, and formally precise solutions that extend *inside* the source charge-current density (as indeed they must for this formalism to be of use in e.g. self-consistent field theories treating extended charge density distributions). In addition to the vector spherical harmonics, it defines and derives the properties of the Hansen multipoles (which are otherwise very nearly a lost art) demonstrating their practical utility with example problems involving antennae. It concludes this part of the exposition with a short description of optical scattering as waves interact with "media", e.g. small spheres intended to model atoms or molecules.

It then procedes to develop relativity theory, first reviewing the elementary theory presumably already familiar to students, then developing the full Lorentz Group. As students tend to *not* be familiar with tensors, the notes contain a special appendix on tensors and tensor notation as a supplement. It also contains a bit of supplemental support on at least those aspects of contour integration relevant to the course for similar reasons. With relativity in hand, relativistic electrodynamics is developed, including the properties of radiation emitted from a point charge as it is accelerated. Finally, the notes conclude with a nice overview of radiation reaction (exploring the work of Lorentz, Dirac, and Wheeler and Feynman) and the puzzles therein – self-interaction versus action at a distance, the need for a *classical* renormalization in a theory based on self-interaction, and a bit more.

One *note-worthy* feature of these notes (sorry, but I do like puns and you'll just have to get used to them:-) is that the electronic/online version of them includes several inventions of my own such as a *wikinote*[1] , a reference to supporting wikipedia articles that appears as a URL and footnote in the text copy but which is an *active link* in a PDF or HTML (online) copy. Similarly, there are google links and ordinary web links presented in the same way.

As noted at the beginning of the text, these are *real* lecture notes and subject to change as they are used, semester by semester. In some cases the changes are quite important, for example when a kind reader gently points out a bone-headed mistake that makes some aspect of the presentation quite incorrect. In others they are smaller improvements: a new link, a slightly improved discussion, fixing clumsy language, a new figure (or putting in one of the missing old ones). As time passes I hope to add a selection of problems that will make the text more of a stand-alone teaching aid as well.

For both of these reasons, students who are using these notes may wish to have *both* a paper snapshot of the notes – that will *inevitably* contain omissions and mistakes or material I don't actually cover in *this* year's class – *and* a (more current) electronic copy. I generally maintain the current snapshot of the electronic copy that I'm actually using to teach from where it is available, for free to all comers, on my personal/class website at:

http://www.phy.duke.edu/∼rgb/Class/phy319.php

(which cleverly and self-consistently demonstrates an active link in action, as did the wikilink above). In this way they can have the convenience of a slightly-out-of-date paper copy to browse or study or follow and mark up during lecture as well as an electronic copy that is up to date and which contains useful active links.

Let it be noted that I'm as greedy and needy as the next human, and

[1]Wikipedia: http://www.wikipedia.org/wiki/wikipedia. A wikinote is basically a footnote that directs a student to a useful article in the Wikipedia. There is some (frankly silly) controversy on just how accurate and useful the Wikipedia is for scholarly work, but for *teaching* or *learning* science and mathematics on your own it is rapidly becoming *indispensible* as some *excellent* articles are constantly being added and improved that cover, basically, all of electrodynamics and the requisite supporting mathematics. Personally, I think the objections to it are largely economic – in a few more years this superb free resource will essentially destroy the lucrative textbook market altogether, which honestly is probably a good thing. At the very least, a textbook will have to add significant value to survive, and maybe will be a bit less expensive than the $100-a-book current standard.

can always use extra money. As I've worked quite hard on these notes (and from observation they go quite beyond what e.g. most of my colleagues make available for their own courses) I have done the work required to transform them into an actual bound book that students *can* elect to purchase all at once instead of downloading the free PDF, printing it out as two-sided pages, punching it, and inserting it into a three ring binder that anonymously joins the rest of their notes and ultimately is thrown away or lost.

This printed book is remarkably inexpensive by the standards of modern textbooks (where e.g Wyld, which I once purchased now at $16 a copy, is not available new for *$70* a copy). At the same site, students can find the actual PDF from which the book is generated available for a very low cost and are at liberty to purchase and keep that on their personal laptops or PDF-capable e-book readers, or for that matter to have *it* printed and bound by a local printer. In both cases I make a small royalty (on the order of $5) from their sale, which is both fair and helps support me so that I can write more texts such as this.

However, students around the world have very different means. Purchasing a $7.50 download in the United States means (for *most* students) that a student has to give up a few Latte Enormes from Starbucks. Purchasing that same download could be a real hardship for students from many countries around the world *including* some from the United States. For this reason students will always have the *option* of using the online notes directly from the class website for free or printing their own copy on paper at cost. All that I ask of students who elect to use them for free is that they "pay it forward" – that one day *they* help others who are less fortunate in some way for free so that we can all keep the world moving along in a positive direction.

These notes begin with *my* course syllabus and class rules and so on. Obviously if you are reading this and are not *in* my class these may be of no use to you. On the other hand, if you are a teacher planning to use these notes to guide a similar course (in which case you should probably contact me to get a free copy of the latex sources so you can modify them according to your own needs) or just a student seeking to learn how to most effectively *use* the notes and learn electrodynamics effectively, you might still find the syllabus and class rules worth at least a peek.

The one restriction I have, and I think it is entirely fair, is that instructors who elect to use these notes to help support the teaching of their own classes (either building them with or without modifications from the sources or using any of the free prebuilt images) may not *resell* these notes to their own students for a profit (recovering printing costs is OK) at least without arranging to send a fair share of that profit back to me, nor may they alter this preface, the authorship or copyright notice (basically all the front-matter) or the license. Everything from the syllabus on is fair game, though, and the notes should easily build on any e.g. linux system.

Anyway, good luck and remember that I *do* cherish feedback of all sorts, corrections, additions (especially in ready-to-build latex with EPS figures:-), suggestions, criticisms, and or course money in the form of the aforementioned small royalties.

Chapter 1

Syllabus and Course Rules

1.1 Contact Information

Instructor: Robert G. Brown
Office: Room 260
Office Phone: 660-2567   quad Cell Phone: 280-8443
Email: rgb@phy.duke.edu
Notes URL: http://www.phy.duke.edu/∼rgb/Class/phy319.php

1.2 Useful Texts and Web References

- J.D Jackson, *Classical Electrodynamics*, 3rd ed.
- Bound paper copy of these notes:
  http://www.lulu.com/content/1144184
- Orfanidi's *Electromagnetic Waves and Antennas*:
  http://www.ece.rutgers.edu/∼orfanidi/ewa/
- H. Wyld, *Methods of Mathematical Physics*, ISBN 978-0738201252, available from e.g. http://amazon.com. Other mathematical physics texts such as Arfken or Morse and Feshback are equivalently useful.
- Donald H. Menzel's *Mathematical Physics*, Dover press, ISBN 0-486-60056-4. This reference has a very nice discussion of dyads and how

to express classical mechanics in tensor form, which is actually quite lovely.

- Fabulous complex variable/contour integration reference by Mark Trodden at Syracuse:
  http://physics.syr.edu/∼trodden/courses/mathmethods/
  This online lecture note/book actually works through the Euler-Lagrange equation as well, but stops unfortunately short of doing EVERYTHING that we need in this course. It is only 70 pages, though – probably unfinished.

- Introduction to tensors by Joseph C. Kolecki at NASA:
  www.grc.nasa.gov/WWW/K-12/Numbers/Math/documents/Tensors_TM2002211716.pdf

- Short review of tensors for a Berkeley cosmology course:
  http://grus.berkeley.edu/∼jrg/ay202/node183.html

- Short review of tensors for a Winnipeg University cosmology course:
  http://io.uwinnipeg.ca/∼vincent/4500.6-001/Cosmology/Tensors.htm

- Wikipedia:
  http://www.wikipedia.org Wikipedia now contains some *excellent* articles on real graduate-level electrodynamics, relativity theory, and more. The math and science community are determined to make it a one stop shop for supporting all sorts of coursework.

- Mathworld:
  http://mathworld.wolfram.com This site, too, is very good although some of the articles tend to be either then or overly technical at the expense of clarity.

- GIYF (Google Is Your Friend). When looking for help on *any* topic, give google a try. I do.

1.3 Course Description

In this year's course we will cover the following basic topics:

- *Very* rapid review of Maxwell's equations, wave equation for EM potentials, Green's functions for the wave and Helmholtz equations, magnetic monopoles. You should all know this already, but it never hurts to go over Maxwell's equations again...

- Plane waves and wave guides. Polarization, propagating modes. (Jackson chapters 7 and 8). This year (fall 2007) Ronen tells me that he got through about the first half of chapter 7, but we'll probably review this quickly for completeness.

- Radiating systems and multipolar radiation (Jackson chapter 9). We will cover this material thoroughly. We'll do lots of really hard problems for homework and you'll all just hate it. But it'll be *soooo* good for you. The new edition of Jackson no longer covers multipoles in two places, but its treatment of vector harmonics is still quite inadequate. We will add a significant amount of material here and go beyond Jackson alone. We may do a tiny bit of material from the beginning of chapter 10 (scattering) – just enough to understand e.g. blue skies and polarization, and perhaps to learn of the existence of e.g. critical opalescence. We will not cover diffraction, apertures, etc. as those are more appropriate to a course in optics.

- Relativity (Jackson chapters 11 and 12). We will do a fairly complete job of at lease special relativity that will hopefully complement the treatments some of you have had or are having in other courses, but those of you who have lived in a Euclidean world all your lives need not be afraid. Yes, I'll continue to beat you to death with problems. It's so easy. Five or six should take you *days*.

- Radiation by moving charges (Jacksom chapters 14 and 16). Basically, this uses the Green's functions deduced during our discussion of relativity to show that accelerated charges radiate, and that as they do so a somewhat mysterious "self-force" is exerted that damps the motion of the particle. This is important, because the (experimental) observation that bound charges (which SHOULD be accelerating) don't radiate leads to the collapse of classical physics and the logical necessity of quantum physics.

- Miscellaneous (Jackson chapters 10, 13, 15). As noted above, we may look a bit at sections here and there in this, but frankly we won't have time to complete the agenda above as it is without working very hard. Stuff in these chapters you'll likely have to learn on your own as you need it.

1.4 Basis of Grade

1.4.1 Percentages

There will be, as you may have guessed, lots of problems. Actually, there will only be a few problems, but they'll seem like a lot. The structure and organization of the course will be (approximately!):

50% of grade Homework.

20% Take-Home Midterm Exam.

20% Take-Home Final Exam.

10% Research/Computing project.

In more detail, Homework is Homework, the Exams are Homework (fancied up a bit) and the Research Project is described below. These figures are only approximate. I may make homework worth a little more or less, but this is about right. Actual grades will be assigned based on performance and experience, curved (if you will) not just across your class but previous graduate classes I have taught the same material to as well. It will be very easy to pass cleanly (B- or better, in a graduate class) *if* you've done *all* the homework and perhaps less easy to get an A.

Note also that grades aside, there is a fundamental need to pass qualifiers. Qualifiers are much easier than the problems we cover in this class, but to comfortably pass it is essential that you learn the *physics* associated with all of the problems and methodologies. Do not miss learning the large scale, intuitive ideas in physics covered (the 'forest') in favor of mastering all sorts of methods for PDEs or transformations for specific problems (the 'trees'). I will do my best to help convey these in lecture, but you should read on your own, ask questions, and so on.

1.4.2 Research Project:

I'm offering you instead the following assignment, with several choices. You may prepare any of:

1. A set of lecture notes on a topic, relevant to the material we will cover, that interests you. If you select this option you may be asked to present the lecture(s), time permitting. This is an especially good option for people who have had courses that have significant overlap with something we will cover, but requires early action!

2. A review paper on a topic, relevant to the material we will cover, that interests you. Typically, in the past, students going into (e.g.) FEL have prepared review papers on the electromechanism of the FEL. That is, relevance to your future research is indicated but not mandated.

3. A computer demonstration or simulation of some important electrodynamical principle or system. Possible projects here include solving the Poisson and inhomogeneous Helmholtz equation numerically, evaluating and plotting radiation patterns and cross-sections for complicated but interesting time dependent charge density distributions, etc. Resources here include Mathematica, maple, SuperMongo, the Gnu Scientific Library, matlab/octave, and more. Obviously now is not the time to learn to program; presumably you are all competent in f77 or C or java or perl or SOMETHING if you select this option, or are willing to work very hard to becomes so. I can provide limited guidance in many (most) of these languages or environments, but will not have time to teach you to code from scratch in this class.

If you choose to do a project, it is due **TWO WEEKS** before the last class[1] so don't blow them off until the end. It is *strongly recommended* that

[1]This is a transparent ploy to make you hand it in on time. But I mean it! Really!

you clear the topic with me beforehand, as a weak topic will get a weak grade even if the presentation itself is adequate.

I will grade you on: doing a decent job (good algebra), picking an interesting topic (somewhat subjective, but I can't help it and that's why I want to talk to you about it ahead of time), adequate preparation (enough algebra), adequate documentation (where did you find the algebra), organization, and Visual Aids (pictures are sometimes worth a thousand equations). Those of you who do numerical calculations (applying the algebra) must also write it up and (ideally) submit some nifty graphics, if possible.

I'm not going to grade you particularly brutally on this — it is supposed to be fun as well as educational. However, if you do a miserable job on the project, it doesn't count. If you do a decent job (evidence of more than 20 hours of work) you get your ten percent of your total grade (which works out to maybe a third-of-a-grade credit and may be promoted from, say, a B+ to a A-).

1.5 Course Rules

The following are the course rules. **Read them and fully understand them!** Violations of the rules below will be considered academic dishonesty and may result in your summary dismissal from our graduate program!

- You **may collaborate with your classmates in any permutation on the homework**. In fact, I *encourage* you to work in groups, as you will probably all learn more that way. However, you must **each** write up *all* the solutions even if they are all the same within a group. Writing them up provides learning reinforcement.

- You **may not get worked out "solutions to Jackson problems" from more advanced graduate students, the solutions book (if you can find it), the web (!) or anyplace else**. It obviously removes the whole point of the homework in the first place. If you do not struggle with these problems (as I did and really, still do) you will not learn. It is a mistake to have too much guidance or to try to avoid the pain.

- You **may ask for help with your homework** from more advanced students, other faculty, personal friends, or your household pets, as long as **no worked–out solutions to the assigned problems are present at any sessions where the problems are discussed**. That way, with or without help, you will *participate in finding the solution of each problem*, which is the idea.

  Obviously your degree of long term success in the class and physics in general will depend largely on the enthusiasm and commitment of your personal participation. Passive is not good, active is good. Take charge of learning this material by doing the work required to do so.

- You **may (indeed must) use the library and all available non–human resources to help solve the problems**. I don't even care if you find the solution to some problem somewhere *in the literature* and **copy it verbatim** provided that you understand it afterwards (which is the goal), **cite your source**, and provided that you do **not** use the solution manual/book for Jackson problems (which exists, floating around somewhere, and which has all sorts of errors in it besides), see second item above.

- You **may NOT collaborate with each other or get outside human help on the take home (midterm, final) exam problems**. They are to be done **alone**. There will be a time limit (typically 24 hours total working time) on the take home exams, spread out over four days or so.

- I reserve the right to specify open or closed book (or note, or library) efforts for the midterm and final exams. In the past I have permitted these exams to be done open book, but there is some advantage to making them closed book as well. Please obey whatever rule is specified in the exam itself.

I will usually be available for questions right after class. Otherwise, it is best to make appointments to see me via e-mail. My third department/university job is helping to manage the computer network, especially with regard to cluster computing (teaching this is my second and doing research is my first) so I'm usually on the computer and always insanely busy.

However, I will nearly *always* try to answer questions if/when you catch me. That doesn't mean that I will know the answers, of course ...

I welcome feedback and suggestions at any time during the year. I would prefer to hear constructive suggestions early so that I have time to implement them this semester.

1.6 The Interplay of Physics and Mathematics

Before we begin, it is worth making one very important remark that can guide a student as they try to make sense of the many, many things developed in this work. As you go through this material, there will be a *strong* tendency to view it all as being nothing but mathematics. For example, we'll spend a *lot* of time studying the wave (partial differential) equation, Green's functions, and the like. This will "feel like" mathematics. This in turn inspires students to at least initially view every homework problem, every class derivation, as being just another piece of algebra.

This is a *bad* way to view it. Don't do this. This is a *physics* course, and the difference between physics and abstract mathematics is that physics *means* something, and the mathematics used in physics is always *grounded in physical law.* This means that solving the very difficult problems assigned throughout the semester, understanding the lectures and notes, developing a conceptual understanding of the *physics* involves a number of mental actions, not just one, and requires your whole brain, not just the symbolic sequential reasoning portions of your left brain.

To develop *insight* as well as problem solving skills, you need to be able to:

- *Visualize* what's going on. Electrodynamics is incredibly geometric. Visualization and spatiotemporal relationships are all *right* brain functions and transcend and guide the parsed logic of the left brain.
- *Care* about what's going on. You are (presumably) graduate students *interested* in physics, and this is some of the coolest physics ever discovered. Even better, it is cool *and* accessible; you *can* master it com-

pletely if you care to and work hard on it this semester. Be *engaged* in class, *participate* in classroom discussions, show *intiative* in your group studies outside of the classroom. Maybe I suck as an instructor – fine, so what? *You* are in charge of your own learning at this point, I'm just the 'facilitator' of a process you could pursue on your own.

- *Recognize* the division between *physics* and *mathematics* and *geometry* in the problem you're working on! This is the most difficult step for most students to achieve.

Most students, alas, will try to solve problems as if they were math problems and not use any physical intuition, geometric visualization, or (most important) *the fundamental physical relationships* upon which the solution is founded. Consequently they'll often start it using some physics, and then try to bull their way through the algebra, not realizing that at they need to add *more* physics from *different* relations at various points on the way through that algebra. This happens, in fact, starting with a student's first introductory physics class when they try to solve a loop-the-loop problem using *only* an expression for centripetal force, perhaps with Newton's laws, but ignore the fact that energy is conserved too. In electrodynamics it more often comes from e.g. starting with the wave equation (correctly) but failing to *re-insert* individual Maxwell equations into the reasoning process, failing to use e.g. charge conservation, failing to recognize a physical constraint.

After a long time and many tries (especially with Jackson problems, which are notorious for this) a student will often reach the perfect level of utter frustration and stop, scratch their head a bit, and decide to stop just doing math and try using a bit of *physics*, and half a page later the problem is solved. This *is* a valuable learning experience, but it is in some sense maximally painful. This short section is designed to help you at minimize that pain to at least some extent.

In the following text some small effort will be made on occasion to differentiate the "mathy" parts of a demonstration or derivation from the "physicsy" parts, so you can see where physics is being injected into a math result to obtain a new understanding, a new constraint or condition on an otherwise general solution, the next critical step on the true path to a desired solution to a problem. Students might well benefit from marking up their texts or notes as they go along in the same way.

What part of what you are writing down is "just math" (and hence something you can reasonably expect your math skills to carry you through later if need be) and what part is *physics* and relies on your knowledge of physical laws, visualizable physical relationships, and intuition? Think about that as you proceed through this text.

Part I

Mathematical Physics

Chapter 2

Mathematical Prelude

When I first started teaching classical electrodynamics, it rapidly became apparent to me that I was spending as much time teaching what amounted to remedial mathematics as I was teaching physics. After all, to even write Maxwell's equations down in either integral or differential form requires multivariate calculus – path integrals, surface integrals, gradients, divergences, curls. These equations are rapidly converted into inhomogeneous partial differential equations and their static and dynamic solutions are expanded in (multipolar) representations, requiring a knowledge of spherical harmonics and various hypergeometric solutions. The solutions are in many cases naturally expressed in terms of complex exponentials, and one requires a certain facility in doing e.g. contour integrals to be able to (for example) understand dispersion or establish representations between various forms of the Green's function. Green's functions themselves and Green's theorem emerge, which in turn requires a student to learn to integrate by parts in vector calculus. This culminates with the development of *vector* spherical harmonics, Hansen functions, and dyadic tensors in the integral equations that allow one to evaluate multipolar fields directly.

Then one hits theory of special relativity and does it all *again*, but now expressing everything in terms of *tensors* and the *theory of continuous groups*. It turns out that all the electrodynamics we worked so hard on is much, much easier to understand if it is expressed in terms of tensors of various rank[1].

[1]Some parts are simpler still if expressed in terms of the geometric extension of the

We discover that it is *essential* to understand tensors and tensor operations and notation in order to follow the formulation of relativity theory and relativistic electrodynamics in a compact, workable form. This is in part because some of the difficulties we have encountered in describing the electric and magnetic fields separately result from the fact that they are *not, in fact, vector fields!* They are components of a second rank field strength tensor and hence *mix* when one changes relativistic frames. Tensors are indeed the natural language of field theories (and much else) in physics, one that is unfortunately not effectively taught where they are taught at all.

The same is true of group theory. Relativity is best and most generally derived by looking for the group of all (coordinate) transformations that preserve a *scalar* form for certain physical quantities, that leave e.g. equations of motion such as the wave equation form invariant. There are strong connections between groups of transformations that conserve a property, the underlying symmetry of the system that requires that property to be conserved, and the labels and coordinatization of the physical description of the system. By effectively exploiting this symmetry, we can often tremendously simplify our mathematical description of a physical system even as we deduce physical laws associated with the symmetry.

Unfortunately, it is the rare graduate student that already knows complex variables and is skilled at doing contour integrals, is very comfortable with multivariate/vector calculus, is familiar with the relevant partial differential equations and their basic solutions, has any idea what you're talking about when you introduce the notion of tensors and manifolds, has worked through the general theory of the generators of groups of continuous transformations that preserve scalar forms, or have even *heard* of either geometric algebra or Hansen multipoles. So rare as to be practically non-existent.

I don't blame the students, of course. I didn't know it, either, when I was a student (if it can honestly be said that I *know* all of this now, for all that I try to teach it). Nevertheless filling in all of the missing pieces, one student at a time, very definitely detracts from the flow of teaching

---

graded division algebra associated with complex numbers: "geometric algebra". This is the algebra of a class of objects that includes the reals, the complex numbers, and the quaternions – as well as generalized objects of what used to be called "Clifford algebra". I urge interested students to check out Lasenby's lovely book on Geometric Algebra, especially the parts that describe the quaternionic formulation of Maxwell's equations.

electrodynamics, while if one *doesn't* bother to fill them in, one might as well not bother trying to teach the course at all.

Over the years in between I've tried many approaches to dealing with the missing math. The most successful one has been to insert little minilectures that focus on the math at appropriate points during the semester, which serve to both prepare the student and to give them a smattering of the basic facts that a good book on mathematical physics would give them, and to *also* require that the students *purchase* a decent book on mathematical physics even though the ones available tend to be encyclopediac and say far too much or omit whole crucial topics and thereby say far too little (or even both).

I'm now trying out a new, semi-integrated approach. This part of the book is devoted to a lightning fast, lecture note-level review of mathematical physics. Fast or not, it will endeavor to be quite complete, at least in terms of what is directly required for this course. However, this is very much a work in progress and I welcome feedback on the idea itself as well as mistakes of omission and commission as always. At the end of I list several readily available sources and references that I'm using myself as I write it and that you might use independently both to learn this material more completely and to check that what I've written is in fact correct and comprehensible.

Chapter 3

Vector Calculus: Integration by Parts

There is one essential theorem of vector calculus that is essential to the development of multipoles – computing the dipole moment. Jackson blithely integrates by parts (for a charge/current density with compact support) thusly:

$$\int_{\mathbf{R}^3} \boldsymbol{J} d^3x = -\int_{\mathbf{R}^3} (\boldsymbol{x} \boldsymbol{\nabla} \cdot \boldsymbol{J}) d^3x \tag{3.1}$$

Then, using the continuity equation and the fact that ρ and $\boldsymbol{J}$ are presumed harmonic with time dependenct $\exp(-i\omega t)$, we substitute $\boldsymbol{\nabla} \cdot \boldsymbol{J} = -\frac{\partial \rho}{\partial t} = i\omega\rho$ to obtain:

$$\int_{\mathbf{R}^3} \boldsymbol{J} d^3x = -i\omega \int_{\mathbf{R}^3} \boldsymbol{x}\rho(\boldsymbol{x})\ d^3x = -i\omega \boldsymbol{p} \tag{3.2}$$

where $\boldsymbol{p}$ is the dipole moment of the fourier component of the charge density distribution.

However, this leaves a nasty question: Just *how* does this integration by parts work? Where does the first equation come from? After all, we can't rely on always being able to look up a result like this, we have to be able to derive it and hence learn a method we can use when we have to do the same thing for a different functional form.

We might guess that deriving it will use the divergence theorem (or Green's theorem(s), if you like), but any naive attempt to make it do so will lead to pain and suffering. Let's see how it goes in this particularly nasty (and yet quite simple) case.

Recall that the idea behind integration by parts is to form the derivative of a product, distribute the derivative, integrate, and rearrange:

$$\begin{aligned} d(uv) &= u\ dv + v\ du \\ \int_a^b d(uv) &= \int_a^b u\ dv + \int_a^b v\ du \\ \int_a^b u\ dv &= (uv)|_a^b - \int_a^b v\ du \end{aligned} \tag{3.3}$$

where if the products $u(a)v(a) = u(b)v(b) = 0$ (as will often be the case when $a = -\infty, b = \infty$ and u and v have compact support) the process "throws the derivative from one function over to the other":

$$\int_a^b u\ dv = -\int_a^b v\ du \tag{3.4}$$

which is often extremely useful in evaluating integrals.

The exact same idea holds for vector calculus, except that the idea is to use the *divergence theorem* to form a *surface integral* instead of a boundary term. Recall that there are many forms of the divergence theorem, but they all map $\boldsymbol{\nabla}$ to $\hat{\boldsymbol{n}}$ in the following integral form:

$$\int_V \boldsymbol{\nabla}\ ...d^3x \to \oint_{S/V} \hat{\boldsymbol{n}}\ ...d^2x \tag{3.5}$$

or in words, if we integrate any form involving the pure gradient operator applied to a (possibly tensor) functional form indicated by the ellipsis ... in this equation, we can convert the result into an integral over the surface that bounds this volume, where the gradient operator is replaced by an outward directed normal but *otherwise the functional form of the expression is preserved*. So while the *divergence* theorem is:

$$\int_V \boldsymbol{\nabla} \cdot \boldsymbol{A}\ d^3x = \oint_{S/V} \hat{\boldsymbol{n}} \cdot \boldsymbol{A}\ d^2x \tag{3.6}$$

there is a "gradient theorem":

$$\int_V \boldsymbol{\nabla} f\ d^3x = \oint_{S/V} \hat{\boldsymbol{n}} f\ d^2x \tag{3.7}$$

and so on.

To prove Jackson's expression we might therefore try to find a suitable product whose divergence contains $\boldsymbol{J}$ as one term. This isn't too easy, however. The problem is finding the right tensor form. Let us look at the following divergence:

$$\begin{aligned}\boldsymbol{\nabla}\cdot(x\boldsymbol{J}) &= \boldsymbol{\nabla}x\cdot\boldsymbol{J}+x\boldsymbol{\nabla}\cdot\boldsymbol{J} \\ &= J_x+x\boldsymbol{\nabla}\cdot\boldsymbol{J} \qquad (3.8)\end{aligned}$$

This looks promising; it is the x-component of a result we might use. However, if try to apply this to a matrix dyadic form in what *looks* like it might be the right way:

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\boldsymbol{x}\boldsymbol{J}) &= (\boldsymbol{\nabla}\cdot\boldsymbol{x})\boldsymbol{J}+\boldsymbol{x}(\boldsymbol{\nabla}\cdot\boldsymbol{J}) \\ &= 3\boldsymbol{J}+\boldsymbol{x}(\boldsymbol{\nabla}\cdot\boldsymbol{J}) \qquad (3.9)\end{aligned}$$

we get the *wrong answer.*

To assemble the *right* answer, we have to sum over the three separate statements:

$$\begin{aligned}(\boldsymbol{\nabla}\cdot(x\boldsymbol{J}))\,\hat{\boldsymbol{x}} &= (J_x+x\boldsymbol{\nabla}\cdot\boldsymbol{J})\,\hat{\boldsymbol{x}} \\ +(\boldsymbol{\nabla}\cdot(y\boldsymbol{J}))\,\hat{\boldsymbol{y}} &= +(J_y+y\boldsymbol{\nabla}\cdot\boldsymbol{J})\,\hat{\boldsymbol{y}} \\ +(\boldsymbol{\nabla}\cdot(z\boldsymbol{J}))\,\hat{\boldsymbol{z}} &= +(J_z+z\boldsymbol{\nabla}\cdot\boldsymbol{J})\,\hat{\boldsymbol{z}}\end{aligned} \qquad (3.10)$$

or

$$\sum_i \hat{\boldsymbol{x}}_{\boldsymbol{i}}\,(\boldsymbol{\nabla}\cdot(x_i\boldsymbol{J})) = \boldsymbol{J}+\boldsymbol{x}(\boldsymbol{\nabla}\cdot\boldsymbol{J}) \qquad (3.11)$$

which is the sum of *three divergences*, not a divergence itself. If we integrate both sides over all space we get:

$$\begin{aligned}\int_{\mathbf{R}^3}\sum_i \hat{\boldsymbol{x}}_{\boldsymbol{i}}\,(\boldsymbol{\nabla}\cdot(x_i\boldsymbol{J}))\,d^3x &= \int_{\mathbf{R}^3}\boldsymbol{J}d^3x+\int_{\mathbf{R}^3}\boldsymbol{x}(\boldsymbol{\nabla}\cdot\boldsymbol{J})d^3x \qquad (3.12)\\ \sum_i \hat{\boldsymbol{x}}_{\boldsymbol{i}}\int_{S(\infty)}(\hat{\boldsymbol{n}}\cdot(x_i\boldsymbol{J}))\,d^3x &= \int_{\mathbf{R}^3}\boldsymbol{J}d^3x+\int_{\mathbf{R}^3}\boldsymbol{x}(\boldsymbol{\nabla}\cdot\boldsymbol{J})d^3x \qquad (3.13)\\ \sum_i \hat{\boldsymbol{x}}_{\boldsymbol{i}}0 &= \int_{\mathbf{R}^3}\boldsymbol{J}d^3x+\int_{\mathbf{R}^3}\boldsymbol{x}(\boldsymbol{\nabla}\cdot\boldsymbol{J})d^3x \qquad (3.14)\\ 0 &= \int_{\mathbf{R}^3}\boldsymbol{J}d^3x+\int_{\mathbf{R}^3}\boldsymbol{x}(\boldsymbol{\nabla}\cdot\boldsymbol{J})d^3x \qquad (3.15)\end{aligned}$$

where we have used the fact that $\boldsymbol{J}$ (and ρ) have *compact support* and are *zero* everywhere on a surface at infinity.

We rearrange this and get:

$$\int_{\mathbb{R}^3} \boldsymbol{J} d^3x = - \int_{\mathbb{R}^3} \boldsymbol{x}(\boldsymbol{\nabla} \cdot \boldsymbol{J}) d^3x \tag{3.16}$$

which is just what we needed to prove the conclusion.

This illustrates one of the most difficult examples of using integration by parts in vector calculus. In general, seek out a tensor form that can be expressed as a pure vector derivative and that evaluates to two terms, one of which is the term you wish to integrate (but can't) and the other the term you want *could* integrate if you could only proceed as above. Apply the generalized divergence theorem, throw out the boundary term (or not – if one keeps it one derives e.g. Green's Theorem(s), which are nothing more than integration by parts in this manner) and rearrange, and you're off to the races.

Note well that the tensor forms may *not be trivial!* Sometimes you do have to work a bit to find just the right combination to do the job.

Chapter 4

Numbers

(Note: This appendix is very much under development. Check back periodically.)

It may seem silly to devote space to numbers as physicists by hypothesis *love* numbers, but the standard undergraduate training of physicists does not include a course in number theory per se, so most of what they are likely to know is gleaned from at most one course in complex numbers (math double majors and math minors excepted). This chapter makes no attempt to present an exhaustive review of number theory (however cool and worthy of a deeper treatment it might be) but instead confines itself to just a few points of direct relevance to electrodynamics.

4.1 Real Numbers

Real numbers are of obvious importance in physics, and electrodynamics is no exception. Measured or counted quantities are almost invariably described in terms of real numbers or their embedded cousins, the integers. Their virtue in physics comes from from the fact that they form a (mathematical) *field*[1] that is, they support the mathematical operations of addition, subtraction, multiplication and division, and it empirically turns out that physical laws turn out to be describable in terms of algebraic forms based on (at least) real numbers. Real numbers form a *group* under or-

[1]Wikipedia: http://www.wikipedia.org/wiki/Field_mathematics. ;

dinary multiplication and, because multiplication is associative and each element possesses a unique inverse, they form a *division algebra*[2]

A division algebra is one where any element other than zero can be divided into any other element to produce a unique element. This property of real numbers is extremely important – indeed it is the property that makes it possible to use algebra per se to solve for many physical quantities from relations expressed in terms of products and sums. The operational steps:

$$\begin{aligned} b \cdot c &= a \\ (b \cdot c) \cdot c^{-1} &= a \cdot c^{-1} \\ b \cdot (c \cdot c^{-1}) &= a \cdot c^{-1} \\ b = b \cdot 1 &= a \cdot c^{-1} \end{aligned} \tag{4.1}$$

are so pervasively implicit in our algebraic operations *because* they are all learned in terms of real numbers that we no longer even think about them until we run into them in other contexts, for example when a, b, c are matrices, with at least c being an invertible matrix.

In any event real numbers are ideally suited for algebra because they form a field, in some sense the archetypical field, whereby physical law can be written down in terms of sums and products with measurable quantities and physical parameters represented by real numbers. Other fields (or rings) are often defined in terms of either subsets of the real numbers or extensions of the real numbers, if only because when we write a symbol for a real number in an algebraic computation we know exactly what we can and cannot do with it.

Real numbers are the basis of real "space" and "time" in physics – they are used to form an algebraic *geometry* wherein real numbers are spatiotemporal *coordinates*. This use is somewhat presumptive – spacetime cannot be probed at distances shorter than the *Planck length* (1.616×10^{-35} meters) – and may be quantized and granular at that scale. Whatever this may or may not mean (close to nothing, lacking a complete quantum theory of gravity) it makes no meaningful difference as far as the applicability of e.g. calculus down *to* that approximate length scale, and so our classical assumption of smooth spacetime will be quite reasonable.

Are real numbers *sufficient* to describe physics, in particular classical

[2]Wikipedia: http://www.wikipedia.org/wiki/Division_algebra. .

electrodynamics? The answer may in *some* sense be yes (because classical measurable quantities are invariably real, as are components of e.g. complex numbers) but as we will see, it will be *far easier* to work over a *different* field: complex numbers, where we will often view real numbers as just the real part of a more general complex number, the real line as just one line in a more general complex plane. As we will see, there is a close relationship between complex numbers and a two-dimensional Euclidean plane that permits us to view certain aspects of the dynamics of the *real* number valued measurable quantities of physics as the real projection of dynamics taking place on the complex plane. Oscillatory phenomena in general are often viewed in this way.

4.2 Complex Numbers

The operation of taking the square root (or any other roots) of a real number has an interesting history which we will not review here. Two aspects of number theory that have grown directly out of exploring square roots are, however, irrational numbers (since the square root of most integers can be shown to be irrational) and imaginary numbers. The former will not interest us as we already work over at *least* the real numbers which include all rationals and irrationals, positive and negative. Imaginary numbers, however, are a true extension of the reals.

Since the product of any two non-negative numbers is non-negative, and the product of any two negative numbers is similarly non-negative, we cannot find any real number that, when squared, is a negative number. This permits us to "imagine" a field of numbers where the square root of a nonzero negative number exists. Such a field cannot be identical to the reals already discussed above. It must *contain* the real numbers, though, in order to be closed under multiplication (as the square of an "imaginary" number is a negative real number, and the square of that real number is a positive real number).

If we define the *unit* imaginary number to be:

$$i = +\sqrt{-1} \tag{4.2}$$

such that

$$\pm\, i^2 = -1 \tag{4.3}$$

we can then form the rest of the field by scaling this imaginary unit through multiplication by a real number (to form the *imaginary axis*) and then generating the field of *complex numbers* by summing all possible combinations of real and imaginary numbers. Note that the imaginary axis alone does *not* form a field or even a multiplicative group as the product of any two imaginary numbers is always real, just as is the product of any two real numbers. However, the product of any real number and an imaginary number is always imaginary, and closure, identity, inverse and associativity can easily be demonstrated.

The easiest way to visualize complex numbers is by orienting the real axis at right angles to the imaginary axis and summing real and imaginary "components" to form all of the complex numbers. There is a one-to-one mapping between complex numbers and a Euclidean two dimensional plane as a consequence that is very useful to us as we seek to understand how this "imaginary" generalization works.

We can write an arbitrary complex number as $z = x+iy$ for real numbers x and y. As you can easily see, this number appears to be a point in a (complex) plane. Addition and subtraction of complex numbers are trivial – add or subtract the real and imaginary components separately (in a manner directly analogous to vector addition).

Multiplication, however, is a bit odd. Given two complex numbers z_1 and z_2, we have:

$$z = z_1 \cdot z_2 = x_1 x_2 + i(x_1 y_2 + y_1 x_2) - y_1 y_2 \tag{4.4}$$

so that the real and imaginary parts are

$$\Re z = x_1 x_2 - y_1 y_2 \tag{4.5}$$

$$\Im z = x_1 y_2 + y_1 x_2 \tag{4.6}$$

This is quite *different* from any of the rules we might use to form the product of two vectors. It also permits us to form the so-called *complex conjugate* of any imaginary number, the number that one can multiply it by to obtain a purely *real* number that appears to be the square of the

Euclidean length of the real and imaginary components

$$z = x + iy \tag{4.7}$$

$$z^* = x - iy \tag{4.8}$$

$$|z|^2 = z^* z = z z^* = x^2 + y^2 \tag{4.9}$$

A quite profound insight into the importance of complex numbers can be gained by representing a complex number in terms of the *plane polar coordinates* of the underlying Euclidian coordinate frame. We can use the product of a number z and its complex conjugate z^* to define the amplitude $|z| = +\sqrt{|z|^2|}$ that is the polar distance of the complex number from the complex origin. The usual polar angle θ can then be swept out from the positive real axis to identify the complex number on the circle of radius $|z|$. This representation can then be expressed in trigonometric forms as:

$$z = x + iy = |z|\cos(\theta) + i|z|\sin(\theta) \tag{4.10}$$

$$= |z|\left(\cos(\theta) + i\sin(\theta)\right) \tag{4.11}$$

$$= |z|e^{i\theta} \tag{4.12}$$

where the final result can be observed any number of ways, for example by writing out the power series of $e^u = 1 + u + u^2/2! + ...$ for complex $u = i\theta$ and matching the real and imaginary subseries with those for the cosine and sine respectively. In this expression

$$\theta = \tan^{-1}\frac{y}{x} \tag{4.13}$$

determines the angle θ in terms of the original "cartesian" complex coordinates.

Trigonometric functions are thus seen to be quite naturally expressible in terms of the exponentials of imaginary numbers. There is a price to pay for this, however. The representation is no longer single valued in θ. In fact, it is quite clear that:

$$z = |z|e^{i\theta \pm 2n\pi} \tag{4.14}$$

for any integer value of n. We usually avoid this problem initially by requiring $\theta \in (-\pi, \pi]$ (the "first leaf") but as we shall see, this leads to problems when considering products and roots.

It is quite easy to multiply two complex numbers in this representation:

$$\begin{aligned} z_1 &= |z_1|e^{i\theta_1} && (4.15) \\ z_2 &= |z_2|e^{i\theta_2} && (4.16) \\ z = z_1 z_2 &= |z_1||z_2|e^{i(\theta_1+\theta_2)} && (4.17) \end{aligned}$$

or the amplitude of the result is the *product* of the amplitudes and the phase of the result is the *sum* of the two phases. Since $\theta_1 + \theta_2$ may well be larger than π even if the two angles individually are not, to stick to our resolution to keep the resultant phase in the range $(\pi, \pi]$ we will have to form a suitable modulus to put it back in range.

Division can easily be represented as multiplication by the inverse of a complex number:

$$z^{-1} = \frac{1}{|z|}e^{-i\theta} \tag{4.18}$$

and it is easy to see that complex numbers are a multiplicative group and division algebra and we can also see that its multiplication is commutative.

One last operation of some importance in this text is the formation of roots of a complex number. It is easy to see that the square root of a complex number can be written as:

$$\sqrt{z} = \pm\sqrt{|z|}e^{i\theta/2} = \sqrt{|z|}e^{i(\theta/2 \pm n\pi)} \tag{4.19}$$

for any integer n. We usually insist on finding roots only within the first "branch cut", and return an answer only with a final phase in the range $(pi, \pi]$.

There is a connection here between the branches, leaves, and topology – there is really only *one* actual point in the complex plane that corresponds to z; the rest of the ways to reach that point are associated with a *winding number* m that tells one how many times one must circle the origin (and in which direction) to reach it from the positive real axis.

Thus there are *two* unique points on the complex plane (on the principle branch) that are square roots (plus multiple copies with different winding numbers on other branches). In problems where the choice doesn't matter we often choose the *first* one reached traversing the circle in a counterclockwise direction (so that it has a positive amplitude). In physics choice often matters for a specific problem – we will often choose the root based on e.g. the direction we wish a solution to propagate as it evolves in time.

Pursuing this general idea it is easy to see that $z^{\frac{1}{n}}$ where n is an integer are the points

$$|z|^{\frac{1}{n}} e^{i(\theta/n \pm 2m\pi/n)} \tag{4.20}$$

where $m = 0, 1, 2...$ as before. Now we will generally have n roots in the principle branch of z and will have to perform a cut to select the one desired while accepting that all of them can work equally well.

4.3 Contour Integration

4.4 Geometric Algebra

Chapter 5

Partial Differential Equations

(Note: This section is very much under development. Check back periodically.)

5.1 The Laplace Equation

5.2 The Helmholtz Equation

5.3 The Wave Equation

Chapter 6

Tensors

6.1 The Dyad and N-adic Forms

There are two very different ways to introduce the notion of a tensor. One is in terms of differential forms, especially the definition of the total differential. This form is ultimately the most useful (and we will dwell upon it below for this reason) but it is also algebraically and intuitively the most complicated. The other way is by contemplating the *outer product* of two vectors, otherwise known as a *dyad*.

We will introduce the dyad in a two dimensional Euclidean space with Cartesian unit vectors, but it is a completely general idea and can be used in an arbitrary n-manifold within a locally Euclidean patch. Suppose one has a vector $\boldsymbol{A} = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}}$ and another vector $\boldsymbol{B} = B_x\hat{\boldsymbol{x}} + B_y\hat{\boldsymbol{y}}$. If one simply multiplies these two vectors together as an *outer product* (ordinary multiplication with the distribution of the terms) one obtains the following result:

$$\boldsymbol{AB} = A_xB_x\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + A_xB_y\hat{\boldsymbol{x}}\hat{\boldsymbol{y}} + A_yB_x\hat{\boldsymbol{y}}\hat{\boldsymbol{x}} + A_yB_y\hat{\boldsymbol{y}}\hat{\boldsymbol{y}} \tag{6.1}$$

This product of vectors is called a *dyadic*, and each pair of unit vectors within is called a *dyad*.

A dyad is an interesting object. Each term appears to be formed out of the ordinary multiplicative product of two numbers (which we can easily and fully compute and understand) followed by a pair of unit vectors that are juxtaposed. What, exactly does this juxtaposition of unit vectors *mean*?

We can visualize (sort of) what $\hat{\boldsymbol{x}}$ by itself is – it is a unit vector in the x direction that we can scale to turn into all possible vectors that are aligned with the x-axis (or into components of general vectors in the two dimensional space). It is not so simple to visualize what a dyad $\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}$ is in this way.

The function of such a product becomes more apparent when we define how it works. Suppose with take the *inner product* (or scalar product, or contraction) of our vector $\boldsymbol{A}$ with the elementary dyad $\hat{\boldsymbol{x}}hx$. We can do this in either order (from either side):

$$\boldsymbol{A} \cdot (\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}) = (\boldsymbol{A} \cdot \hat{\boldsymbol{x}})\hat{\boldsymbol{x}} = A_x \hat{\boldsymbol{x}} \tag{6.2}$$

or

$$(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}) \cdot \boldsymbol{A} = \hat{\boldsymbol{x}}(\hat{\boldsymbol{x}} \cdot \boldsymbol{A}) = A_x \hat{\boldsymbol{x}} \tag{6.3}$$

We see that the inner product of a unit dyad $\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}$ with a vector serves to project out the *vector* that is the x-component of $\boldsymbol{A}$ (not the scalar magnitude of that vector A_x). The inner product of a dyad with a vector is a vector.

What about the product of other dyads with $\boldsymbol{A}$?

$$(\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}) \cdot \boldsymbol{A} = \hat{\boldsymbol{x}}(\hat{\boldsymbol{y}} \cdot \boldsymbol{A}) = A_y \hat{\boldsymbol{x}} \tag{6.4}$$

$$\boldsymbol{A} \cdot (\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}) = (\boldsymbol{A} \cdot \hat{\boldsymbol{x}})\hat{\boldsymbol{y}} = A_x \hat{\boldsymbol{y}} \tag{6.5}$$

which are *not* equal. In fact, these terms seem to create the *new* vector components that might result from the interchange of the x and y components of the vector $\boldsymbol{A}$, as do $(\hat{\boldsymbol{y}}\hat{\boldsymbol{x}}) \cdot \boldsymbol{A} = A_x \hat{\boldsymbol{y}}$ etc.

Note well! Some of the dyads commute with respect to an inner product of the dyad with a vector, others (e.g. $\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}$) do not! Our generalized dyadic multiplication produces what appear to be "intrinsically" non-commutative results when contracted with vectors on the left or the right respectively.

This is in fact a break point – if we pursue this product in one direction we could easily motivate and introduce *Geometric Algebra*, in terms of which Maxwell's equations can be written in a compact and compelling form. However, even without doing this, we can arrive at that a compelling form (that is, in fact, quaternionic), so we will restrain ourselves and only learn enough about tensors to be able to pursue the usual tensor form without worrying about whether or how it can be decomposed in a division algebra.

The thing to take out of the discussion so far is that in general the inner product of a dyad with a vector serves to project out the *scalar amplitude* of the vector on the left or the right and reconstruct a possibly *new* vector out of the remaining unit vector. Very shortly we are going to start writing relations that sum over basis vectors where the basis is not necessarily orthonormal (as this isn't really necessary or desireable when discussing curvilinear coordinate systems). To do this, I will introduce at this point the *Einstein summation convention* where writing a product with repeated indices implies summation over those indices:

$$\boldsymbol{A} = \sum_i A_i \hat{\boldsymbol{x}}_i = A_i \hat{\boldsymbol{x}}_i \tag{6.6}$$

You can see how the summation symbol is in some sense redundant unless for some reason we wish to focus on a single term in the sum. In tensor analysis this is almost never the case, so it is easier to just specify the exceptions.

Note that we can form general dyadic forms directly from the unit dyads without the intermediate step of taking the outer product of particular vectors, producing terms like $\{\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}, \hat{\boldsymbol{y}}\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}\}$. We can also take another outer product from the left or right with all of these forms producing *tryads*, terms like $\{\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}\hat{\boldsymbol{x}}, ...\hat{\boldsymbol{y}}\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}, \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}\hat{\boldsymbol{y}}\}$ (eight terms total). Furthermore we can repeat all of the arguments above in higher dimensional spaces, e.g. $\{\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}, \hat{\boldsymbol{x}}\hat{\boldsymbol{z}}, ..., \hat{\boldsymbol{z}}\hat{\boldsymbol{z}}\}$.

There is a clear one-to-one correspondance of these monad unit vectors to specific column vectors, e.g.:

$$\hat{\boldsymbol{x}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{6.7}$$

$$\hat{\boldsymbol{y}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{6.8}$$

$$\hat{\boldsymbol{z}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{6.9}$$

This correspondance *continues* through the various unit dyads, tryads:

$$\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.10}$$

$$\hat{\boldsymbol{x}}\hat{\boldsymbol{y}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.11}$$

and so on.

We will call all of these unit monads, dyads, tryads, and so on, as well as the quantities formed by multiplying them by ordinary numbers and summing them according to similar -adic type, *tensors*. As we can see, there are several ways of *representing* tensors that all lead to identical algebraic results, where one of the most compelling is the matrix representation illustrated above. Note well that the feature that differentiates tensors from "ordinary" matrices is that the components correspond to particular -adic combinations of *coordinate directions* in some linear vector space; tensors will *change*, as a general rule, when the underlying coordinate description is changed. Let us define some of the terms we will commonly use when working with tensors.

The dimension of the matrix in a matrix representation of a tensor quantity we call its *rank*. We have special (and yet familiar) names for the first few tensor ranks:

0th rank tensor or *scalar*. This is an "ordinary number", which may at the very least be real or complex, and possibly could be numbers associated with geometric algebras of higher grade. It's characteristic defining feature is that is is *invariant* under transformations of the underlying coordinate system. All of the following are algebraic examples of scalar quantities: $x, 1.182, \pi, A_x, \boldsymbol{A}\cdot\boldsymbol{B}$...

1st rank tensor or *vector*. This is a set of scalar numbers, each an amplitude corresponding to a particular unit vector or monad, and inherits its transformational properties from those of the underlying unit vectors. Examples: $\boldsymbol{A} = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}}$, $\{x_i\}$, $\{x^i\}$,

$$\hat{\boldsymbol{z}} = \left(\begin{array}{c} A_x \\ A_y \\ A_z \end{array} \right)$$

where the i in e.g. x^i does *not correspond to a power* but is rather a coordinate index corresponding to a *contra*variant (ordinary) vector where x_i similarly corresponds to a *co*variant vector, and where covariance and contravariance will be defined below.

2nd rank tensor or $D \times D$ *matrix* (where D is the dimension of the space, so the matrix has D^2 components). Examples: $C_{xy}\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}$, $\boldsymbol{AB}$, $\overset{\Leftrightarrow}{\boldsymbol{C}}$, A_{ij}, A_i^j, A^{ij},

$$\overset{\Leftrightarrow}{\boldsymbol{A}} = \left(\begin{array}{ccc} A_{xx} & A_{xy} & A_{xz} \\ A_{yx} & A_{yy} & A_{yz} \\ A_{zx} & A_{zy} & A_{zz} \end{array} \right)$$

where again in matrix context the indices may be raised or lowered to indicate covariance or contravariance in the particular row or column.

3rd and higher rank tensors are the $D \times D \times D$... matrices with a rank corresponding to the number of indices required to describe it. In physics we will have occassion to use tensors through the fourth rank occasionally, through the third rank fairly commonly, although most of the physical quantities of interest will be tensors of rank 0-2. For examples we will simply generalize that of the examples above, using $\overset{\Leftrightarrow}{\boldsymbol{T}}$ as a generic tensor form or (more often) explicitly indicating its indicial form as in $T_{111}, T_{112}, \ldots$ or ϵ_{ijk}.

Using an indicial form with the Einstein summation convention is very powerful, as we shall see, and permits us to fairly simply represent forms that would otherwise involve a large number of nested summations over all coordinate indices. To understand precisely how to go about it, however, we have to first examine coordinate transformations.

6.2 Coordinate Transformations

Suppose we have a coordinate frame K in D dimensions, where D will typically be 4 for relativistic spacetime (with the 0th coordinate equal to ct

as usual) or 3 for just the spatial part. To simplify our notation, we will use roman characters such as i, j, k for the three-vector spatial-only part of a four-vector, and use greek characters such as μ, ν, γ, δ for the entire four-vector (where recall, repeated indices imply summation over e.g. $i = 1, 2, 3$ or $\mu = 0, 1, 2, 3$, hence the distinction as it can be used to de-facto restrict the summation range).

Now suppose that we wish to transform to a new coordinate frame K'. At this time we place very few restrictions on this transformation. The transformation might, therefore, translate, rotate, rescale or otherwise alter the original coordinate description. As we do this, our *description* of physical quantities expressed in the old coordinates must systematically change to a description in the new coordinates, since the actual physical situation being described is not altered by the change in coordinate frames. All that is altered is our point of view.

Our first observation might be that it may not be *possible* to describe our physical quantities in the new frame if the transformation were completely general. For example, if the dimension of K' were different (either larger or smaller than that of K) we might well be unable to represent some of the physics that involved the missing coordinate or have a certain degree of arbitrariness associated with a new coordinate added on. A second possible problem involves regions of the two coordinate frames that cannot be made to correspond – if there is a patch of the K frame that simply does not map into a corresponding patch of the K' frame we cannot expect to correctly describe any physics that depends on coordinates inside the patch in the new frame.

These are *not irrelevant mathematical issues* to the physicist. A perpetual *open question* in physics is whether or not any parts of it involve *additional variables*. Those variables might just be "parameters" that can take on some range of values, or they might be supported only within spacetime scales that are too small to be directly observed (leaving us to infer what happens in these microscale "patches" from observations made on the macroscale), they may be macroscopic domains over which frame transformations are singular (think "black holes") or they may be actual *extra dimensions* – hidden variables, if you like – in which interactions and structure can occur that is only visible to us in our four dimensional spacetime in *projection*. With no *a priori* reason to include or exclude any of these possibilities, the wise scientist must be prepared to believe or disbelieve them all and to include them in the "mix" of possible explanations for otherwise difficult to understand phenomena.

However, our purposes here are more humble. We only want to be able to describe the relatively mundane coordinate transformations that do *not* involve singularities, unmatched patches, or additional or missing coordinate dimensions. We will therefore require that our coordinate transformations be *one-to-one* – each point in the spacetime frame K corresponds to one and only one point in the spacetime frame K' – and *onto* – no missing or extra patches in the K' frame. This suffices to make the transformations *invertible*. There will be two very general classes of transformation that satisfy these requirements to consider. In one of them, the new coordinates can be reached by means of a *parametric transformation* of the original ones where the parameters can be *continuously varied* from a set of 0 values that describe "no transformation". In the other, this is not the case.

For the *moment*, let's stick to the first kind, and start our discussion by looking at our friends the coordinates themselves. By definition, the untransformed coordinates of an inertial reference frame are *contravariant* vectors. We symbolize contravariant components (not just 4-vectors – this discussion applies to tensor quantities on all manifolds on the patch of coordinates that is locally flat around a point) with *superscript* indices:

$$x_{\text{contravariant}} = (x^0, x^1, x^2, x^3 ...) \tag{6.12}$$

where we are not going to discuss manifolds, curved spaces, tangent or cotangent bundles (much) although we will still use a few of these terms in a way that is hopefully clear in context. I encourage you to explore the references above to find discussions that extend into these areas. Note that I'm using a non-bold x to stand for a four-vector, which is pretty awful, but which is also very common.

Now let us define a mapping between a point (event) x in the frame K and the *same* point x' described in the K' frame. x in K consists of a set of four scalar numbers, its frame coordinates, and we need to transform these four numbers into four new numbers in K'. From the discussion above, we want this mapping to be a *continuous function* in both directions. That is:

$$x^{0'} = x^{0'}(x^0, x^1, x^2 ...) \tag{6.13}$$

$$
\begin{aligned}
x^{1'} &= x^{1'}(x^0, x^1, x^2 ...) && (6.14)\\
x^{2'} &= x^{2'}(x^0, x^1, x^2 ...) && (6.15)\\
... & && (6.16)
\end{aligned}
$$

and

$$
\begin{aligned}
x^0 &= x^0(x^{0'}, x^{1'}, x^{2'} ...) && (6.17)\\
x^1 &= x^1(x^{0'}, x^{1'}, x^{2'} ...) && (6.18)\\
x^2 &= x^2(x^{0'}, x^{1'}, x^{2'} ...) && (6.19)\\
... & && (6.20)
\end{aligned}
$$

have to both exist and be well behaved (continuously differentiable and so on). In the most general case, the coordinates have to be *linearly independent* and span the K or K' frames but are not necessarily orthonormal. We'll go ahead and work with orthonormal coordinate bases, however, which is fine since non-orthnormal bases can always be othogonalized with Gram-Schmidt and normalized anyway.

Given this formal transformation, we can write the following relation using the chain rule and definition of derivative:

$$dx^{0'} = \frac{\partial x^{0'}}{\partial x^0} dx^0 + \frac{\partial x^{0'}}{\partial x^1} dx^1 + \frac{\partial x^{0'}}{\partial x^2} dx^2 + \ldots \tag{6.21}$$

$$dx^{1'} = \frac{\partial x^{1'}}{\partial x^0} dx^0 + \frac{\partial x^{1'}}{\partial x^1} dx^1 + \frac{\partial x^{1'}}{\partial x^2} dx^2 + \ldots \tag{6.22}$$

$$dx^{2'} = \frac{\partial x^{2'}}{\partial x^0} dx^0 + \frac{\partial x^{2'}}{\partial x^1} dx^1 + \frac{\partial x^{2'}}{\partial x^2} dx^2 + \ldots \tag{6.23}$$

$$\vdots$$

where again, superscripts stand for indices and not powers in this context. We can write this in a tensor-matrix form:

$$\begin{pmatrix} dx^{0'} \\ dx^{1'} \\ dx^{2'} \\ \vdots \end{pmatrix} = \begin{pmatrix} \frac{\partial x^{0'}}{\partial x^0} & \frac{\partial x^{0'}}{\partial x^1} & \frac{\partial x^{0'}}{\partial x^1} & \cdots \\ \frac{\partial x^{1'}}{\partial x^0} & \frac{\partial x^{1'}}{\partial x^1} & \frac{\partial x^{1'}}{\partial x^1} & \cdots \\ \frac{\partial x^{2'}}{\partial x^0} & \frac{\partial x^{2'}}{\partial x^1} & \frac{\partial x^{2'}}{\partial x^1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} dx^0 \\ dx^1 \\ dx^2 \\ \vdots \end{pmatrix} \tag{6.24}$$

The determinant of the matrix above is called the *Jacobean* of the transformation and must not be zero (so the transformation is invertible. This matrix defines the differential transformation between the coordinates in the K and K' frame, given the invertible maps defined above. All first rank tensors that transform like the coordinates, that is to say according to this transformation matrix linking the two coordinate systems, are said to be *contravariant vectors* where obviously the coordinate vectors themselves are contravariant by this construction.

We can *significantly compress* this expression using Einsteinian summation:

$$dx^{i'} = \frac{\partial x^{i'}}{\partial x^j} dx^j \tag{6.25}$$

in which compact notation we can write the definition of an arbitrary contravariant vector A as being one that transforms according to:

$$A^{i'} = \frac{\partial x^{i'}}{\partial x^j} A^j \tag{6.26}$$

There, that was easy!

Chapter 7

Group Theory

(Note: This appendix is very much under development. Check back periodically.)

7.1 Definition of a Group

7.2 Groups of Transformation

Chapter 8

Math References

- www.grc.nasa.gov/WWW/K-12/Numbers/Math/documents/...
  ...Tensors_TM2002211716.pdf. This is a NASA white paper by Joseph C. Kolecki on the use of tensors in physics (including electrodynamics) and is quite lovely. It presents the modern view of tensors as entities linked both traditional bases and manifolds much as I hope to do here.

- *Mathematical Physics* by Donald H. Menzel, Dover Press, ISBN 0-486-60056-4. This book was written in 1947 and hence presents both the "old way" and the "new way" of understanding tensors. It is cheap (as are all Dover Press books) and actually is a really excellent desk reference for both undergraduate and graduate level classical physics in general! Section 27 in this book covers simple cartesian tensors, section 31 tensors defined in terms of transformations.

- *Schaum's Outline* series has a volume on vectors and tensors. Again an excellent desk reference, it has very complete sections on vector calculus (e.g. divergence theorem, stokes theorem), multidimensional integration (including definitions of the Jacobian and coordinate transformations between curvilinear systems) and tensors (the old way).

- http://www.mathpages.com/rr/s5-02/5-02.htm This presents tensors in terms of the manifold coordinate description and is actually quite lovely. It is also just a part of http://www.mathpages.com/, a rather huge collection of short articles on all sorts of really cool problems with absolutely no organization as far as I can tell. Fun to look over and sometimes very useful.

- Wikipedia: http://www.wikipedia.org/wiki/Manifold Tensors tend to be described in terms of coordinates on a *manifold*. An n-dimensional manifold is basically a mathematical space which can be covered with locally Euclidean "patches" of coordinates. The patches must overlap so that one can move about from patch to patch without ever losing the ability to describe position in local "patch coordinates" that are Euclidean (in mathematese, this sort of neighborhood is said to be "homeomorphic to an open Euclidean n-ball"). The manifolds of interest to us in our discussion of tensors are *differentiable* manifolds, manifolds on which one can do calculus, as the transformational definition of tensors requires the ability to take derivatives on the underlying manifold.

- Wikipedia: http://www.wikipedia.org/wiki/Tensor This reference is (for Wikipedia) somewhat lacking. The better material is linked to this page, see e.g.
  Wikipedia: http://www.wikipedia.org/wiki/Covariant vector and
  Wikipedia: http://www.wikipedia.org/wiki/Contravariant vector
  and much more.

- http://www.mth.uct.ac.za/omei/gr/chap3/frame3.html This is a part of a "complete online course in tensors and relativity" by Peter Dunsby. It's actually pretty good, and is definitely modern in its approach.

- http://grus.berkeley.edu/~jrg/ay202/node183.html This is a section of an online astrophysics text or set of lecture notes. The tensor review is rather brief and not horribly complete, but it is adequate and is in the middle of other useful stuff.

Anyway, you get the idea – there are plentiful resources in the form of books both paper and online, white papers, web pages, and wikipedia articles that you can use to *really* get to where you understand tensor algebra, tensor calculus (differential geometry), and group theory. As you do so you'll find that many of the things you've learned in mathematics and physics classes in the past become simplified notationally (even as their core content of course does not change).

As footnoted above, this simplification becomes even greater when some of the ideas are further extended into a general geometric division algebra, and I strongly urge interested readers to obtain and peruse Lasenby's book on *Geometric Algebra*. One day I may attempt to add a section on it here as well and try to properly unify the geometric algebraic concepts embedded in the particular tensor forms of relativistic electrodynamics.

Part II

Non-Relativistic Electrodynamics

Chapter 9

Plane Waves

9.1 The Free Space Wave Equation

9.1.1 Maxwell's Equations

We begin with Maxwell's Equations (ME):

$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = \rho \tag{9.1}$$

$$\boldsymbol{\nabla} \times \boldsymbol{H} - \frac{\partial \boldsymbol{D}}{\partial t} = \boldsymbol{J} \tag{9.2}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \tag{9.3}$$

$$\boldsymbol{\nabla} \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \tag{9.4}$$

in SI units, where $\boldsymbol{D} = \epsilon \boldsymbol{E}$ and $\boldsymbol{H} = \boldsymbol{B}/\mu$. By this point, remembering these should be second nature, and you should really be able to freely go back and forth between these and their integral formulation, and derive/justify the Maxwell Displacement current in terms of charge conservation, etc. Note that there are two inhomogeneous (source-connected) equations and two homogeneous equations, and that the inhomogeneous forms are the ones that are medium-dependent. This is significant for later, remember it.

For the moment, let us express the inhomogeneous MEs in terms of just $\boldsymbol{E} = \epsilon \boldsymbol{D}$ and $\boldsymbol{B} = \boldsymbol{H}/\mu$, explicitly showing the *permittivity* ϵ and the *permeability* μ[1]:

[1] In SI units, now that Jackson 3rd finally dropped the curséd evil of Gaussian units.

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\epsilon} \tag{9.5}$$

$$\nabla \times \boldsymbol{B} - \mu\epsilon\frac{\partial \boldsymbol{E}}{\partial t} = \mu \boldsymbol{J} \tag{9.6}$$

It is difficult to convey to you how important these four equations are going to be to us over the course of the semester. Over the next few months, then, we will make Maxwell's Equations dance, we will make them sing, we will "mutilate" them (turn them into distinct coupled equations for transverse and longitudinal field components, for example) we will couple them, we will transform them into a manifestly covariant form, we will solve them microscopically for a point-like charge in general motion. We will (hopefully) *learn* them.

For the next two chapters we will primarily be interested in the properties of the field in regions of space without charge (sources). Initially, we'll focus on a vacuum, where there is no dispersion at all; later we'll look a bit at dielectric media and dispersion. In a source-free region, $\rho = 0$ and $\boldsymbol{J} = 0$ and we obtain **Maxwell's Equations in a Source Free Region of Space:**

$$\nabla \cdot \boldsymbol{E} = 0 \tag{9.7}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{9.8}$$

$$\nabla \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \tag{9.9}$$

$$\nabla \times \boldsymbol{B} - \epsilon\mu\frac{\partial \boldsymbol{E}}{\partial t} = 0 \tag{9.10}$$

where for the moment we ignore any possibility of dispersion (frequency dependence in ϵ or μ).

9.1.2 The Wave Equation

After a little work (two curls together, using the identity:

$$\nabla \times (\nabla \times \mathbf{a}) = \nabla(\nabla \cdot \mathbf{a}) - \nabla^2 \mathbf{a} \tag{9.11}$$

Mostly, anyway. Except for relativity.

and using Gauss' Laws) we can easily find that $\boldsymbol{E}$ and $\boldsymbol{B}$ satisfy the **wave equation**:

$$\nabla^2 u - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0 \tag{9.12}$$

(for $u = \boldsymbol{E}$ or $u = \boldsymbol{B}$) where

$$v = \frac{1}{\sqrt{\mu\epsilon}}. \tag{9.13}$$

The wave equation separates[2] for harmonic waves and we can actually write the following homogeneous PDE for just the spatial part of $\boldsymbol{E}$ or $\boldsymbol{B}$:

$$\left(\nabla^2 + \frac{\omega^2}{v^2}\right)\boldsymbol{E} = \left(\nabla^2 + k^2\right)\boldsymbol{E} = 0 \tag{9.14}$$

$$\left(\nabla^2 + \frac{\omega^2}{v^2}\right)\boldsymbol{B} = \left(\nabla^2 + k^2\right)\boldsymbol{B} = 0 \tag{9.15}$$

where the time dependence is implicitly $e^{-i\omega t}$ and where $v = \omega/k$.

This is called the *homogeneous Helmholtz equation* (HHE) and we'll spend a lot of time studying it and its inhomogeneous cousin. Note that it reduces in the $k \to 0$ limit to the familiar homogeneous Laplace equation, which is basically a special case of this PDE.

Observing that[3]:

$$\boldsymbol{\nabla} e^{ik\boldsymbol{n}\cdot\boldsymbol{x}} = ik\boldsymbol{n}e^{ik\boldsymbol{n}\cdot\boldsymbol{x}} \tag{9.16}$$

we can easily see that the wave equation has (among many, many others) *a* solution on $\mathbb{R}^3$ that looks like:

$$u(\mathbf{x}, t) = u_0 e^{i(k\boldsymbol{n}\cdot\boldsymbol{x} - \omega t)} \tag{9.17}$$

where the **wave number** $\boldsymbol{k} = k\boldsymbol{n}$ has the magnitude

$$k = \frac{\omega}{v} = \sqrt{\mu\epsilon}\,\omega \tag{9.18}$$

and determines the propagation direction of this *plane wave*.

[2] In case you've forgotten: Try a solution such as $u(\mathbf{x}, t) = X(x)Y(y)Z(z)T(t)$, or (with a bit of inspiration) $\boldsymbol{E}(\mathbf{x})e^{-i\omega t}$ in the differential equation. Divide by u. You end up with a bunch of terms that can each be identified as being constant as they depend on x, y, z, t separately. For a suitable choice of constants one obtains the following PDE for spatial part of harmonic waves.

[3] Yes, you should work this out termwise if you've never done so before. Don't just take my word for *anything*.

9.1.3 Plane Waves

Plane waves can propagate in any direction. Any superposition of these waves, for all possible $\omega, \boldsymbol{k}$, is also a solution to the wave equation. However, recall that $\boldsymbol{E}$ and $\boldsymbol{B}$ are not independent, which restricts the solution in electrodynamics somewhat.

To get a feel for the interdependence of $\boldsymbol{E}$ and $\boldsymbol{B}$, let's pick $\boldsymbol{k} = \pm k\hat{\boldsymbol{x}}$ so that e.g.:

$$\begin{aligned} \boldsymbol{E}(x,t) &= \boldsymbol{E}_+ e^{i(kx-\omega t)} + \boldsymbol{E}_- e^{i(-kx-\omega t)} && (9.19) \\ \boldsymbol{B}(x,t) &= \boldsymbol{B}_+ e^{i(kx-\omega t)} + \boldsymbol{B}_- e^{i(-kx-\omega t)} && (9.20) \end{aligned}$$

which are plane waves travelling to the right or left along the x-axis for any complex $\boldsymbol{E}_+, \boldsymbol{E}_-$, $\boldsymbol{B}_+, \boldsymbol{B}_-$. In one dimension, at least, if there is no dispersion we can construct a fourier series of these solutions for various k that converges to any well–behaved function of a single variable.

[Note in passing that:

$$u(x,t) = f(x - vt) + g(x + vt) \tag{9.21}$$

for arbitrary smooth $f(z)$ and $g(z)$ is the most general solution of the 1-dimensional wave equation. *Any* waveform that preserves its shape and travels along the x-axis at speed v is a solution to the one dimensional wave equation (as can be verified directly, of course). How boring! These particular harmonic solutions have this form (verify this).]

If there is dispersion (velocity a function of frequency) then the fourier superposition is no longer stable and the last equation no longer holds. *Each* fourier component is still an exponential, but their velocity is different, and a wave packet spreads out it propagates. We'll look at this shortly to see how it works for a very simple (gaussian) wave packet but for now we'll move on.

Note that $\boldsymbol{E}$ and $\boldsymbol{B}$ are connected by having to satisfy Maxwell's equations even if the wave is travelling in just one direction (say, in the direction of a unit vector $\boldsymbol{n}$); we cannot choose the wave amplitudes separately. Suppose

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{x},t) &= \boldsymbol{\mathcal{E}} e^{i(k\boldsymbol{n}\cdot\boldsymbol{x}-\omega t)} \\ \boldsymbol{B}(\boldsymbol{x},t) &= \boldsymbol{\mathcal{B}} e^{i(k\boldsymbol{n}\cdot\boldsymbol{x}-\omega t)} \end{aligned}$$

where $\boldsymbol{\mathcal{E}}$, $\boldsymbol{\mathcal{B}}$, and $\boldsymbol{n}$ are constant vectors (which may be complex, at least for the moment).

Note that applying $(\nabla^2 + k^2)$ to these solutions in the HHE leads us to:

$$k^2 \boldsymbol{n} \cdot \boldsymbol{n} = \mu\epsilon\omega^2 = \frac{\omega^2}{v^2} \tag{9.22}$$

as the condition for a solution. Then a *real* $\boldsymbol{b} \cdot \boldsymbol{n} = 1$ leads to the plane wave solution indicated above, with $k = \frac{\omega}{v}$, which is the most familiar form of the solution (but not the only one)!

This has mostly been "mathematics", following more or less directly from the wave equation. The same reasoning might have been applied to sound waves, water waves, waves on a string, or "waves" $u(x,t)$ of nothing in particular. Now let's use some *physics* in the spirit suggested in the last section of the Syllabus and see what it tells us about the *particular* electromagnetic waves that follow from Maxwell's equations turned into the wave equation. These waves all satisfy *each* of Maxwell's equations separately.

For example, from Gauss' Laws we see e.g. that:

$$\begin{aligned} \boldsymbol{\nabla} \cdot \boldsymbol{E} &= 0 \\ \boldsymbol{\nabla} \cdot \boldsymbol{\mathcal{E}} e^{i(k\boldsymbol{n}\cdot\boldsymbol{x} - \omega t)} &= 0 \\ \boldsymbol{\mathcal{E}} \cdot \boldsymbol{\nabla} e^{i(k\boldsymbol{n}\cdot\boldsymbol{x} - \omega t)} &= 0 \\ ik\boldsymbol{\mathcal{E}} \cdot \boldsymbol{n} e^{i(k\boldsymbol{n}\cdot\boldsymbol{x} - \omega t)} &= 0 \end{aligned} \tag{9.23}$$

or (dividing out nonzero terms and then repeating the reasoning for $\boldsymbol{B}$):

$$\boldsymbol{n} \cdot \boldsymbol{\mathcal{E}} = 0 \quad \text{and} \quad \boldsymbol{n} \cdot \boldsymbol{\mathcal{B}} = 0. \tag{9.24}$$

Which basically means for a **real unit vector** $\boldsymbol{n}$ that $\boldsymbol{E}$ and $\boldsymbol{B}$ are *perpendicular* to $\boldsymbol{n}$, the direction of propagation! A plane electromagnetic wave is therefore a **transverse** wave. This seems like it is an important thing to know, and is not at all a *mathematical* conclusion of the wave equation per se.

Repeating this sort of thing using one of the the curl eqns (say, Faraday's law) one gets:

$$\boldsymbol{\mathcal{B}} = \sqrt{\mu\epsilon}\,(\boldsymbol{n} \times \boldsymbol{\mathcal{E}}) \tag{9.25}$$

(the i cancels, $k/\omega = 1/v = \sqrt{\epsilon\mu}$). This means that $\boldsymbol{E}$ and $\boldsymbol{B}$ have the **same phase** if $\boldsymbol{n}$ is **real**[4]

If $\boldsymbol{n}$ is real (and hence a unit vector), then we can introduce three real, mutually orthogonal unit vectors $(\hat{\boldsymbol{\epsilon_1}}, \hat{\boldsymbol{\epsilon_2}}, \hat{\boldsymbol{n}})$ and use them to express the field strengths:

$$\boldsymbol{\mathcal{E}}_1 = \hat{\boldsymbol{\epsilon}}_1 E_0, \quad \boldsymbol{\mathcal{B}}_1 = \hat{\boldsymbol{\epsilon}}_2 \sqrt{\mu\epsilon} E_0 \tag{9.26}$$

and

$$\boldsymbol{\mathcal{E}}_2 = \hat{\boldsymbol{\epsilon}}_2 E_0', \quad \boldsymbol{\mathcal{B}}_2 = -\hat{\boldsymbol{\epsilon}}_1 \sqrt{\mu\epsilon} E_0' \tag{9.27}$$

where E_0 and E_0' are constants that may be complex. It is worth noting that

$$|E| = v|B| \tag{9.28}$$

have the same dimensions and that the magnitude of the electric field is greater than that of the magnetic field to which it is coupled via Maxwell's Equations by a factor of the speed of light in the medium, as this will be used a *lot* in electrodynamics.

These relations describe a wave propagating in the direction $\hat{\boldsymbol{n}} = \hat{\boldsymbol{\epsilon}}_1 \times \hat{\boldsymbol{\epsilon}}_2 = \hat{\boldsymbol{\epsilon}}_3$. This follows from the (time-averaged) Poynting vector (for any particular component pair):

$$\begin{aligned}
\boldsymbol{S} &= \frac{1}{2}(\boldsymbol{E} \times \boldsymbol{H}^*) && (9.29)\\
&= \frac{1}{2\mu}(\boldsymbol{E} \times \boldsymbol{B}^*) && (9.30)\\
&= \frac{\sqrt{\epsilon\mu}}{2\mu}(\boldsymbol{E} \times v\boldsymbol{B}^*) && (9.31)\\
&= \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}} \mid E_0 \mid^2 \hat{\boldsymbol{n}} && (9.32)
\end{aligned}$$

Now, kinky as it may seem, there is no real[5] reason that $\boldsymbol{k} = k\boldsymbol{n}$ cannot be complex (while k remains real!) As an exercise, figure out the complex vector of your choice such that

$$\boldsymbol{n} \cdot \boldsymbol{n} = 1. \tag{9.33}$$

[4] Whoops! You mean $\boldsymbol{n}$ doesn't have to be real? See below. Note also that we are assuming ϵ and μ are real as well, and they don't have to be either.

[5] Heh, heh.

Since I don't really expect you to do that (gotta save your strength for the real problems later) I'll work it out for you. Note that this is:

$$\boldsymbol{n} = \boldsymbol{n}_R + i\boldsymbol{n}_I \tag{9.34}$$

$$n_R^2 - n_I^2 = 1 \tag{9.35}$$

$$\boldsymbol{n}_R \cdot \boldsymbol{n}_I = 0. \tag{9.36}$$

So, $\boldsymbol{n}_R$ must be orthogonal to $\boldsymbol{n}_I$ and the difference of their squares must be one. For example:

$$\boldsymbol{n}_R = \sqrt{2}\,\hat{\mathbf{i}} \quad \boldsymbol{n}_I = 1\,\hat{\mathbf{j}} \tag{9.37}$$

works, as do infinitely more More generally (recalling the properties of hyberbolics functions):

$$\boldsymbol{n} = \hat{\boldsymbol{e}}_1 \cosh\theta + i\hat{\boldsymbol{e}}_2 \sinh\theta \tag{9.38}$$

where the unit vectors are orthogonal should work for any θ.

Thus the **most general $\boldsymbol{\mathcal{E}}$** such that $\mathbf{n} \cdot \boldsymbol{\mathcal{E}} = 0$ is

$$\boldsymbol{\mathcal{E}} = (i\hat{\boldsymbol{e}}_1 \sinh\theta - \hat{\boldsymbol{e}}_2 \cosh\theta)A + \hat{\boldsymbol{e}}_3 B \tag{9.39}$$

where (sigh) A and B are again, arbitrary complex constants. Note that if $\boldsymbol{n}$ is complex, the exponential part of the fields becomes:

$$e^{i(k\boldsymbol{n}\cdot\boldsymbol{x}-\omega t)} = e^{-k\boldsymbol{n}_I\cdot\boldsymbol{x}} e^{i(k\boldsymbol{n}_R\cdot\boldsymbol{x}-\omega t)}. \tag{9.40}$$

This *inhomogeneous plave wave* exponentially grows or decays in some direction while remaining a "plane wave" in the other (perpendicular) direction.

Fortunately, nature provides us with few sources that produce this kind of behavior (Imaginary $\boldsymbol{n}$? Just imagine!) in electrodynamics. So let's forget it for the moment, but remember that it is there for when you run into it in field theory, or mathematics, or catastrophe theory.

Instead we'll concentrate on kiddy physics descriptions of polarization when $\boldsymbol{n}$ is a real unit vector, continuing the reasoning above.

9.1.4 Polarization of Plane Waves

We've really done all of the hard work already in setting things up above (and it wasn't too hard). Indeed, the $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ defined a few equations back are just two independent polarizations of a transverse plane wave. However, we need to explore the rest of the *physics*, and understand just what is going on in the whole electrodynamic field and not just the electric field component of same.

Let's start by writing $\boldsymbol{E}$ in a fairly general way:

$$\boldsymbol{E}_i = \hat{\boldsymbol{\epsilon}}_i E_i e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}, \tag{9.41}$$

Then we can return (as we will, over and over) to the curl equations to find:

$$\boldsymbol{B}_i = \sqrt{\mu\epsilon}\frac{\boldsymbol{k}\times\boldsymbol{E}_i}{k} \tag{9.42}$$

for $i = 1, 2$ and $\hat{\boldsymbol{\epsilon}}_i$ a unit vector perpendicular to the direction of propagation $\boldsymbol{n}$.

Then generally,

$$\boldsymbol{E}(\boldsymbol{x}, t) = (\hat{\boldsymbol{\epsilon}}_1 E_1 + \hat{\boldsymbol{\epsilon}}_2 E_2)e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} \tag{9.43}$$

$$\boldsymbol{B}(\boldsymbol{x}, t) = \frac{1}{v}(\hat{\boldsymbol{\epsilon}}_2 E_1 - \hat{\boldsymbol{\epsilon}}_1 E_2)e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} \tag{9.44}$$

where E_1 and E_2 are (as usual) complex amplitudes since there is no reason (even in nature) to assume that the fields polarized in different directions have the same phase. (Note that a complex E corresponds to a simple phase shift in the exponential.)

The *polarization* of the plane wave describes the relative direction, magnitude, and phase of the *electric* part of the wave. We have several well-known cases:

1. If E_1 and E_2 have the same phase (but different magnitude) we have **Linear Polarization** of the $\boldsymbol{E}$ field with the polarization vector making an angle $\theta = \tan^{-1}(E_2/E_1)$ with ϵ_1 and magnitude $E = \sqrt{E_1^2 + E_2^2}$. Frequently we will choose coordinates in this case so that (say) $E_2 = 0$.

2. If E_1 and E_2 have different phases and different magnitudes, we have **Elliptical Polarization**. It is fairly easy to show that the electric field strength traces out an *ellipse* in the $1, 2$ plane.

3. A special case of elliptical polarization results when the amplitudes are out of phase by $\pi/2$ and the magnitudes are equal. In this case we

have **Circular Polarization**. Since $e^{i\pi/2} = i$, in this case we have a wave of the form:

$$\boldsymbol{E} = \frac{E_0}{\sqrt{2}} \left(\hat{\boldsymbol{\epsilon_1}} \pm i \hat{\boldsymbol{\epsilon}}_2 \right) = E_0 \hat{\boldsymbol{\epsilon}}_\pm . \tag{9.45}$$

where we have introduced complex *unit helicity vectors* such that:

$$\hat{\boldsymbol{\epsilon}}_\pm^* \cdot \hat{\boldsymbol{\epsilon}}_\mp = 0 \tag{9.46}$$
$$\hat{\boldsymbol{\epsilon}}_\pm^* \cdot \hat{\boldsymbol{\epsilon}}_3 = 0 \tag{9.47}$$
$$\hat{\boldsymbol{\epsilon}}_\pm^* \cdot \hat{\boldsymbol{\epsilon}}_\pm = 1 \tag{9.48}$$

As we can see from the above, elliptical polarization can have positive or negative **helicity** depending on whether the polarization vector swings around the direction of propagation counterclockwise or clockwise when looking into the oncoming wave.

Another completely general way to represent a polarized wave is via the unit helicity vectors:

$$\boldsymbol{E}(\mathbf{x}, t) = \left(E_+ \hat{\boldsymbol{\epsilon}}_+ + E_- \hat{\boldsymbol{\epsilon}}_- \right) e^{i\boldsymbol{k}\cdot\boldsymbol{x} - \omega t} \tag{9.49}$$

It is left as an exercise to prove this. Note that as always, $E_\pm$ are *complex* amplitudes!

I'm leaving Stokes parameters out, but you should read about them on your own in case you ever need them (or at least need to know what they are). They are relevant to the issue of *measuring* mixed polarization states, but are no more general a description of polarization itself than either of those above.

9.2 Reflection and Refraction at a Plane Interface

Suppose a plane wave is incident upon a plane surface that is an interface between two materials, one with μ, ϵ and the other with μ', ϵ'.

Incident Wave

$$\boldsymbol{E} = \boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} \tag{9.50}$$
$$\boldsymbol{B} = \sqrt{\mu\epsilon}\frac{\boldsymbol{k}\times\boldsymbol{E}}{k} \tag{9.51}$$

Refracted Wave

$$\boldsymbol{E'} = \boldsymbol{E'}_0 e^{i(\boldsymbol{k'}\cdot\boldsymbol{x}-\omega t)} \tag{9.52}$$
$$\boldsymbol{B'} = \sqrt{\mu'\epsilon'}\frac{\boldsymbol{k'}\times\boldsymbol{E'}}{k'} \tag{9.53}$$

Reflected Wave

$$\boldsymbol{E''} = \boldsymbol{E''}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} \tag{9.54}$$
$$\boldsymbol{B''} = \sqrt{\mu\epsilon}\frac{\boldsymbol{k}\times\boldsymbol{E''}}{k'} \tag{9.55}$$

where the reflected wave and incident wave do not leave the first medium and hence retain speed $v = 1/\sqrt{\mu\epsilon}$, μ, ϵ and $k = \omega\sqrt{\mu\epsilon} = \omega/v$. The refracted wave changes to speed $v' = 1/\sqrt{\mu'\epsilon'}$, μ', $k' = \omega\sqrt{\mu'\epsilon'} = \omega/v'$.

[Note that the *frequency* ω of the wave is the same in both media! Ask yourself why this must be so as a kinematic constraint...]

Our goal is to *completely understand* how to compute the reflected and refracted wave from the incident wave. This is done by matching the wave across the boundary interface. There are two aspects of this matching – a *static* or kinematic matching of the waveform itself and a *dynamic* matching associated with the (changing) polarization in the medium. These two kinds of matching lead to two distinct and well-known results.

9.2.1 Kinematics and Snell's Law

The phase factors of all three waves must be *equal* on the actual boundary itself, hence:

$$(\boldsymbol{k} \cdot \boldsymbol{x})_{z=0} = (\boldsymbol{k'} \cdot \boldsymbol{x})_{z=0} = (\boldsymbol{k''} \cdot \boldsymbol{x})_{z=0} \tag{9.56}$$

as a *kinematic constraint* for the wave to be consistent. That is, this has nothing to do with "physics" per se, it is just a mathematical requirement for the wave *description* to work. Consequently it is generally covered even in kiddy-physics classes, where one can derive Snell's law just from pictures of incident waves and triangles and a knowledge of the wavelength shift associated with the speed shift with a fixed frequency wave.

At $z = 0$, the three $\boldsymbol{k}$'s must lie in a plane and we obtain:

$$\begin{aligned} k \sin\theta_{\text{incident}} &= k' \sin\theta_{\text{refracted}} = k \sin\theta_{\text{reflected}} \\ n \sin\theta_{\text{incident}} &= n' \sin\theta_{\text{refracted}} = n \sin\theta_{\text{reflected}} \end{aligned} \tag{9.57}$$

which is both *Snell's Law* and the *Law of Reflection*, where we use $k = \omega/v = n\omega/c$ to put it in terms of the *index of refraction*, defined by $v = c/n$. Note that we cancel ω/c, using the fact that the frequency is the same in both media.

9.2.2 Dynamics and Reflection/Refraction

Now we do the dynamics, that is to say, the real physics. Real physics is associated with the equations of motion of the EM field, that is, with Maxwell's equations, which in turn become the wave equation, so dynamics is associated with the boundary value problem satisfied by the (wave equation) PDEs.

So what *are* those boundary conditions? Recall that the electric *displacement* perpendicular to the surface must be continuous, that the electric *field* parallel to the surface must be continuous, that the magnetic *field* parallel to the surface must be continuous and the magnetic *induction* perpendicular to the surface must be continuous.

To put it another (more physical) way, the perpendicular components of the electric field will be discontinous at the surface due to the surface charge layer associated with the local polarization of the medium in response to the wave. This polarization is actually *not instantaneous*, and is a *bulk response* but here we will assume that the medium can react instantly as the wave arrives and that the wavelength includes many atoms so that the response is a collective one. These assumptions are valid for e.g. visible light incident on ordinary "transparent" matter. Similarly, surface current loops cause magnetic induction components parallel to the surface to be discontinuously changed.

Algebraically, this becomes (for $\boldsymbol{E}$):

$$\epsilon(\boldsymbol{E}_0 + \boldsymbol{E}_0'') \cdot \hat{\boldsymbol{n}} = \epsilon' \boldsymbol{E}_0' \cdot \hat{\boldsymbol{n}} \tag{9.58}$$

$$(\boldsymbol{E}_0 + \boldsymbol{E}_0'') \times \hat{\boldsymbol{n}} = \boldsymbol{E}_0' \times \hat{\boldsymbol{n}} \tag{9.59}$$

where the latter cross product is just a fancy way of finding $E_\perp$ components. In most cases one wouldn't actually "do" this decomposition algebraically, one would just inspect the problem and write down the $||$ and $\perp$ components directly using a sensible coordinate system (such as one where $\hat{\boldsymbol{n}} = \hat{\boldsymbol{z}}$).

Similarly for $\boldsymbol{B}$:

$$(\boldsymbol{B}_0 + \boldsymbol{B}_0'') \cdot \hat{\boldsymbol{n}} = \boldsymbol{B}_0' \cdot \hat{\boldsymbol{n}} \tag{9.60}$$

$$\frac{1}{\mu}(\boldsymbol{B}_0 + \boldsymbol{B}_0'') \times \hat{\boldsymbol{n}} = \frac{1}{\mu'} \boldsymbol{B}_0' \times \hat{\boldsymbol{n}} \tag{9.61}$$

(where, recall, $\boldsymbol{B} = (\boldsymbol{k} \times \boldsymbol{E})/(vk)$ etc.) Again, one usually would not use this cross product algebraically, but would simply formulate the problem in a convenient coordinate system and take advantage of the fact that:

$$|\boldsymbol{B}_0| = \frac{|\boldsymbol{E}_0|}{v} = \sqrt{\mu\epsilon}|\boldsymbol{E}_0| \tag{9.62}$$

Coordinate choice and Brewster's Law

What, then, is a "convenient coordinate system"? One where $\hat{\boldsymbol{n}} = \hat{\boldsymbol{z}}$ is perpendicular to the surface is good for starters. The remaining two coordinates are selected to define the *plane of reflection and refraction* and its perpendicular. This is particularly useful because (as we shall see) the reflected and refracted intensities *depend on their polarization* relative to the plane of scattering.

Again, to motivate this before messing with the algebra, you hopefully are all familiar with the result taught at the kiddy-physics level known as

Brewster's Law. The argument works like this: because the refracted ray consists of (basically) dipole re-radiation of the incident field at the surface and because dipoles do not radiate along the direction of the dipole moment, the polarization component with $\boldsymbol{E}$ *in the scattering plane* has a component in this direction.

This leads to the insight that at certain angles the refracted ray will be completely polarized perpendicular to the scattering plane (Brewster's Law)! Our algebra needs to have this decomposition built in from the beginning or we'll have to work very hard indeed to obtain this as a result!

Let us therefore treat rays polarized in or perpendicular to the plane of incidence/reflection/refraction separately.

$\boldsymbol{E}$ Perpendicular to Plane of Incidence

The electric field in this case is perforce parallel to the surface and hence $\boldsymbol{E} \cdot \hat{\boldsymbol{n}} = 0$ and $|\boldsymbol{E} \times \hat{\boldsymbol{n}}| = 1$ (for incident, reflected and refracted waves). Only two of the four equations above are thus useful. The $\boldsymbol{E}$ equation is trivial. The $\boldsymbol{B}$ equation requires us to determine the magnitude of the cross product of $\boldsymbol{B}$ of each wave with $\hat{\boldsymbol{n}}$. Let's do one component as an example. Examining the triangle formed between $\boldsymbol{B}_0$ and $\hat{\boldsymbol{n}}$ for the incident waves (where θ_i is the angle of incidence and/or reflection, and θ_r is the angle of refraction), we see that:

$$\begin{aligned} \frac{1}{\mu}|\boldsymbol{B}_0 \times \hat{\boldsymbol{n}}| &= \frac{1}{\mu} B_0 \cos(\theta_i) \\ &= \frac{\sqrt{\mu\epsilon}}{\mu} E_0 \cos(\theta_i) \\ &= \sqrt{\frac{\epsilon}{\mu}} E_0 \cos(\theta_i) \end{aligned} \tag{9.63}$$

Repeating this for the other two waves and collecting the results, we obtain:

$$E_0 + E_0'' = E_0' \tag{9.64}$$

$$\sqrt{\frac{\epsilon}{\mu}}(E_0 - E_0'') \cos(\theta_i) = \sqrt{\frac{\epsilon'}{\mu'}} E_0' \cos(\theta_r) \tag{9.65}$$

This is two equations with two unknowns. Solving it is a bit tedious. We need:

$$\begin{aligned}
\cos(\theta_r) &= \sqrt{1-\sin^2(\theta_r)} && (9.66)\\
&= \sqrt{1-\frac{n^2}{n'^2}\sin^2(\theta_i)} && (9.67)\\
&= \frac{\sqrt{n'^2-n^2\sin^2(\theta_i)}}{n'} && (9.68)
\end{aligned}$$

Then we (say) eliminate E_0' using the first equation:

$$\sqrt{\frac{\epsilon}{\mu}}(E_0 - E_0'')\cos(\theta_i) = \sqrt{\frac{\epsilon'}{\mu'}}(E_0 + E_0'')\frac{\sqrt{n'^2-n^2\sin^2(\theta_i)}}{n'} \tag{9.69}$$

Collect all the terms:

$$\begin{aligned}
E_0\left(\sqrt{\frac{\epsilon}{\mu}}\cos(\theta_i) - \sqrt{\frac{\epsilon'}{\mu'}}\frac{\sqrt{n'^2-n^2\sin^2(\theta_i)}}{n'}\right) &= \\
E_0''\left(\sqrt{\frac{\epsilon'}{\mu'}}\frac{\sqrt{n'^2-n^2\sin^2(\theta_i)}}{n'} + \sqrt{\frac{\epsilon}{\mu}}\cos(\theta_i)\right) & \qquad (9.70)
\end{aligned}$$

Solve for E_0'':

$$E_0'' = E_0\frac{\left(\sqrt{\frac{\epsilon}{\mu}}\cos(\theta_i) - \sqrt{\frac{\epsilon'}{\mu'}}\frac{\sqrt{n'^2-n^2\sin^2(\theta_i)}}{n'}\right)}{\left(\sqrt{\frac{\epsilon}{\mu}}\cos(\theta_i) + \sqrt{\frac{\epsilon'}{\mu'}}\frac{\sqrt{n'^2-n^2\sin^2(\theta_i)}}{n'}\right)} \tag{9.71}$$

This expression can be simplified after some tedious cancellations involving

$$\frac{n}{n'} = \sqrt{\frac{\mu\epsilon}{\mu'\epsilon'}} \tag{9.72}$$

and either repeating the process or back-substituting to obtain :

$$\begin{aligned}
E_0'' &= E_0\frac{\left(n\cos(\theta_i) - \frac{\mu}{\mu'}\sqrt{n'^2-n^2\sin^2(\theta_i)}\right)}{\left(n\cos(\theta_i) + \frac{\mu}{\mu'}\sqrt{n'^2-n^2\sin^2(\theta_i)}\right)} && (9.73)\\
E_0' &= E_0\frac{2n\cos(\theta_i)}{\left(n\cos(\theta_i) + \frac{\mu}{\mu'}\sqrt{n'^2-n^2\sin^2(\theta_i)}\right)} && (9.74)
\end{aligned}$$

$\boldsymbol{E}$ Parallel to Plane of Incidence

Now the *magnetic* field is parallel to the surface so $\boldsymbol{B}\cdot\hat{\boldsymbol{n}} = 0$ and $|\boldsymbol{B}\times\hat{\boldsymbol{n}}| = 1$. This time *three* equations survive, but they cannot all be independent as we have only two unknowns (given Snell's law above for the reflected/refracted waves). We might as well use the simplest possible forms, which are clearly the ones where we've already worked out the geometry, e.g. $\boldsymbol{E}_0 \times \hat{\boldsymbol{n}} = E_0 \cos(\theta_i)$ (as before for $\boldsymbol{B}_0$). The two simplest ones are clearly:

$$\begin{aligned}
(E_0 - E_0'')\cos(\theta_i) &= E_0'\cos(\theta_r) &\quad (9.75)\\
\sqrt{\frac{\epsilon}{\mu}}(E_0 + E_0'') &= \sqrt{\frac{\epsilon'}{\mu'}}E_0' &\quad (9.76)
\end{aligned}$$

(from the second matching equations for both $\boldsymbol{E}$ and $\boldsymbol{B}$ above).

It is left as a moderately tedious exercise to repeat the reasoning process for these two equations – eliminate either E_0' or E_0'' and solve/simplify for the other, repeat or backsubstitute to obtain the originally eliminated one (or use your *own* favorite way of algebraically solving simultaneous equations) to obtain:

$$\begin{aligned}
E_0' &= E_0 \frac{2nn'\cos(\theta_i)}{\frac{\mu}{\mu'}n'^2\cos(\theta_i) + n\sqrt{n'^2 - n^2\sin^2(\theta_i)}} &\quad (9.77)\\
E_0'' &= E_0 \frac{\frac{\mu}{\mu'}n'^2\cos(\theta_i) - n\sqrt{n'^2 - n^2\sin^2(\theta_i)}}{\frac{\mu}{\mu'}n'^2\cos(\theta_i) + n\sqrt{n'^2 - n^2\sin^2(\theta_i)}} &\quad (9.78)
\end{aligned}$$

The last result that one should note before moving on is the important case of *normal incidence* (where $\cos\theta_i = 1$ and $\sin(\theta_i) = 0$). Now there should only be perpendicular solutions. Interestingly, either the parallel or perpendicular solutions above simplify with obvious cancellations and tedious eliminations to:

$$\begin{aligned}
E_0' &= E_0 \frac{2n}{n' + n} &\quad (9.79)\\
E_0'' &= E_0 \frac{n' - n}{n' + n} &\quad (9.80)
\end{aligned}$$

Note well that the reflected wave changes *phase* (is negative relative to the incident wave in the plane of scattering) if $n > n'$. This of course makes sense – there are many intuitive reasons to expect a wave to invert its phase when reflecting from a "heavier" medium starting with things one learns studying wave pulses on a string. If this doesn't make sense to you please ask for help.

Intensity

Without wanting to get all tedious about it, you should be able to compute the *transmission coefficient* and *reflection coefficient* for all of these waves from these results. These are basically the fraction of the energy (per unit area per unit time) in the incident wave that is transmitted vs being reflected by the surface.

This is a simple idea, but it is a bit tricky to actually compute for a couple of reasons. One is that we only care about energy that makes it *through* the surface. The directed intensity of the wave (energy per unit area per unit time) is the Poynting vector $\boldsymbol{S}$. We therefore have to compute the magnitude of the energy flux through the surface $\boldsymbol{S} \cdot \hat{\boldsymbol{n}}$ for each wave:

$$S_n = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}|E_0|^2 \cos(\theta_i) \tag{9.81}$$

$$S'_n = \frac{1}{2}\sqrt{\frac{\epsilon'}{\mu'}}|E'_0|^2 \cos(\theta_r) \tag{9.82}$$

$$S''_n = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}|E''_0|^2 \cos(\theta_i) \tag{9.83}$$

$$\tag{9.84}$$

This is easy only if the waves are incident $\perp$ to the surface, in which case one gets:

$$T = \frac{I'_0}{I_0} = \sqrt{\frac{\epsilon'\mu}{\epsilon\mu'}}\frac{|E'_0|^2}{|E_0|^2} \tag{9.85}$$

$$= \frac{4nn'}{(n'+n)^2} \tag{9.86}$$

$$R = \frac{I''_0}{I_0} = \frac{|E''_0|^2}{|E_0|^2} \tag{9.87}$$

$$= \frac{(n'-n)^2}{(n'+n)^2} \tag{9.88}$$

In this case it is easy to verify that $T + R = 1$ as it should.

Polarization Revisited: The Brewster Angle

Note well the expression for the reflected wave amplitude for in-plane polarization:

$$E_0'' = E_0 \frac{\frac{\mu}{\mu'} n'^2 \cos(\theta_i) - n\sqrt{n'^2 - n^2 \sin^2(\theta_i)}}{\frac{\mu}{\mu'} n'^2 \cos(\theta_i) + n\sqrt{n'^2 - n^2 \sin^2(\theta_i)}} \tag{9.89}$$

This amplitude will be *zero* for certain angles, namely those such that:

$$\frac{\mu}{\mu'} n'^2 \cos(\theta_i) = n\sqrt{n'^2 - n^2 \sin^2(\theta_i)} \tag{9.90}$$

Squaring both sides and restoring cosine term to its original form:

$$\left(\frac{\mu}{\mu'}\right)^2 n'^2 \cos^2(\theta_i) = n^2 \cos^2(\theta_r) \tag{9.91}$$

We therefore expect the reflected wave to *vanish* when

$$\frac{\mu n'}{\mu' n} = \frac{\cos(\theta_r)}{\cos(\theta_i)} \tag{9.92}$$

For optical frequencies $\mu \approx \mu'$ (to simplify the algebra somewhat) and this is equivalent to:

$$n' \cos(\theta_i) = n \cos(\theta_r) \tag{9.93}$$

From Snell's law this in turn is:

$$\frac{n}{n'} \tan(\theta_i) = \frac{n'}{n} \tan(\theta_r) \tag{9.94}$$

This trancendental equation can be solved by observation from its symmetry. It is true if and only if:

$$\tan(\theta_i) = \frac{n'}{n} = \cot(\theta_r) \tag{9.95}$$

The angle of incidence

$$\theta_b = \tan^{-1}\left(\frac{n'}{n}\right) \tag{9.96}$$

is called *Brewster's angle*. At this reflected and refracted wave travel at *right angles* with respect to one another according to Snell's law. This means that the dipoles in the second medium that are *responsible* for the reflected wave are *parallel* to the direction of propagation and (as we shall see) oscillating dipoles to *not radiate* in the direction of their dipole moment! However, the result above was obtained without any appeal to the microscopic properties of the dielectric moments that actually coherently scatter the incident wave at the surface – it follows strictly as the result of solving a boundary value problem for electromagnetic plane waves.

Students interested in optical fibers are encouraged to read further in Jackson, 7.4 and learn how the cancellation and reradiation of the waves to produce a reflected wave at angles where total internal reflection happens does not occur instantaneously at the refracting surface but in fact involves the penetration of the second medium some small distance by non-propagating fields. This in turn is related to polarization, dispersion, and skin depth, which we will now treat in some detail.

9.3 Dispersion

Up to now, we have obtained all of our results with the *assumption* that the medium was *free from dispersion*. This just meant that we assumed that the index of refraction was *constant* as a function of frequency, so all wavelengths were similarly affected. Of course none of our results dependent particular strongly on this result, but in any event it *is not correct*. The permittivity (and to a lesser extent for transparent materials, the permeability) is a *function of the frequency* and thus the speed of light changes as we pass from one dielectric medium to another. Let us now figure out what and how dispersion works.

By the way, when I say that it "isn't correct" I'm not asserting an opinion or mathematical conclusion. That's not how physics works. Rather it is *always* ultimately empirical: rainbows and prisms (not to mention opaque objects) remind us that most physical media are *not* free from dispersion. Understanding and modelling the *dynamics* of dispersion in a way that correctly explains these observed phenomena is a key step in the understanding of much modern physics, which involves the measurement and prediction of various *susceptibilities* (another form of the permittivity, basically, as you can see below) in both classical and quantum circumstances. A full understanding of the *particular* dispersion of a physical medium is possible only in the context of quantum theory, but to understand the phenomenon *itself* we can fortunately rely on a rather simple model that exhibits all the essential features observed in macroscopic media.

9.3.1 Static Case

Recall, (from sections 4.5 and 4.6) that when the electric field penetrates a medium made of bound charges, it *polarizes* those charges. The charges themselves then produce a field that opposes, and hence by superposition reduces, the applied field. The key assumption in these sections was that the polarization of the medium was a *linear function* of the total field in the vicinity of the atoms.

Linearity response was easily modelled by assuming a harmonic (linear) *restoring force*:

$$\boldsymbol{F} = -m\omega_0^2\boldsymbol{x} \tag{9.97}$$

acting to pull a charge e from a neutral equilibrium in the presence of an electric field:

$$m\omega_0^2\boldsymbol{x} = e\boldsymbol{E} \tag{9.98}$$

where $\boldsymbol{E}$ is the applied external field. The dipole moment of this (presumed) system is $\boldsymbol{p}_{\text{mol}} = e\boldsymbol{x} = \frac{e^2}{m\omega_0^2}\boldsymbol{E}$. Real molecules, of course, have *many* bound charges, each of which at equilibrium has an approximately linear restoring force with its own natural frequency, so a more general model of molecular polarizability is:

$$\gamma_{\text{mol}} = \frac{1}{\epsilon_0}\sum_i \frac{e_i^2}{m_i\omega_i^2} \tag{9.99}$$

From the linear approximation you obtained an equation for the total *polarization* (dipole moment per unit volume) of the material:

$$\boldsymbol{P} = N\gamma_{\text{mol}}\left(\epsilon_0\boldsymbol{E} + \frac{1}{3}\boldsymbol{P}\right) \tag{9.100}$$

(equation 4.68). This can be put in many forms. For example, using the definition of the (dimensionless) *electric susceptibility*:

$$\boldsymbol{P} = \epsilon_0\chi_e\boldsymbol{E} \tag{9.101}$$

we find that:

$$\chi_e = \frac{N\gamma_{\rm mol}}{1 - \frac{N\gamma_{\rm mol}}{3}}. \tag{9.102}$$

The susceptibility is one of the most often measured or discussed quantities of physical media in many contexts of physics.

However, as we've just seen, in the context of waves we will most often have occasion to use polarizability in terms of the *permittivity* of the medium, ϵ. In term of χ_e, this is:

$$\epsilon = \epsilon_0(1 + \chi_e) \tag{9.103}$$

From a knowledge of ϵ (in the regime of optical frequencies where $\mu \approx \mu_0$ for many materials of interest) we can easily obtain, e. g. the index of refraction:

$$n = \frac{c}{v} = \frac{\sqrt{\mu\epsilon}}{\sqrt{\mu_0\epsilon_0}} \approx \sqrt{\frac{\epsilon}{\epsilon_0}} \approx \sqrt{1 + \chi_e} \tag{9.104}$$

or

$$n = \sqrt{\frac{1 + \frac{2N\gamma_{\rm mol}}{3}}{1 - \frac{N\gamma_{\rm mol}}{3}}} \tag{9.105}$$

if N and $\gamma_{\rm mol}$ are known or at least approximately computable using the (surprisingly accurate) expression above.

So much for static polarizability of insulators – it is readily understandable in terms of real physics of pushes and pulls, and the semi-quantitative models one uses to understand it work quite well. However, *real* fields aren't static, and *real materials* aren't all insulators. So we gotta

1. Modify the model to make it *dynamic*.

2. Evaluate the model (more or less as above, but we'll have to work harder.

3. Understand what's going on.

Let's get started.

9.3.2 Dynamic Case

The obvious generalization of the static model for the polarization is to assume a *damped linear response* to a *harmonic* (plane wave) driving electric field. That is, every molecule will be viewed as a collection of damped, driven (charged) harmonic oscillators. Magnetic and non–linear effects will be neglected. This is valid for a variety of materials subjected to "weak" harmonic EM fields[6] which in practice (with optical frequencies) means nearly everything but laser light.

The equation of motion[7] for a single damped, driven harmonically bound charged electron is:

$$m\left[\ddot{\boldsymbol{x}} + \gamma\dot{\boldsymbol{x}} + \omega_0^2\boldsymbol{x}\right] = -eE(\boldsymbol{x}, t) \tag{9.106}$$

where γ is the damping constant (so $-m\gamma\dot{\boldsymbol{x}}$ is the velocity dependent damping force). If we assume that the electric field $\boldsymbol{E}$ and $\boldsymbol{x}$ are harmonic in time at frequency ω (or fourier transform the equation and find its solution for a single fourier component) and neglect the transients we get:

$$\boldsymbol{p} = -e\boldsymbol{x} = \frac{e^2}{m}\frac{\boldsymbol{E}_\omega}{(\omega_0^2 - \omega^2 - i\omega\gamma)} \tag{9.107}$$

for each electron[8].

Actually, we have N molecules/unit volume each with Z electrons where f_i of them have frequencies and damping constants ω_i and γ_i, respectively (whew!) then (since we will stick in the definitions $\boldsymbol{P}_\omega = \epsilon_0\chi_e\boldsymbol{E}_\omega$ and $\epsilon = 1 + \chi_e$)

$$\epsilon(\omega) = \epsilon_0\left(1 + \frac{Ne^2}{m}\sum_i \frac{f_i}{(\omega_i^2 - \omega^2 - i\omega\gamma_i)}\right) \tag{9.108}$$

where the *oscillator strengths* satisfy the sum rule:

$$\sum_i f_i = Z. \tag{9.109}$$

These equations (within suitable approximations) are valid for quantum theories, and indeed, since quantum oscillators have certain discrete frequencies, they seem to "naturally" be quantum mechanical.

[6]Why? If you don't understand this, you need to go back to basics and think about expanding a potential well in a Taylor series about a particle's equilibrium position. The linear term vanishes because it is equilibrium, so the first surviving term is likely to be quadratic. Which is to say, proportional to x^2 where x is the displacement from equilibrium, corresponding to a linear restoring force to lowest order.

[7]You do remember Newton's law, don't you? Sure hope so...

[8]I certainly hope you can derive this result, at least if your life depends on it. In qualifiers, while teaching kiddy physics, whenever.

9.3.3 Things to Note

Before we go on, we should understand a few things:

1. ϵ is now *complex!* The imaginary part is explicitly connected to the damping constant.

2. Consequently we can now see how the index of refraction

$$n = \frac{c}{v} = \frac{\sqrt{\mu\epsilon}}{\sqrt{\mu_0\epsilon_0}}, \qquad (9.110)$$

   can be also be *complex*. A complex index of refraction describes *absorption* (or amplification!) and arises from the *damping* term in the electrons' EOM (or non–linear, non–equilibrium effects in lasers, which we will *not* consider here). This makes energy conservation kind of sense. Energy absorbed by the electrons and *dissipated* via the "frictional" damping force is removed from the EM field as it propagates through the medium. This (complex dispersion of incident waves) is the basis for the "optical" description of scattering which is useful to nuclear physicists.

3. The term

$$\frac{1}{\omega_i^2 - \omega^2 - i\omega\gamma}$$

   has a form that you will see again and again and again in your studies. It should be meditated upon, studied, dreamed about, mentally masticated and enfolded into your beings until you *understand* it. It is a complex equation with *poles* in the imaginary/real plane. It describes (very generally speaking) *resonances.*

   It is useful to convert this into a form which has manifest real and imaginary parts, since we will have occasion to compute them in real problems one day. A bit of algebra gives us:

$$\frac{1}{\omega_i^2 - \omega^2 - i\omega\gamma} = \frac{(\omega_i^2 - \omega^2) + i\omega\gamma}{(\omega_i^2 - \omega^2)^2 + \omega^2\gamma^2}$$

4. If N is "small" ($\sim 10^{19}$ molecules/cc for a gas) χ_e is small (just like in the static case) and the medium is nearly transparent at most frequencies.

5. if N is "large" ($\sim 10^{23}$ molecules/cc for a liquid or solid) χ_e can be quite large in principle, and near a resonance can be quite large and *complex!*

These points and more require a new language for their convenient description. We will now pause a moment to develop one.

9.3.4 Anomalous Dispersion, and Resonant Absorption

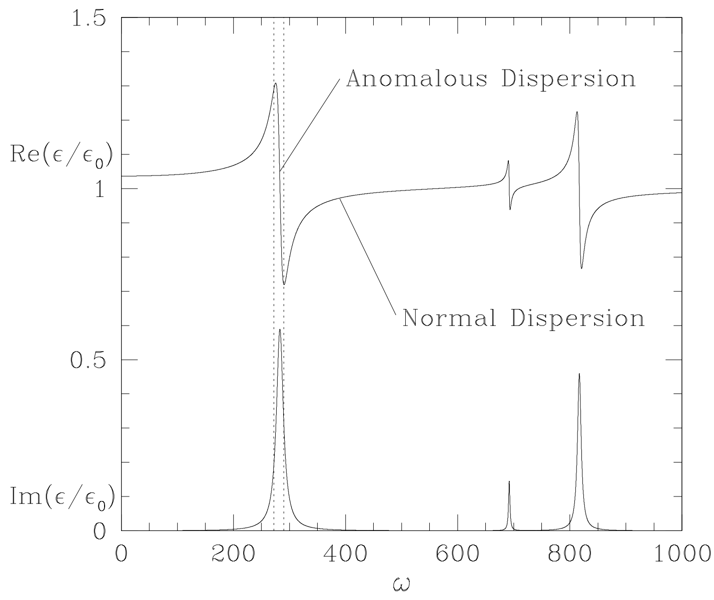

Figure 9.1: Typical curves indicating the real and imaginary parts of ϵ/ϵ_0 for an atom with three visible resonances. Note the regions of *anomalous* (descending) real dispersion in the immediate vicinity of the *resonances*, separated by large regions of *normal* (ascending) dispersion.

The γ_i are typically small compared to the oscillator frequencies ω_i. (Just to give you an idea, $\gamma_i \sim 10^9$ sec^{-1} to $\omega_i \sim 10^{15}$ sec^{-1} for optical transitions in atoms, with similar proportionalities for the other relevant transitions.) That means that at most frequencies, $\epsilon(\omega)$ is nearly real

Suppose we only have a few frequencies. Below the smallest ω_i, all the (real) terms in the sum are positive and Re $\epsilon(\omega) > 1$. As we increase ω, one by one the terms in the sum become negative (in their real part) until beyond the highest frequency the entire sum and hence Re $\epsilon(\omega) < 1$.

As we sweep past each "pole" (where the real part in the denominator of a single term is zero) that term *increases rapidly* in the real part, then dives through zero to become large and negative, then increases monotonically to zero. Meanwhile, its (usually small) imaginary part grows, reaching a peak just where the real part is zero (when $\epsilon(\omega)$ is pure imaginary). In the vicinity of the pole, the contribution of this term can dominate the rest of the sum. We define:

Normal dispersion as strictly *increasing* Re $\epsilon(\omega)$ with increasing ω. This is the normal situation everywhere but near a pole.

Anomalous dispersion as *decreasing* Re $\epsilon(\omega)$ with increasing ω. This is true only near a sufficiently strong pole (one that dominates the sum). At that point, the imaginary part of the index of refraction becomes (relatively) appreciable.

Resonant Absorption occurs in the regions where Im ϵ is large. We will parametrically describe this next.

9.3.5 Attenuation by a complex ϵ

Suppose we write (for a given frequency)

$$k = \beta + i\frac{\alpha}{2}. \tag{9.111}$$

Then

$$\boldsymbol{E}_\omega(\boldsymbol{x}) = e^{ikx} = e^{i\beta x} e^{-\frac{\alpha}{2}x} \tag{9.112}$$

and the *intensity* of the (plane) wave falls off like $e^{-\alpha x}$. α measures the damping of the plane wave in the medium.

Let's think a bit about k:

$$k = \frac{\omega}{v} = \frac{\omega}{c} n \tag{9.113}$$

where:

$$n = c/v = \frac{\sqrt{\mu\epsilon}}{\sqrt{\mu_0\epsilon_0}} \tag{9.114}$$

In most "transparent" materials, $\mu \approx \mu_0$ and this simplifies to $n = \sqrt{\epsilon/\epsilon_0}$. Thus:

$$k^2 = \frac{\omega^2}{c^2}\frac{\epsilon}{\epsilon_0} \tag{9.115}$$

Nowever, *now* ϵ has real and imaginary parts, so k may as well! In fact, using the expression for k in terms of β and α above, it is easy to see that:

$$\text{Re } k^2 = \beta^2 - \frac{\alpha^2}{4} = \frac{\omega^2}{c^2}\text{Re } \frac{\epsilon}{\epsilon_0} \tag{9.116}$$

and

$$\text{Im } k^2 = \beta\alpha = \frac{\omega^2}{c^2}\text{Im } \frac{\epsilon}{\epsilon_0}. \tag{9.117}$$

As long as $\beta^2 >> \alpha^2$ (again, true most of the time in trasparent materials) we can thus write:

$$\frac{\alpha}{\beta} \approx \frac{\text{Im } \epsilon(\omega)}{\text{Re } \epsilon(\omega)} \tag{9.118}$$

and

$$\beta \approx (\omega/c)\sqrt{\text{Re } \frac{\epsilon}{\epsilon_0}} \tag{9.119}$$

This ratio can be interpreted as a quantity similar to Q, the fractional decrease in intensity per wavelength travelled through the medium (as opposed to the fractional decrease in intensity per period).

To find α in some useful form, we have to examine the details of $\epsilon(\omega)$, which we will proceed to do next.

When ω is in among the resonances, there is little we can do besides work out the details of the behavior, since the properties of the material can be dominated strongly by the *local* dynamics associated with the nearest, strongest resonance. However, there are two limits that are of particular interest to physicists where the "resonant" behavior can be either evaluated or washed away. They are the *low frequency behavior* which determines the conduction properties of a material far away from the electron resonances per se, and the *high frequency behavior* which is "universal".

9.3.6 Low Frequency Behavior

Near $\omega = 0$ the qualitative behavior depends upon whether or not there is a "resonance" there. If there is, then $\epsilon(\omega \approx 0)$ can begin with a complex component that attenuates the propagation of EM energy in a (nearly static) applied electric field. This (as we shall see) accurately describes *conduction* and *resistance*. If there isn't, then ϵ is nearly all real and the material is a dielectric insulator.

Suppose there are both "free" electrons (counted by f_f) that are "resonant" at zero frequency, and "bound" electrons (counted by f_b). Then if we start out with:

$$\begin{aligned}
\epsilon(\omega) &= \epsilon_0 \left(1 + \frac{Ne^2}{m}\sum_i \frac{f_i}{(\omega_i^2 - \omega^2 - i\omega\gamma_i)}\right) \\
&= \epsilon_0 \left(1 + \frac{Ne^2}{m}\sum_b \frac{f_b}{(\omega_b^2 - \omega^2 - i\omega\gamma_b)}\right) \\
&\qquad + \frac{Ne^2}{m}\sum_f \frac{f_f}{(-\omega^2 - i\omega\gamma_f)} \\
&= \epsilon_b + i\epsilon_0 \frac{Ne^2 f_f}{m\omega(\gamma_0 - i\omega)} \qquad (9.120)
\end{aligned}$$

where ϵ_b is now *only* the contribution from all the "bound" dipoles.

We can understand this from

$$\boldsymbol{\nabla} \times \boldsymbol{H} = \boldsymbol{J} + \frac{d\boldsymbol{D}}{dt} \qquad (9.121)$$

(Maxwell/Ampere's Law). Let's first of all think of this in terms of a plain old static current, sustained according to Ohm's Law:

$$\boldsymbol{J} = \sigma \boldsymbol{E}. \qquad (9.122)$$

If we assume a harmonic time dependence and a "normal" dielectric constant ϵ_b, we get:

$$\begin{aligned}
\boldsymbol{\nabla} \times \boldsymbol{H} &= (\sigma - i\omega\epsilon_b)\,\boldsymbol{E} \\
&= -i\omega\left(\epsilon_b + i\frac{\sigma}{\omega}\right)\boldsymbol{E}. \qquad (9.123)
\end{aligned}$$

On the other hand, we can instead set the static current to *zero* and consider all "currents" present to be the *result* of the polarization response

$\boldsymbol{D}$ to the field $\boldsymbol{E}$. In this case:

$$\begin{aligned}\boldsymbol{\nabla}\times\boldsymbol{H} &= -i\omega\epsilon\boldsymbol{E}\\ &= -i\omega\left(\epsilon_b + i\epsilon_0\frac{Ne^2}{m}\frac{f_f}{(\gamma_0 - i\omega)}\right)\boldsymbol{E}\end{aligned}\tag{9.124}$$

Equating the two latter terms in the brackets and simplifying, we obtain the following relation for the conductivity:

$$\sigma = \epsilon_0\,\frac{n_f e^2}{m}\frac{1}{(\gamma_0 - i\omega)}.\tag{9.125}$$

This is the **Drude Model** with $n_f = f_f N$ the number of "free" electrons per unit volume. It is primarily useful for the insight that it gives us concerning the "conductivity" being closely related to the zero-frequency complex part of the permittivity. Note that at $\omega = 0$ it is purely real, as it should be, recovering the usual Ohm's Law.

We conclude that the distinction between dielectrics and conductors is a matter of perspective away from the purely static case. Away from the static case, "conductivity" is simply a feature of resonant amplitudes. It is a matter of taste whether a description is better made in terms of dielectric constants and conductivity or complex dielectric.

9.3.7 High Frequency Limit; Plasma Frequency

Way *above* the highest resonant frequency the dielectric constant takes on a simple form (factoring out $\omega >> \omega_i$ and doing the sum to the lowest surviving order in ω_p/ω. As before, we start out with:

$$\begin{aligned}\epsilon(\omega) &= \epsilon_0\left(1 + \frac{Ne^2}{m}\sum_i\frac{f_i}{(\omega_i^2 - \omega^2 - i\omega\gamma_i)}\right)\\ &= \epsilon_0\left(1 - \frac{Ne^2}{\omega^2 m}\sum_i\frac{f_i}{(1 + i\frac{\gamma_i}{\omega} - \frac{\omega_i^2}{\omega^2})}\right)\\ &\approx \epsilon_0\left(1 - \frac{NZe^2}{\omega^2 m}\right)\\ &\approx \epsilon_0\left(1 - \frac{\omega_p^2}{\omega^2}\right)\end{aligned}\tag{9.126}$$

where

$$\omega_p^2 = \frac{ne^2}{m}. \tag{9.127}$$

This is called the **plasma frequency**, and it depends only on $n = NZ$, the total number of electrons per unit volume.

The wave number in this limit is given by:

$$ck = \sqrt{\omega^2 - \omega_p^2} \tag{9.128}$$

(or $\omega^2 = \omega_p^2 + c^2k^2$). This is called a **dispersion relation** $\omega(k)$. A large portion of contemporary and famous physics involves calculating dispersion relations (or equivalently susceptibilities, right?) from first principles.

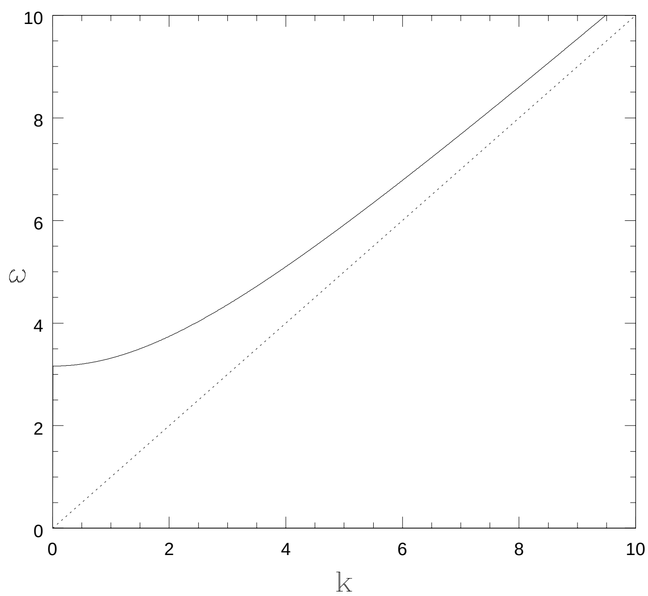

Figure 9.2: The dispersion relation for a plasma. Features to note: Gap at $k = 0$, asymptoticallyx linear behavior.

The dielectric constant is essentially real and slightly less than one.

In certain physical situations (such as a plasma or the ionosphere) all the electrons are essentially "free" (in a degenerate "gas" surrounding the positive charges) and resonant damping is neglible. In that case this relation can hold for frequencies well below ω_p (but well above the static limit, since plasmas are low frequency "conductors"). Waves incident on a plasma are reflected and the fields inside fall off exponentially away from the surface. Note that

$$\alpha_p \approx \frac{2\omega_p}{c} \tag{9.129}$$

shows how electric flux is expelled by the "screening" electrons.

The reflectivity of metals is caused by essentially the same mechanism. At high frequencies, the dielectric constant of a metal has the form

$$\epsilon(\omega) \approx \epsilon_0(\omega) - \frac{\omega_p^2}{\omega^2} \tag{9.130}$$

where $\omega_p^2 = ne^2/m^*$ is the "plasma frequency" of the conduction electrons. m^* is the "effective mass" of the electrons, introduced to describe the effects of binding phenomenologically.

Metals reflect according to this rule (with a very small field penetration length of "skin depth") as long as the dielectric constant is negative; in the ultraviolet it becomes positive and metals can become transparent. Just one of many problems involved in making high ultraviolet, x–ray and gamma ray lasers — it is so hard to make a mirror!

9.4 Penetration of Waves Into a Conductor – Skin Depth

9.4.1 Wave Attenuation in Two Limits

Recall from above that:

$$\boldsymbol{\nabla} \times \boldsymbol{H} = -i\omega\epsilon\boldsymbol{E} = -i\omega\left(\epsilon_b + i\frac{\sigma}{\omega}\right)\boldsymbol{E}. \tag{9.131}$$

Then:

$$k^2 = \frac{\omega^2}{v^2} = \mu\epsilon\omega^2 = \mu\epsilon_b\omega^2\left(1 + i\frac{\sigma}{\omega\epsilon_b}\right) \tag{9.132}$$

Also $k = \beta + i\frac{\alpha}{2}$ so that

$$k^2 = \left(\beta^2 - \frac{\alpha^2}{4}\right) + i\alpha\beta = \mu\epsilon_b\omega^2\left(1 + i\frac{\sigma}{\omega\epsilon_b}\right) \tag{9.133}$$

Oops. To determine α and β, we have to take the square root of a complex number. How does that work again? See the appendix on *Complex Numbers...*

In many cases we can pick the right branch by selecting the one with the right (desired) behavior on physical grounds. If we restrict ourselves to the two simple cases where ω is large or σ is large, it is the one in the principle branch (upper half plane, above a branch cut along the real axis. From the last equation above, if we have a poor conductor (or if the frequency is much higher than the plasma frequency) and $\alpha \ll \beta$, then:

$$\beta \approx \sqrt{\mu\epsilon_b}\omega \tag{9.134}$$

$$\alpha \approx \sqrt{\frac{\mu}{\epsilon_b}}\sigma \tag{9.135}$$

and the attenuation (recall that $\boldsymbol{E} = \boldsymbol{E}_0 e^{-\frac{\alpha}{2}} e^{i\beta\hat{\boldsymbol{n}}\cdot\boldsymbol{E}}$) is independent of frequency.

The other limit that is relatively easy is a *good* conductor, $\sigma \gg \omega\epsilon_b$. In that case the *imaginary* term dominates and we see that

$$\beta \approx \frac{\alpha}{2} \tag{9.136}$$

or

$$\beta \approx \sqrt{\frac{\mu\sigma\omega}{2}} \tag{9.137}$$

$$\alpha \approx \sqrt{2\mu\sigma\omega} \tag{9.138}$$

Thus

$$k = (1+i)\sqrt{\frac{\mu\sigma\omega}{2}} \tag{9.139}$$

Recall that if we apply the $\boldsymbol{\nabla}$ operator to $\boldsymbol{E}e^{ik(\boldsymbol{n}\cdot\boldsymbol{x}-i\omega t}$ we get:

$$\begin{aligned} \boldsymbol{\nabla}\cdot\boldsymbol{E} &= 0 \\ ik\boldsymbol{E}_0\cdot\hat{\boldsymbol{n}} &= 0 \\ \boldsymbol{E}_0\cdot\hat{\boldsymbol{n}} &= 0 \end{aligned} \tag{9.140}$$

and

$$\begin{aligned}
-\frac{\partial \boldsymbol{B}}{\partial t} &= \boldsymbol{\nabla} \times \vec{E} \\
i\omega\mu \boldsymbol{H}_0 &= i(\hat{\boldsymbol{n}} \times \boldsymbol{E}_0)(1+i)\sqrt{\frac{\mu\sigma\omega}{2}} \\
\boldsymbol{H}_0 &= \frac{1}{\omega}\sqrt{\frac{\sigma\omega}{\mu}}(\hat{\boldsymbol{n}} \times \boldsymbol{E}_0)\frac{1}{\sqrt{2}}(1+i) \\
&= \frac{1}{\omega}\sqrt{\frac{\sigma\omega}{\mu}}(\hat{\boldsymbol{n}} \times \boldsymbol{E}_0)e^{i\pi/4}
\end{aligned} \tag{9.141}$$

so $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$ are *not in phase* (using the fact that $i = e^{i\pi/2}$).

In the case of superconductors, $\sigma \to \infty$ and the phase angle between them is $\pi/4$. In this case $\boldsymbol{H}_0 \gg \boldsymbol{E}$ (show this!) and the energy is mostly *magnetic*.

Finally, note well that the quantity $\left(\frac{\alpha}{2}\right)^{-1} = \delta$ is an *exponential damping length* that describes how rapidly the wave attenuates as it moves into the conducting medium. δ is called the *skin depth* and we see that:

$$\delta = \frac{2}{\alpha} = \frac{1}{\beta} = \sqrt{\frac{2}{\mu\sigma\omega}} \tag{9.142}$$

We will examine this quantity in some detail in the sections on waveguides and optical cavities, where it plays an important role.

9.5 Kramers-Kronig Relations

We find KK relations by playing looped games with Fourier Transforms. We begin with the relation between the electric field and displacement at some particular frequency ω:

$$\boldsymbol{D}(\boldsymbol{x},\omega) = \epsilon(\omega)\boldsymbol{E}(\boldsymbol{x},\omega) \tag{9.143}$$

where we note the two (forward and backward) fourier transform relations:

$$\boldsymbol{D}(\boldsymbol{x},t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \boldsymbol{D}(\boldsymbol{x},\omega)e^{-i\omega t}d\omega \tag{9.144}$$

$$\boldsymbol{D}(\boldsymbol{x},\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \boldsymbol{D}(\boldsymbol{x},t')e^{i\omega t'}dt' \tag{9.145}$$

and of course:

$$\boldsymbol{E}(\boldsymbol{x},t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\boldsymbol{E}(\boldsymbol{x},\omega)e^{-i\omega t}d\omega \tag{9.146}$$

$$\boldsymbol{E}(\boldsymbol{x},\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\boldsymbol{E}(\boldsymbol{x},t')e^{i\omega t'}dt' \tag{9.147}$$

Therefore:

$$\begin{aligned}
\boldsymbol{D}(\boldsymbol{x},t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\epsilon(\omega)\boldsymbol{E}(\boldsymbol{x},\omega)e^{-i\omega t}d\omega \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\epsilon(\omega)e^{-i\omega t}d\omega\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\boldsymbol{E}(\boldsymbol{x},t')e^{i\omega t'}dt' \\
&= \epsilon_0\left\{\boldsymbol{E}(\boldsymbol{x},t) + \int_{-\infty}^{\infty}G(\tau)\boldsymbol{E}(\boldsymbol{x},t-\tau)d\tau\right\}
\end{aligned} \tag{9.148}$$

where we have introduced the *susceptibility kernel*:

$$G(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{\frac{\epsilon(\omega)}{\epsilon_0} - 1\right\}e^{-i\omega\tau}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\chi_e(\omega)e^{-i\omega\tau}d\omega \tag{9.149}$$

(noting that $\epsilon(\omega) = \epsilon_0(1+\chi_e(\omega))$). This equation is nonlocal in time unless $G(\tau)$ is a delta function, which in turn is true only if the dispersion is constant.

To understand this, consider the susceptibility kernel for a simple one resonance model (more resonances are just superposition). In this case, recall that:

$$\chi_e = \frac{\epsilon}{\epsilon_0} - 1 = \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma_0\omega} \tag{9.150}$$

so

$$G(\tau) = \frac{\omega_p^2}{2\pi}\int_{-\infty}^{\infty}\frac{1}{\omega_0^2 - \omega^2 - i\gamma_0\omega}e^{-i\omega\tau}d\omega \tag{9.151}$$

This is an integral we can do using contour integration methods. We use the quadratic formula to find the roots of the denominator, then write the factored denominator in terms of the roots:

$$\omega_{1,2} = \frac{-i\gamma \pm \sqrt{-\gamma^2 + 4\omega_0^2}}{2} \tag{9.152}$$

or

$$\omega_{1,2} = \frac{-i\gamma}{2} \pm \omega_0\sqrt{1 - \frac{\gamma^2}{4\omega_0^2}} = \frac{-i\gamma}{2} \pm \nu_0 \tag{9.153}$$

where $\nu_0 \approx \omega_0$ as long as $\omega_0 \gg \gamma/2$ (as is usually the case, remember β and $\alpha/2$). Note that these poles are in the *lower half plane* (LHP) because of the sign of γ in the original harmonic oscillator – it was *dissipative*. This is important.

Then

$$G(\tau) = (2\pi i)\frac{\omega_p^2}{2\pi}\oint_C \frac{1}{(\omega-\omega_1)(\omega-\omega_2)}e^{-i\omega\tau}d\omega \tag{9.154}$$

If we close the contour in the upper half plane (UHP), we have to restrict $\tau < 0$ (why? because otherwise the integrand will not vanish on the contour at infinity where ω has a positive imaginary part. Since it encloses no poles, $G(\tau < 0)$ vanishes, and we get *no contribution from the future* in the integral above for $\boldsymbol{E}$. The result appears to be *causal*, but really we cheated – the "causality" results from the *damping* term, which represents *entropy* and yeah, gives time an arrow here. But it doesn't really break the symmetry of time in this problem and if our model involved a dynamically pumped medium so that the wave experienced *gain* moving through it (an imaginary term that was *positive*) we would have had poles in the UHP and our expression for $\boldsymbol{E}$ would *not* be "causal". Really it is equally causal in both cases, because the fourier transforms involved sample all times anyway.

If we close the integrand in the LHP, $\tau > 0$ and if we do the rest of the (fairly straightforward) algebra we get:

$$G(\tau) = \omega_p^2 e^{-\frac{\gamma\tau}{2}}\frac{\sin(\nu_0)}{\nu_0}\Theta(\tau) \tag{9.155}$$

where the latter is a Heaviside function to enforce the $\tau > 0$ constraint.

Our last little exercise is to use complex variables and Cauchy's theorem again. We start by noting that $\boldsymbol{D}$ and $\boldsymbol{E}$ and $G(\tau)$ are all *real*. Then we can integrate by parts and find things like:

$$\frac{\epsilon(\omega)}{\epsilon_0} - 1 = i\frac{G(0)}{\omega} - \frac{G'(0)}{\omega^2} + ... \tag{9.156}$$

from which we can conclude that $\epsilon(-\omega) = \epsilon^*(\omega^*)$ and the like. Note the even/odd imaginary/real oscillation in the series. $\epsilon(\omega)$ is therefore analytic in the UHP and we can write:

$$\frac{\epsilon(z)}{\epsilon_0} - 1 = \frac{1}{2\pi i}\oint_C \frac{\frac{\epsilon(\omega')}{\epsilon_0} - 1}{\omega' - z}d\omega' \tag{9.157}$$

We let $z = \omega + i\delta$ where $\delta \to 0_+$ (or deform the integral a bit below the singular point on the $\mathrm{Re}(\omega)$ axis). From the Plemlj Relation:

$$\frac{1}{\omega' - \omega - i\delta} = P\frac{1}{\omega' - \omega} + i\pi\delta(\omega' - \omega) \tag{9.158}$$

(see e.g. Wyld, Arfkin). If we substitute this into the integral above along the real axis only, do the delta-function part and subtract it out, cancel a factor of 1/2 that thus appears, we get:

$$\frac{\epsilon(\omega)}{\epsilon_0} = 1 + \frac{1}{i\pi}P\int_{-\infty}^{\infty}\frac{\frac{\epsilon(\omega')}{\epsilon_0} - 1}{\omega' - \omega}d\omega' \tag{9.159}$$

Although this looks like a single integral, because of the i in the denominator it is really *two*. The real part of the integrand becomes the imaginary part of the result and vice versa. That is:

$$\mathrm{Re}\left(\frac{\epsilon(\omega)}{\epsilon_0}\right) = 1 + \frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{\mathrm{Im}\left(\frac{\epsilon(\omega')}{\epsilon_0}\right)}{\omega' - \omega}d\omega' \tag{9.160}$$

$$\mathrm{Im}\left(\frac{\epsilon(\omega)}{\epsilon_0}\right) = -\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{\mathrm{Re}\left(\frac{\epsilon(\omega')}{\epsilon_0}\right) - 1}{\omega' - \omega}d\omega' \tag{9.161}$$

These are the *Kramers-Kronig Relations*. They tell us that the dispersive and absorptive properties of the medium are *not independent*. If we know the entire absorptive spectrum we can compute the dispersive spectrum and vice versa. There is one more form of the KK relations given in Jackson, derived from the discovery above that the real part of $\epsilon(\omega)$ is even in ω while the imaginary part is odd. See if you can derive this on your own for the fun of it all...

9.6 Plane Waves Assignment

To start off the semester right, visit the Wikipedia and Mathworld websites and look up and understand:

1. Separation of variables

2. Spherical Harmonics

3. Bessel Functions
4. Spherical Bessel Functions
5. Green's Functions
6. Wave Equation
7. Plane Wave

Just explore the kinds of things you can find there – I'm discovering that these web references are rapidly becoming THE universal free textbook. It is actually amazing to watch it happen (and participate in it as time permits).

Jackson, problems:

7.4, 7.6, 7.19, 7.21

Also, derive on your own *all* the principal results presented in these online lecture notes. It is easy to read and see me do it. It is not so easy to do it, even for me. Working through this, possibly several times until you really "get it", will truly improve your understanding of how everything works.

Chapter 10

Wave Guides

10.1 Boundary Conditions at a Conducting Surface: Skin Depth

Let us consider for a moment what time dependent EM fields look like at the surface of a "perfect" conductor. A perfect conductor can move as much charge instantly as is required to cancel all fields inside. The skin depth $\delta = \lim_{\sigma\to\infty} \sqrt{2/\mu\epsilon_b\sigma} = 0$ as α diverges – effectively all frequencies are "static" to a perfect conductor. This is how type I superconductors expel all field flux.

If we examine the fields in the vicinity of a boundary between a perfect conductor and a normal dielectric/diamagnetic material, we get:

$$(\boldsymbol{D} - \boldsymbol{D}_c) \cdot \hat{\boldsymbol{n}} = \hat{\boldsymbol{n}} \cdot \boldsymbol{D} = \Sigma \tag{10.1}$$

where $\boldsymbol{D}_c$ and $\boldsymbol{E}_c$ inside the conductor vanish. Similarly,

$$\hat{\boldsymbol{n}} \times (\boldsymbol{H} - \boldsymbol{H}_c) = \hat{\boldsymbol{n}} \times \boldsymbol{H} = \boldsymbol{K} \tag{10.2}$$

(where in these expressions, Σ is the surface charge density so we don't confuse it with the conductivity σ, sigh, and similarly $\boldsymbol{K}$ is the surface current density).

In addition to these two inhomogeneous equations that normal and parallel fields at the surface to sources, we have the usual two homogeneous equations:

$$
\begin{aligned}
\hat{\boldsymbol{n}} \cdot (\boldsymbol{B} - \boldsymbol{B}_c) &= 0 \qquad (10.3)\\
\hat{\boldsymbol{n}} \times (\boldsymbol{E} - \boldsymbol{E}_c) &= 0 \qquad (10.4)
\end{aligned}
$$

Note that these are pretty much precisely the boundary conditions for a static field and should come as no surprise. For perfect conductors, we expect the fields inside to vanish, which in turn implies that $\boldsymbol{E}$ outside must be normal to the conducting surface and $\boldsymbol{B}$ outside must lie only parallel to the conducting surface, as usual.

However, for materials that are *not* perfect conductors, the fields *don't* vanish instantly "at" the mathematical surface. Instead they die off exponentially within a few multiples of the skin depth δ. On scales *large* with respect to this, they will "look" like the static field conditions above, but of course within this cutoff things are very different.

For one thing, Ohm's law tells us that we cannot have an actual "surface layer of charge" because for any finite conductivity, the resistance scales like the cross-sectional area through which charge flows. Consequently the *real* boundary condition on $\boldsymbol{H}$ *precisely* at the surface is:

$$
\begin{aligned}
\hat{\boldsymbol{n}} \times (\boldsymbol{H} - \boldsymbol{H}_c) &= 0 \qquad (10.5)\\
\boldsymbol{H}_{||} &= \boldsymbol{H}_{c,||} \qquad (10.6)
\end{aligned}
$$

where $\boldsymbol{H}_{||} = (\hat{\boldsymbol{n}} \times \boldsymbol{H}) \times \hat{\boldsymbol{n}}$. However, this creates a problem! If this field varies rapidly in some direction (and it does) it will generate an electric field according to Faraday's law! If the direction of greatest variation is "into the conductor" (as the field is being screened by induced surface currents) then it will generate a small electric field parallel to the surface, one which is neglected (or rather, cannot occur) in the limit that the conductivity is infinite. This electric field, in turn, generates a current, which causes the gradual cancellation of $\boldsymbol{H}_{||}$ as less and less the total bulk current is enclosed by a decending loop boundary.

If the conductivity is large but not infinite, one way to figure out what happens is to employ a series of successive approximations starting with the assumption of perfect conductivity and using it to generate a first order correction based on the actual conductivity and wavelength. The way it works is:

1. First, we assume that outside the conductor we have only $\boldsymbol{E}_\perp$ and $\boldsymbol{H}_{||}$ from the statement of the boundary conditions assuming that the fields are instantly cancelled at the surface.

2. Assume $\delta \ll k^{-1}$ *along* the surface – the skin depth is much less than a wavelength and the fields (whatever they may be) *vanish* across roughly this length scale, so we can neglect variation (derivatives) with respect to coordinates that lie along the surface compared to the coordinate perpendicular to the surface.

3. Use this approximation in Maxwell's Equations, along with the assumed boundary conditions for a perfect conductor, to derive relations between the fields in the transition layer.

4. These relations determine the *small* corrections to the presumed boundary fields both just outside and just inside the surface.

The assumption of rapid variation only as one decends into the conductor is a key step, as we shall see.

Thus (from 1):

$$\hat{\boldsymbol{n}} \times (\boldsymbol{H} - \boldsymbol{H}_c) = 0 \tag{10.7}$$

or $\boldsymbol{H}_{||}(\text{outside}) = \boldsymbol{H}_{||}(\text{inside}) = \boldsymbol{H}_{||} \neq 0$, where the latter assumption is because the result is boring if there are no fields, right?

We both Ampere's law (assuming no displacement in the conductor to leading order) and Faraday's law to obtain relations for the harmonic fields in terms of curls of each other:

$$\begin{aligned} \boldsymbol{\nabla} \times \boldsymbol{H}_c &= \sigma \boldsymbol{E}_c = \boldsymbol{J} & (10.8) \\ \boldsymbol{\nabla} \times \boldsymbol{E}_c &= -\frac{\partial \boldsymbol{B}_c}{\partial t} = i\omega\mu_c \boldsymbol{H}_c & (10.9) \end{aligned}$$

become

$$\begin{aligned} \boldsymbol{E}_c &= \frac{1}{\sigma} \boldsymbol{\nabla} \times \boldsymbol{H}_c & (10.10) \\ \boldsymbol{H}_c &= -i\frac{1}{\mu_c \omega} \boldsymbol{\nabla} \times \boldsymbol{E}_c & (10.11) \end{aligned}$$

As we might expect, high frequencies create relatively large induced electric fields as the magnetic fields change, but high conductivity limits the size of the supported electric field for any given magnetic field strength in a *frequency independent way.*

Now we need to implement assumption 2 on the $\boldsymbol{\nabla}$ operator. If we pick a coordinate ξ to be perpendicular to the surface pointing into the conductor (in the $-\hat{\boldsymbol{n}}$ direction) and insist that only variations in this direction will be significant only on length scales of δ:

$$\boldsymbol{\nabla} \approx -\hat{\boldsymbol{n}}\frac{\partial}{\partial \xi} \tag{10.12}$$

then we get:

$$\begin{aligned} \boldsymbol{E}_c &\approx -\frac{1}{\sigma}\left(\hat{\boldsymbol{n}} \times \frac{\partial \boldsymbol{H}_c}{\partial \xi}\right) \\ \boldsymbol{H}_c &\approx i\frac{1}{\mu_c \omega}\left(\hat{\boldsymbol{n}} \times \frac{\partial \boldsymbol{E}_c}{\partial \xi}\right) \end{aligned} \tag{10.13}$$

(Note well the deliberate use of *approx* to emphasize that there may well be components of the fields in the normal direction or other couplings between the components in the surface, but those components do not *vary* particularly rapidly *along* the surface and so are not large contributors to the curl.)

These two equations are very interesting. They show that while the *magnitude* of the fields in the vicinity of the conducting surface may be large or small (depending on the charge and currents near the surface) the curls themselves are dominated by the particular components of $\boldsymbol{E}_c$ and $\boldsymbol{H}_c$ that are *in the plane* perpendicular to $\hat{\boldsymbol{n}}$ (and each other) because the field strengths (whatever they are) are most rapidly varying across the surface.

What this pair of equations ultimately does is show that *if* there is a magnetic field just inside the conductor parallel to its surface (and hence perpendicular to $\hat{\boldsymbol{n}}$) $\boldsymbol{H}_{||}$ that rapidly varies as one descends, *then* there must be an electric field $\boldsymbol{E}_{||}$ that is its partner. Our zeroth approximation boundary condition on $\boldsymbol{H}_{||}$ above shows that it is actually *continuous* across the mathematical surface of the boundary and does not have to be zero either just outside or just inside of it. However, in a good conductor the $\boldsymbol{E}_{||}$ field it produces is *small.*

This gives us a bit of an intuitive foundation for the manipulations of Maxwell's equations below. They should lead us to expressions for the coupled EM fields parallel to the surface that self-consistently result from these two equations.

We start by determining the component of $\boldsymbol{H}_c$ (the total vector magnetic field just inside the conductor) in the direction perpendicular to the surface:

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{H}_c = \frac{i}{\mu_c \omega} \hat{\boldsymbol{n}} \cdot (\hat{\boldsymbol{n}} \times \frac{\partial \boldsymbol{E}_c}{\partial \xi}) = 0 \tag{10.14}$$

This tells us that $\boldsymbol{H}_c = \boldsymbol{H}_{||} = (\hat{\boldsymbol{n}} \times \boldsymbol{H}_c) \times \hat{\boldsymbol{n}}$ – the magnetic field coupled by $\boldsymbol{E}_c$ by Faraday's law lies *in the plane of the conducting surface* to lowest order.

Next we form a vector that lies *perpendicular* to both the normal and the magnetic field. We expect $\boldsymbol{E}_c$ to lie along this direction one way or the other.

$$\begin{aligned}
\hat{\boldsymbol{n}} \times \boldsymbol{H}_c &= \hat{\boldsymbol{n}} \times \left(i \frac{1}{\mu_c \omega} \hat{\boldsymbol{n}} \times \frac{\partial \boldsymbol{E}_c}{\partial \xi} \right) \\
&= i \frac{1}{\mu_c \omega} \frac{\partial}{\partial \xi} \left(\hat{\boldsymbol{n}} (\hat{\boldsymbol{n}} \cdot \boldsymbol{E}_c) - \boldsymbol{E}_c \right) \\
&= -i \frac{1}{\mu_c \omega} \frac{\partial \boldsymbol{E}_{c,||}}{\partial \xi}
\end{aligned}$$

(where $\boldsymbol{E}_{c,\perp} = \hat{\boldsymbol{n}}(\hat{\boldsymbol{n}} \cdot \boldsymbol{E})$ and $\boldsymbol{E}_c = \boldsymbol{E}_{c,\perp} + \boldsymbol{E}_{c,||}$) and find that it does! The fact that the electric field varies most rapidly in the $-\hat{\boldsymbol{n}}$ $(+\xi)$ direction picks out its component in the plane *whatever it might be* and relates it to the magnetic field direction also in the plane.

However, this does not show that the two conditions can lead to a self-sustaining solution in the absence of driving external currents (for example). To show that we have to substitute *Ampere's* law back into this:

$$\begin{aligned}
\hat{\boldsymbol{n}} \times \boldsymbol{H}_c &= -i \frac{1}{\mu_c \omega} \frac{\partial}{\partial \xi} \left(-\frac{1}{\sigma} (\hat{\boldsymbol{n}} \times \frac{\partial \boldsymbol{H_c}}{\partial \xi}) \right) \\
\hat{\boldsymbol{n}} \times \boldsymbol{H}_c &= i \frac{1}{\mu_c \omega \sigma} (\hat{\boldsymbol{n}} \times \frac{\partial^2 \boldsymbol{H_c})}{\partial \xi^2}) \\
(\hat{\boldsymbol{n}} \times \frac{\partial^2 \boldsymbol{H_c})}{\partial \xi^2}) &= -i \mu_c \omega \sigma \hat{\boldsymbol{n}} \times \boldsymbol{H}_c \\
\frac{\partial^2}{\partial \xi^2} (\hat{\boldsymbol{n}} \times \boldsymbol{H_c}) &+ i \mu_c \omega \sigma (\hat{\boldsymbol{n}} \times \boldsymbol{H}_c) = 0
\end{aligned}$$

or

$$\frac{\partial^2}{\partial \xi^2}(\hat{\boldsymbol{n}} \times \boldsymbol{H}_c) + \frac{2i}{\delta^2}(\hat{\boldsymbol{n}} \times \boldsymbol{H}_c) = 0 \tag{10.15}$$

where we used the first result and substituted $\delta^2 = 2/(\mu_c \omega \sigma)$.

This is a well-known differential equation that can be written any of several ways. Let $\kappa^2 = \frac{2i}{\delta^2}$. It is equivalent to all of:

$$(\frac{\partial^2}{\partial \xi^2} + \kappa^2)(\hat{\boldsymbol{n}} \times \boldsymbol{H}_c) = 0 \tag{10.16}$$

$$(\frac{\partial^2}{\partial \xi^2} + \kappa^2)(\hat{\boldsymbol{n}} \times \boldsymbol{H}_c) \times \hat{\boldsymbol{n}} = 0 \tag{10.17}$$

$$(\frac{\partial^2}{\partial \xi^2} + \kappa^2)\boldsymbol{H}_{||} = 0 \tag{10.18}$$

$$(\frac{\partial^2}{\partial \xi^2} + \kappa^2)\boldsymbol{H}_c = 0 \tag{10.19}$$

Where:

$$(\hat{\boldsymbol{n}} \times \boldsymbol{H}_c) \times \hat{\boldsymbol{n}} = \boldsymbol{H}_{||} \tag{10.20}$$

as noted above. The solution to this form is then:

$$\boldsymbol{H}_c(\xi) = \boldsymbol{H}_0 e^{\pm\sqrt{-\kappa^2}\xi} \tag{10.21}$$

where $\boldsymbol{H}_0$ is the magnetic field vector *in the plane of the conductor* at the surface and where this equation indicates how this value is attenuated as one decends into the conductor.

As always, we have two linearly independent solutions. Either of them will work, and (given the already determined sign/branch associated with the time dependence $e^{-i\omega t}$) will ultimately have the *physical* interpretation of waves moving in the direction of $+\xi$ $(-\hat{\boldsymbol{n}})$ or in the direction of $-\xi$ $(\hat{\boldsymbol{n}})$. Let us pause for a moment to refresh our memory of taking the square root of complex numbers (use the subsection that treats this in the last chapter of these notes or visit Wikipedia of there is any problem understanding).

For *this* particular problem,

$$\sqrt{-\kappa^2} = \sqrt{-\frac{2i}{\delta^2}} = \pm\frac{1}{\delta}(-1 + i) \tag{10.22}$$

(draw this out in pictures). We want the solution that propagates *in*to the surface of the conductor, decending *from* the dielectric medium, which is the positive branch:

$$
\begin{aligned}
\boldsymbol{H}_c = \boldsymbol{H}_0 e^{\sqrt{-\kappa^2}\xi} &= \boldsymbol{H}_0 e^{\frac{1}{\delta}(-1+i)\xi} \\
&= \boldsymbol{H}_0 e^{-\frac{\xi}{\delta}} e^{i\frac{\xi}{\delta}} \qquad (10.23)
\end{aligned}
$$

(consider $e^{i\xi/\delta - \omega t}$).

Now we need to find an expression for $\boldsymbol{E}_c$, which we do by backsubstituting into Ampere's Law:

$$
\begin{aligned}
\boldsymbol{E}_c &= -\frac{1}{\sigma}\left(\hat{\boldsymbol{n}} \times \frac{\partial \boldsymbol{H}_c}{\partial \xi}\right) \\
&= -\frac{1}{\delta\sigma}(-1+i)\left(\hat{\boldsymbol{n}} \times \boldsymbol{H}_0\right) e^{\frac{1}{\delta}(-1+i)\xi} \\
\boldsymbol{E}_c &= \sqrt{\frac{\mu_c\omega}{2\sigma}}(1-i)(\hat{\boldsymbol{n}} \times \boldsymbol{H}_0) e^{-\frac{\xi}{\delta}} e^{i\frac{\xi}{\delta}} \qquad (10.24)
\end{aligned}
$$

Note well the direction! Obviously $\hat{\boldsymbol{n}} \cdot \boldsymbol{E}_c = 0$, (in this approximation) so $\boldsymbol{E}_c$ must lie in the plane of the conductor surface, just like $\boldsymbol{H}_{||}$!

As before (when we discussed fields in a good conductor):

- $\boldsymbol{E}_c, \boldsymbol{H}_c$ not in phase, but out of phase by $\pi/4$.
- Rapid decay as wave penetrates surface.
- $\boldsymbol{H}_c \gg \boldsymbol{E}_c$ (σ "large", δ "small") so energy is primarily magnetic.
- $\hat{\boldsymbol{n}} \perp \boldsymbol{E}_c \perp \boldsymbol{H}_c \perp \hat{\boldsymbol{n}}$ – fields are predominantly parallel to the surface and mutually transverse, they propagate "straight into" surface, attenuating rapidly as they go.
- Recall:

  $$\hat{\boldsymbol{n}} \times (\boldsymbol{E} - \boldsymbol{E}_c) = 0 \qquad (10.25)$$

  at the surface. Since $\boldsymbol{E}_c$ lies approximately in the surface, this yields

  $$\boldsymbol{E} \approx \boldsymbol{E}_c \approx \sqrt{\frac{\mu_c\omega}{2\sigma}}(1-i)(\hat{\boldsymbol{n}} \times \boldsymbol{H}_0) e^{-\frac{\xi}{\delta}} e^{i\frac{\xi}{\delta}} \qquad (10.26)$$

  just *outside* the surface – the field is approximately continuous! At this level of approximation, $\boldsymbol{\nabla} \times \boldsymbol{E} = i\omega\boldsymbol{B}$, $\boldsymbol{E}$ is parallel to the surface, and there is a small $\boldsymbol{B}_\perp$ to the surface of the same general order of magnitude as $\boldsymbol{E}$.

Since both $\boldsymbol{E}_{||} \neq 0$ and $\boldsymbol{H}_{||} \neq 0$ at the surface ($\xi = 0$) there must be a power flow *into* the conductor!

$$\frac{dP_{\rm in}}{dA} = -\frac{1}{2}\mathrm{Re}\left(\boldsymbol{n}\cdot(\boldsymbol{E}_c \times \boldsymbol{H}_c^*)\right) = \frac{\mu_c \omega \delta}{4}|\boldsymbol{H}_0|^2 \tag{10.27}$$

where we HOPE that it turns into heat. Let's see:

$$\boldsymbol{J} = \sigma \boldsymbol{E} = \sqrt{\frac{\mu_c \omega \sigma}{2}}(1-i)(\hat{\boldsymbol{n}} \times \boldsymbol{H}_0)e^{-\xi(1-i)/\delta} \tag{10.28}$$

so that the time averaged power loss is (from Ohm's Law):

$$\begin{aligned}
\frac{dP}{dV} = \frac{1}{\Delta A}\frac{dP}{d\xi} &= \frac{1}{2}\boldsymbol{J}\cdot\boldsymbol{E}^* = \frac{1}{2\sigma}\boldsymbol{J}\cdot\boldsymbol{J}^* & (10.29)\\
\Delta P &= \Delta A \frac{1}{2\sigma}\int_0^\infty d\xi \boldsymbol{J}\cdot\boldsymbol{J}^* \\
&= \Delta A \frac{\mu_c \omega}{2}|H_0|^2 \int_0^\infty d\xi e^{-2\xi/\delta} \\
&= \Delta A \frac{\mu_c \omega}{4}|H_0|^2 & (10.30)
\end{aligned}$$

which just happens to correspond to the flux of the pointing vector through a surface ΔA!

Finally, we need to define the "surface current":

$$\boldsymbol{K}_{\rm eff} = \int_0^\infty \boldsymbol{J} d\xi = (\hat{\boldsymbol{n}} \times \boldsymbol{H}) \tag{10.31}$$

where $\boldsymbol{H}$ is determined just outside(inside) of the surface of a "perfect" conductor in an idealized limit – note that we are just adding up the total current in the surface layer and that it all works out.

Hopefully this exposition is complete enough (and correct enough) that any bobbles from lecture are smoothed out. You can see that although Jackson blithely pops all sorts of punch lines down in the text, the actual algebra of getting them, while straightforward, is not trivial!

10.2 Mutilated Maxwell's Equations (MMEs)

We are now prepared to look at the propagation of waves in volumes of space bounded in some way by conducting surfaces. We'll generally assume that the conductors in question are "perfect" as far as boundary conditions on the dimensions of the volume in question are concerned. The place where this will lead to error is in the gradual attenuation of a propagating wave as it loses energy to the Joule heating of the surface of the bounding conductor, but this process will be slow relative to a wavelength and using the results of the previous section we can add this attenuation in by hand afterwards if necessary.

Since we are going to have to solve boundary value problems for the wave equations for the coupled field components, we'd better select a relatively simple geometry or we'll be here all semester. The two geometries we will examine are *cylindrical* waveguides where propagation is along the z axis of the cylinder and *rectangular* waveguides where the propagation is along the z axis of a waveguide with a rectangular cross-section in the $x - y$ plane of dimension $a \times b$. The transverse coordinates are therefore (ρ, ϕ) or (x, y), respectively.

As usual, we will start by assuming that we're dealing with a harmonic wave with time dependence $e^{-i\omega t}$, write down Maxwell's equations in free space (the cavity volume), turn them into wave equations for the field separately, note that the fields are coupled by Maxwell's equations themselves, and impose boundary conditions. The only thing that is "special" about a cylinder is the form of the Laplacian and how we separate the laplacian to respect the boundary conditions. Let's skip ahead to the wave equation since by now everybody should be able to do this in their sleep:

$$(\nabla^2 + \mu\epsilon\omega^2)\,\{\boldsymbol{E} \text{ or } \boldsymbol{B}\} = 0 \tag{10.32}$$

We look at propagation along z, making it "plane-wave-like":

$$\boldsymbol{E}(\boldsymbol{x}, t) = \boldsymbol{E}(\rho, \phi)e^{\pm ikz - i\omega t} \tag{10.33}$$

$$\boldsymbol{B}(\boldsymbol{x}, t) = \boldsymbol{B}(\rho, \phi)e^{\pm ikz - i\omega t} \tag{10.34}$$

so that the wave equation becomes:

$$\left(\nabla_\perp^2 + (\mu\epsilon\omega^2 - k^2)\right)\{\boldsymbol{E} \text{ or } \boldsymbol{B}\} = 0 \tag{10.35}$$

(Note that $\nabla_\perp^2 = \nabla^2 - \frac{\partial^2}{\partial z^2}$).

Resolve fields into components $\perp$ and $||$ to z:

$$\boldsymbol{E} = E_z\hat{\boldsymbol{z}} + (\hat{\boldsymbol{z}} \times \boldsymbol{E}) \times \hat{\boldsymbol{z}}$$

$$
\begin{aligned}
&= \boldsymbol{E}_z + \boldsymbol{E}_\perp && (10.36)\\
\boldsymbol{B} &= B_z\hat{\boldsymbol{z}} + (\hat{\boldsymbol{z}}\times\boldsymbol{B})\times\hat{\boldsymbol{z}} &&\\
&= \boldsymbol{B}_z + \boldsymbol{B}_\perp && (10.37)\\
& && (10.38)
\end{aligned}
$$

(defining $\boldsymbol{E}_z$ and $\boldsymbol{E}_\perp$ etc. in fairly obvious ways). Now we try to write Maxwell's equations in terms of these field components, assuming that the *only* z-dependence permitted is $e^{\pm ikz}$.

This isn't trivial to do – let's start with Faraday's law, for example:

$$
\boldsymbol{\nabla}\times\boldsymbol{E} = -\frac{\partial\boldsymbol{B}}{\partial t} = i\omega\boldsymbol{B} \tag{10.39}
$$

If we project out the z component of both sides we get:

$$
\begin{aligned}
\hat{\boldsymbol{z}}\cdot(\boldsymbol{\nabla}\times\boldsymbol{E}) &= i\omega B_z\\
\hat{\boldsymbol{z}}\cdot\left\{\left(\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}\right)\hat{\boldsymbol{x}}+\left(\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}\right)\hat{\boldsymbol{y}}\right. &+ \\
\left.\left(\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right)\hat{\boldsymbol{z}}\right\} &= i\omega B_z\\
\left(\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right) &= i\omega B_z\\
\hat{\boldsymbol{z}}\cdot(\boldsymbol{\nabla}_\perp\times\boldsymbol{E}_\perp) &= i\omega B_z && (10.40)
\end{aligned}
$$

as only the $\perp$ components of the curl contribute to the z direction. Similarly:

$$
\begin{aligned}
\hat{\boldsymbol{z}}\times(\boldsymbol{\nabla}\times\boldsymbol{E}) &= i\omega(\hat{\boldsymbol{z}}\times\boldsymbol{B})\\
\hat{\boldsymbol{z}}\times\left\{\left(\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}\right)\hat{\boldsymbol{x}}+\left(\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}\right)\hat{\boldsymbol{y}}\right. &+ \\
\left.\left(\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right)\hat{\boldsymbol{z}}\right\} &= i\omega(\hat{\boldsymbol{z}}\times\boldsymbol{B}_\perp)\\
\left(\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}\right)\hat{\boldsymbol{y}}-\left(\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}\right)\hat{\boldsymbol{x}} &= i\omega(\hat{\boldsymbol{z}}\times\boldsymbol{B}_\perp)\\
\frac{\partial\boldsymbol{E}_\perp}{\partial z}+i\omega(\hat{\boldsymbol{z}}\times\boldsymbol{B}_\perp) &= \boldsymbol{\nabla}_\perp E_z && (10.41)
\end{aligned}
$$

(where $\hat{\boldsymbol{z}}\times\boldsymbol{B} = \hat{\boldsymbol{z}}\times\boldsymbol{B}_\perp$, of course).

Ouch! Looks like working through the curl termwise is a certain amount of pain! However, now that we've done it once (and see how it goes) Ampere's law should be straightforward:

$$\begin{aligned}\hat{\boldsymbol{z}} \cdot (\boldsymbol{\nabla} \times \boldsymbol{H}) &= -i\omega D_z \\ \hat{\boldsymbol{z}} \cdot (\boldsymbol{\nabla}_\perp \times \boldsymbol{B}_\perp) &= -i\omega\mu\epsilon E_z\end{aligned}$$

and

$$\begin{aligned}\hat{\boldsymbol{z}} \times (\boldsymbol{\nabla} \times \boldsymbol{H}) &= -i\omega(\hat{\boldsymbol{z}} \times \boldsymbol{D}) \\ \frac{\partial \boldsymbol{B}_\perp}{\partial z} - i\omega\mu\epsilon(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) &= \boldsymbol{\nabla}_\perp B_z\end{aligned}$$

Finally, we have Gauss's Law(s):

$$\begin{aligned}\boldsymbol{\nabla} \cdot \boldsymbol{E} &= 0 \\ \boldsymbol{\nabla}_\perp \cdot \boldsymbol{E}_\perp + \frac{\partial E_z}{\partial z} &= 0 \\ \boldsymbol{\nabla}_\perp \cdot \boldsymbol{E}_\perp &= -\frac{\partial E_z}{\partial z}\end{aligned}$$

and identically,

$$\boldsymbol{\nabla}_\perp \cdot \boldsymbol{B}_\perp = -\frac{\partial B_z}{\partial z} \tag{10.42}$$

Let's collect all of these in just one place now:

$$\boldsymbol{\nabla}_\perp \cdot \boldsymbol{E}_\perp = -\frac{\partial E_z}{\partial z} \tag{10.43}$$

$$\boldsymbol{\nabla}_\perp \cdot \boldsymbol{B}_\perp = -\frac{\partial B_z}{\partial z} \tag{10.44}$$

$$\hat{\boldsymbol{z}} \cdot (\boldsymbol{\nabla}_\perp \times \boldsymbol{B}_\perp) = -i\omega\mu\epsilon E_z \tag{10.45}$$

$$\hat{\boldsymbol{z}} \cdot (\boldsymbol{\nabla}_\perp \times \boldsymbol{E}_\perp) = i\omega B_z \tag{10.46}$$

$$\frac{\partial \boldsymbol{B}_\perp}{\partial z} - i\omega\mu\epsilon(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) = \boldsymbol{\nabla}_\perp B_z \tag{10.47}$$

$$\frac{\partial \boldsymbol{E}_\perp}{\partial z} + i\omega(\hat{\boldsymbol{z}} \times \boldsymbol{B}_\perp) = \boldsymbol{\nabla}_\perp E_z \tag{10.48}$$

Gee, only a few *pages* of algebra to obtain in a *shortened* way what Jackson just puts down in three short lines. Hopefully the point is clear – to "get" a lot of this you have to sooner or later work it all out, however long it may take you, or you'll end up *memorizing* (or trying to) all of Jackson's results. Something that most normal humans could never do in a lifetime of trying...

Back to work, as there is still plenty to do.

10.3 TEM Waves

Now we can start looking at waveforms in various cavities. Suppose we let $E_z = B_z = 0$. Then the wave in the cavity is a pure *transverse electromagnetic (TEM) wave* just like a plane wave, *except* that it has to satisfy the boundary conditions of a perfect conductor at the cavity boundary!

Note from the equations above that:

$$\begin{aligned} \boldsymbol{\nabla}_\perp \cdot \boldsymbol{E}_\perp &= 0 \\ \boldsymbol{\nabla}_\perp \times \vec{E}_\perp &= 0 \end{aligned}$$

from which we can immediately see that:

$$\nabla_\perp^2 \boldsymbol{E}_\perp = 0 \tag{10.49}$$

and that

$$\boldsymbol{E}_\perp = -\boldsymbol{\nabla}\phi \tag{10.50}$$

for some suitable potential that satisfies $\nabla_\perp^2 \phi = 0$. The solution looks like a propagating *electrostatic* wave. From the wave equation we see that:

$$\mu\epsilon\omega^2 = k^2 \tag{10.51}$$

or

$$k = \pm\omega\sqrt{\mu\epsilon} \tag{10.52}$$

which is just like a plane wave (which can propagate in either direction, recall).

Again referring to our list of mutilated Maxwell equations above, we see that:

$$\begin{aligned} ik\boldsymbol{E}_\perp &= -i\omega(\hat{\boldsymbol{z}} \times \boldsymbol{B}_\perp) \\ \boldsymbol{D}_\perp &= -\frac{\omega\mu\epsilon}{k}(\hat{\boldsymbol{z}} \times \boldsymbol{H}_\perp) \\ \boldsymbol{D}_\perp &= \pm\sqrt{\mu\epsilon}(\hat{\boldsymbol{z}} \times \boldsymbol{H}_\perp) \end{aligned} \tag{10.53}$$

or working the other way, that:

$$\boldsymbol{B}_\perp = \pm\sqrt{\mu\epsilon}(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) \tag{10.54}$$

so we can easily find one from the other.

TEM waves cannot be sustained in a cylinder because the surrounding (perfect, recall) conductor is equipotential. Therefore $\boldsymbol{E}_\perp$ is zero as is $\boldsymbol{B}_\perp$. However, they are the *dominant* way energy is transmitted down a coaxial cable, where a potential difference is maintained between the central conductor and the coaxial sheathe. In this case the fields are very simple, as the $\boldsymbol{E}$ is purely radial and the $\boldsymbol{B}$ field circles the conductor (so the energy goes which way?) with no z components.

Finally, note that all frequencies are permitted for a TEM wave. It is not "quantized" by the appearance of eigenvalues due to a constraining boundary value problem.

10.4 TE and TM Waves

Note well that we have written the mutilated Maxwell Equations so that the z components are *all on the right hand side*. If they are known functions, and if the only z dependence is the complex exponential (so we can do all the z-derivatives and just bring down a $\pm ik$) then *the transverse components $\boldsymbol{E}_\perp$ and $\boldsymbol{B}_\perp$ are determined!*

In fact (for propagation in the $+z$ direction, $e^{+ikz-i\omega t}$):

$$\begin{aligned}
ik\boldsymbol{E}_\perp + i\omega(\hat{\boldsymbol{z}} \times \boldsymbol{B}_\perp) &= \boldsymbol{\nabla}_\perp E_z \\
ik(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) + i\omega\hat{\boldsymbol{z}} \times (\hat{\boldsymbol{z}} \times \boldsymbol{B}_\perp) &= \hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp E_z \\
ik(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) &= i\omega\boldsymbol{B}_\perp + \hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp E_z \qquad (10.55) \\
\hat{\boldsymbol{z}} \cdot (\boldsymbol{\nabla}_\perp \times \boldsymbol{B}_\perp) &= -i\omega\mu\epsilon E_z \qquad (10.56)
\end{aligned}$$

and

$$\begin{aligned}
ik\boldsymbol{B}_\perp - i\omega\mu\epsilon(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) &= \boldsymbol{\nabla}_\perp B_z \\
ik\boldsymbol{B}_\perp - \boldsymbol{\nabla}_\perp B_z &= i\omega\mu\epsilon(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) \\
i\frac{k^2}{\omega\mu\epsilon}\boldsymbol{B}_\perp - \frac{k}{\omega\mu\epsilon}\boldsymbol{\nabla}_\perp B_z &= ik(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp)
\end{aligned}$$

$$
i\frac{k^2}{\omega\mu\epsilon}\boldsymbol{B}_\perp - \frac{k}{\omega\mu\epsilon}\boldsymbol{\nabla}_\perp B_z = i\omega\boldsymbol{B}_\perp + \hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp E_z \tag{10.57}
$$

or

$$
\boldsymbol{B}_\perp = \frac{i}{\mu\epsilon\omega^2 - k^2}\left(k\boldsymbol{\nabla}_\perp B_z + \mu\epsilon\omega(\hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp E_z)\right) \tag{10.58}
$$

$$
\boldsymbol{E}_\perp = \frac{i}{\mu\epsilon\omega^2 - k^2}\left(k\boldsymbol{\nabla}_\perp E_z - \omega(\hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp B_z)\right) \tag{10.59}
$$

(where we started with the second equation and eliminated $\hat{\boldsymbol{z}} \times \boldsymbol{B}_\perp$ to get the second equation just like the first).

Now comes the relatively tricky part. Recall the boundary conditions for a perfect conductor:

$$
\begin{aligned}
\hat{\boldsymbol{n}} \times (\boldsymbol{E} - \boldsymbol{E}_c) = \hat{\boldsymbol{n}} \times \boldsymbol{E} &= 0 \\
\hat{\boldsymbol{n}} \cdot (\boldsymbol{B} - \boldsymbol{B}_c) = \hat{\boldsymbol{n}} \cdot \boldsymbol{B} &= 0 \\
\hat{\boldsymbol{n}} \times \boldsymbol{H} &= \boldsymbol{K} \\
\hat{\boldsymbol{n}} \cdot \boldsymbol{D} &= \Sigma
\end{aligned}
$$

They tell us basically that $\boldsymbol{E}$ ($\boldsymbol{D}$) is strictly perpendicular to the surface and that $\boldsymbol{B}$ ($\boldsymbol{H}$) is strictly parallel to the surface of the conductor *at* the surface of the conductor.

This means that it is *not* necessary for E_z or B_z *both* to vanish everywhere inside the dielectric (although both can, of course, and result in a TEM wave or no wave at all). All that is strictly required by the boundary conditions is for

$$
E_z|_S = 0 \tag{10.60}
$$

on the conducting surface S (it can only have a normal component so the z component must vanish). The condition on B_z is even weaker. It must lie parallel to the surface and be continuous across the surface (where $\boldsymbol{H}$ can discontinuously change because of $\boldsymbol{K}$). That is:

$$
\frac{\partial B_z}{\partial n}|_S = 0 \tag{10.61}
$$

We therefore have two possibilities for *non*-zero E_z or B_z that can act as source term in the mutilated Maxwell Equations.

10.4.1 TM Waves

$$\begin{aligned} B_z &= 0 \\ E_z|_S &= 0 \end{aligned} \tag{10.62, 10.63}$$

The magnetic field is strictly transverse, but the electric field in the z direction only has to vanish at the boundary – elsewhere it can have a z component.

Thus:

$$\begin{aligned} \boldsymbol{E}_\perp &= \frac{i}{\mu\epsilon\omega^2 - k^2}\left(k\boldsymbol{\nabla}_\perp E_z - \omega(\hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp B_z)\right) \\ (\mu\epsilon\omega^2 - k^2)\boldsymbol{E}_\perp &= ik\boldsymbol{\nabla}_\perp E_z \\ \frac{1}{ik}(\mu\epsilon\omega^2 - k^2)\boldsymbol{E}_\perp &= \boldsymbol{\nabla}_\perp E_z \end{aligned} \tag{10.64}$$

which looks just perfect to substitute into:

$$\begin{aligned} \boldsymbol{B}_\perp &= \frac{i}{\mu\epsilon\omega^2 - k^2}\left(k\boldsymbol{\nabla}_\perp B_z + \mu\epsilon\omega(\hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp E_z)\right) \\ (\mu\epsilon\omega^2 - k^2)\boldsymbol{B}_\perp &= i\mu\epsilon\omega(\hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp E_z) \\ (\mu\epsilon\omega^2 - k^2)\boldsymbol{B}_\perp &= \frac{\mu\epsilon\omega}{k}(\mu\epsilon\omega^2 - k^2)(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) \end{aligned} \tag{10.65}$$

giving us:

$$\boldsymbol{B}_\perp = \pm\frac{\mu\epsilon\omega}{k}(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) \tag{10.66}$$

or (as the book would have it):

$$\boldsymbol{H}_\perp = \pm\frac{\epsilon\omega}{k}(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) \tag{10.67}$$

(where as usual the two signs indicate the direction of wave propagation).

Of course, we still have to find at least one of the two fields for this to do us any good. Or do we? Looking above we see:

$$\begin{aligned} (\mu\epsilon\omega^2 - k^2)\boldsymbol{E}_\perp &= ik\boldsymbol{\nabla}_\perp\psi \\ \boldsymbol{E}_\perp &= \frac{\pm ik}{(\mu\epsilon\omega^2 - k^2)}\boldsymbol{\nabla}_\perp\psi \end{aligned} \tag{10.68}$$

Where $\psi(x,y)e^{ikz} = E_z$. This must satisfy the transverse wave function:

$$\left(\nabla_\perp^2 + (\mu\epsilon\omega^2 - k^2)\right)\psi = 0 \tag{10.69}$$

and the boundary conditions for a TM wave:

$$\psi|_S = 0 \tag{10.70}$$

TE Waves

$$E_z = 0 \tag{10.71}$$
$$\frac{\partial B_z}{\partial n}|_S = 0 \tag{10.72}$$

The electric field is strictly transverse, but the magnetic field in the z-direction can be nonzero. Doing exactly the same algebra on the same two equations as we used in the TM case, we get instead:

$$\boldsymbol{H}_\perp = \pm\frac{k}{\mu\omega}(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp) \tag{10.73}$$

along with

$$\boldsymbol{B}_\perp = \frac{\pm ik}{(\mu\epsilon\omega^2 - k^2)}\boldsymbol{\nabla}_\perp\psi \tag{10.74}$$

where $\psi(x,y)e^{ikz} = B_z$ and

$$\left(\nabla_\perp^2 + (\mu\epsilon\omega^2 - k^2)\right)\psi = 0 \tag{10.75}$$

and the boundary conditions for a TE wave:

$$\frac{\partial\psi}{\partial n}|_S = 0 \tag{10.76}$$

10.4.2 Summary of TE/TM waves

The transverse wave equation and boundary condition (dirichlet or neumann) are an *eigenvalue problem*. We can see two things right away. First of all:

$$\mu\epsilon\omega^2 \geq k^2 \tag{10.77}$$

or we no longer have a wave, we have an exponential function that cannot be made to satisfy the boundary conditions on the entire surface. Alternatively,

$$v_p^2 = \frac{\omega^2}{k^2} \geq \frac{1}{\mu\epsilon} = v^2 \tag{10.78}$$

which has the lovely property (as a phase velocity) of being faster than the speed of light in the medium!

To proceed further in our understanding, we need to look at an actual example – we'll find that only certain $k_n = k_0 n$ for $n = 1, 2, 3 ... n_{\mathrm{cutoff}}$ will permit the boundary conditions to be solved, and we'll learn some important things about the propagating solutions at the same time.

10.5 Rectangular Waveguides

Rectangular waveguides are important for two reasons. First of all, the Laplacian operator separates nicely in Cartesian coordinates, so that the boundary value problem that must be solved is both familiar and straightforward. Second, they are extremely common in actual application in physics laboratories for piping e.g. microwaves around as experimental probes.

In Cartesian coordinates, the wave equation becomes:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + (\mu\epsilon\omega^2 - k^2) \right) \psi = 0 \tag{10.79}$$

This wave equation separates and solutions are products of sin, cos or exponential functions in each variable separately. To determine which combination to use it suffices to look at the BC's being satisfied. For TM waves, one solves for $\psi = E_z$ subject to $E_z|_S = 0$, which is automatically true if:

$$E_z(x, y) = \psi_{mn}(x, y) = E_0 \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \tag{10.80}$$

where a and b are the dimensions of the x and y sides of the boundary rectangle and where in principle $m, n = 0, 1, 2....$

However, the wavenumber of any given mode (given the frequency) is determined from:

$$k^2 = \mu\epsilon\omega^2 - \pi^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2} \right) + \tag{10.81}$$

where $k^2 > 0$ for a "wave" to exist to propagate at all. If either index m or n is zero, there is no wave, so the first mode that can propagate has a dispersion relation of:

$$k_{11}^2 = \mu\epsilon\omega^2 - \pi^2(\frac{1}{a^2} + \frac{1}{b^2}) \tag{10.82}$$

so that:

$$\omega \geq \frac{\pi}{\sqrt{\mu\epsilon}}\sqrt{\frac{1}{a^2} + \frac{1}{b^2}} = \omega_{c,\mathrm{TM}}(11) \tag{10.83}$$

Each combination of permitted m and n is associated with a cutoff of this sort – waves with frequencies greater than or equal to the cutoff can support propogation in all the modes with lower cutoff frequencies.

If we repeat the argument above for TE waves (as is done in Jackson, which is why I did TM here so you could see them both) you will be led by nearly identical arguments to the conclusion that the lowest frequency mode cutoff occurs for $a > b$, $m = 1$ and $n = 0$ to produce the $H_z(x, y) = \psi(x, y)$ solution to the wave equation above. The cutoff in this case is:

$$\omega \geq \frac{\pi}{\sqrt{\mu\epsilon}}\frac{1}{a} = \omega_{c,\mathrm{TE}}(10) < \omega_{c,\mathrm{TM}}(11) \tag{10.84}$$

There exists, therefore, a range of frequencies in between where *only* one TE mode is supported with dispersion:

$$k^2 = k_{10}^2 = \mu\epsilon\omega^2 - \frac{\pi^2}{a^2}. \tag{10.85}$$

Note well that this mode and cutoff corresponds to exactly one-half a free-space wavelength across the long dimension of the waveguide. The wave solution for the right-propagating TE mode is:

$$H_z = H_0 \cos\left(\frac{\pi x}{a}\right) e^{ikz - i\omega t} \tag{10.86}$$

$$H_x = \frac{ik}{\mu\epsilon\omega^2 - k^2}\frac{\partial H_z}{\partial x} = -\frac{ika}{\pi} H_0 \sin\left(\frac{\pi x}{a}\right) e^{ikz - i\omega t} \tag{10.87}$$

$$E_y = \frac{\mu\omega}{k} H_x = \frac{i\mu\omega a}{\pi} H_0 \sin\left(\frac{\pi x}{a}\right) e^{ikz - i\omega t} \tag{10.88}$$

We used $\gamma^2 = \mu\epsilon\omega^2 - k^2 = \pi^2/a^2$ and $\boldsymbol{E}_\perp = ik/\gamma^2 \boldsymbol{\nabla}_\perp \psi$ to get the second of these, and $\boldsymbol{H}_\perp = \frac{k}{\omega\mu}(\hat{\boldsymbol{z}} \times \boldsymbol{E}_\perp))$ to get the last one.

There is a lot more one can study in Jackson associated with waveguides, but we must move on at this time to a brief look at resonant cavities (another important topic) and multipoles.

10.6 Resonant Cavities

We will consider a resonant cavity to be a waveguide of length d with caps at both ends. As before, we must satisfy TE or TM boundary conditions on the cap surfaces, either Dirichlet in E_z or Neumann in B_z. In between, we expect to find harmonic standing waves instead of travelling waves.

Elementary arguments for presumed standing wave z-dependence of:

$$A \sin kz + B \cos kz \tag{10.89}$$

such that the solution has nodes or antinodes at both ends lead one to conclude that only:

$$k = p\frac{\pi}{d} \tag{10.90}$$

for $p = 0, 1, 2...$ are supported by the cavity. For TM modes $\boldsymbol{E}_\perp$ must vanish on the caps because the nonzero E_z field must be the only E field component sustained, hence:

$$E_z = \psi(x, y) \cos\left(\frac{p\pi z}{d}\right) \tag{10.91}$$

For TE modes H_z must vanish as the only permitted field component is a non-zero $\boldsymbol{H}_\perp$, hence:

$$H_z = \psi(x, y) \sin\left(\frac{p\pi z}{d}\right) \tag{10.92}$$

Given these forms and the relations already derived for e.g. a rectangular cavity, one can easily find the formulae for the permitted transverse fields, e.g.:

$$\boldsymbol{E}_\perp = -\frac{p\pi}{d(\mu\epsilon\omega^2 - k^2)} \sin\left(\frac{p\pi z}{d}\right) \boldsymbol{\nabla}_\perp\psi \tag{10.93}$$

$$\boldsymbol{H}_\perp = -\frac{i\epsilon\omega}{\mu\epsilon\omega^2 - k^2} \cos\left(\frac{p\pi z}{d}\right) (\hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp\psi) \tag{10.94}$$

for TM fields and

$$\boldsymbol{E}_\perp = -\frac{i\mu\omega}{\mu\epsilon\omega^2 - k^2} \sin\left(\frac{p\pi z}{d}\right) (\hat{\boldsymbol{z}} \times \boldsymbol{\nabla}_\perp\psi) \tag{10.95}$$

$$\boldsymbol{H}_\perp = \frac{p\pi}{d(\mu\epsilon\omega^2 - k^2)} \cos\left(\frac{p\pi z}{d}\right) \boldsymbol{\nabla}_\perp\psi \tag{10.96}$$

for TE fields, with $\psi(x, y)$ determined as before for cavities.

However, now k is *doubly* determined as a function of *both* p and d *and* as a function of m and n. The only frequencies that lead to acceptable solutions are ones where the two *match*, where the resonant k in the z direction corresponds to a permitted $k(\omega)$ associated with a waveguide mode.

I leave you to read about the definition of Q:

$$Q = \frac{\omega_0}{\Delta\omega} \tag{10.97}$$

or the fractional energy loss per cycle of the cavity oscillator in the limit where this quantity is small compared to the total energy. Note that $\Delta\omega$ is the full width at half maximum of the presumed resonant form (basically the same as was presumed in our discussions of dispersion, but for energy instead of field).

I strongly advise that you go over this on your own – Q describes the damping of energy stored in a cavity mode due to e.g. the finite conductivity of the walls or the partial transparency of the end caps to energy (as might exist in the case of a laser cavity). If you go into laser physics, you will very much need this. If not, you'll need to understand the general idea of Q to teach introductory physics and e.g. LRC circuits or damped driven harmonic oscillators, where it also occurs and should know it at *least* qualitatively for e.g. qualifiers. I added an optional problem for resonant cavities to the homework assignment in case you wanted something specific to work on while studying this.

10.7 Wave Guides Assignment

Jackson 8.2,8.4(,8.6 optional)

Chapter 11

Radiation

Well, now we have learned a little about how to describe waves propagating through "free" space – dielectric media, possibly bounded by a conducting surface. But how did they get there? Well, sit yourselves down and I'll tell you. They were radiated there by *accelerating, time dependent charge–current distributions!*

And now we'll learn how...

Note well! This treatment differs *substantially* from Jackson's, which actually kinda sucks. Ultimately it will be much simpler to understand and is consistently developed. However, it really is the same thing and one gets the same general expressions for the multipole fields or potentials.

11.1 Maxwell's Equations, Yet Again

Suppose we are given a system of classical charges that oscillate harmonically with time. Note that, as before, this can be viewed as the special case of the Fourier transform at a particular frequency of a general time dependent distribution; however, this is a very involved issue that we will examine in detail later in the semester.

The form of the charge distribution we will study for the next few weeks is:

$$\rho(\boldsymbol{x}, t) = \rho(\boldsymbol{x}) e^{-i\omega t} \tag{11.1}$$

$$\boldsymbol{J}(\boldsymbol{x}, t) = \boldsymbol{J}(\boldsymbol{x}) e^{-i\omega t}. \tag{11.2}$$

The spatial distribution is essentially "arbitrary". Actually, we want it to have **compact support** which just means that it doesn't extend to infinity in any direction. Later we will also want it to be small with respect to a wavelength.

11.1.1 Quickie Review of Chapter 6

Recall the following morphs of Maxwell's equations, this time *with* the sources and expressed in terms of potentials by means of the homogeneous equations. Gauss's Law for magnetism is:

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \tag{11.3}$$

This is an identity if we define $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$:

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{A}) = 0 \tag{11.4}$$

Similarly, Faraday's Law is

$$\boldsymbol{\nabla} \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \tag{11.5}$$

$$\boldsymbol{\nabla} \times \boldsymbol{E} + \frac{\partial \boldsymbol{\nabla} \times \boldsymbol{A}}{\partial t} = 0 \tag{11.6}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{E} + \frac{\partial \boldsymbol{A}}{\partial t}) = 0 \tag{11.7}$$

and is satisfied as an identity by a scalar potential such that:

$$\boldsymbol{E} + \frac{\partial \boldsymbol{A}}{\partial t} = -\boldsymbol{\nabla}\phi \tag{11.8}$$

$$\boldsymbol{E} = -\boldsymbol{\nabla}\phi - \frac{\partial \boldsymbol{A}}{\partial t} \tag{11.9}$$

Now we look at the inhomogeneous equations in terms of the potentials. Ampere's Law:

$$\boldsymbol{\nabla} \times \boldsymbol{B} = \mu(\boldsymbol{J} + \epsilon \frac{\partial \boldsymbol{E}}{\partial t}) \tag{11.10}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{A}) = \mu(\boldsymbol{J} + \epsilon \frac{\partial \boldsymbol{E}}{\partial t}) \tag{11.11}$$

$$\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A} = \mu\boldsymbol{J} + \mu\epsilon\frac{\partial\boldsymbol{E}}{\partial t} \tag{11.12}$$

$$\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A} = \mu\boldsymbol{J} - \mu\epsilon\boldsymbol{\nabla}\frac{\partial\phi}{\partial t} - \mu\epsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2} \tag{11.13}$$

$$\nabla^2\boldsymbol{A} - \mu\epsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2} = -\mu\boldsymbol{J} + \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{A} + \mu\epsilon\frac{\partial\phi}{\partial t}) \tag{11.14}$$

Similarly Gauss's Law for the electric field becomes:

$$\boldsymbol{\nabla}\cdot\boldsymbol{E} = \frac{\rho}{\epsilon} \tag{11.15}$$

$$\boldsymbol{\nabla}\cdot\left(-\boldsymbol{\nabla}\phi - \frac{\partial\boldsymbol{A}}{\partial t}\right) = \frac{\rho}{\epsilon} \tag{11.16}$$

$$\nabla^2\phi + \frac{\partial\boldsymbol{\nabla}\cdot\boldsymbol{A}}{\partial t} = -\frac{\rho}{\epsilon} \tag{11.17}$$

In the the Lorentz gauge,

$$\boldsymbol{\nabla}\cdot\mathbf{A} + \mu\epsilon\frac{\partial\Phi}{\partial t} = 0 \tag{11.18}$$

the potentials satisfy the following *inhomogeneous* wave equations:

$$\nabla^2\Phi - \mu\epsilon\frac{\partial^2\Phi}{\partial t^2} = -\frac{\rho}{\epsilon} \tag{11.19}$$

$$\nabla^2\mathbf{A} - \mu\epsilon\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu\boldsymbol{J} \tag{11.20}$$

where ρ and $\boldsymbol{J}$ are the charge density and current density distributions, respectively. For the time being we will stick with the Lorentz gauge, although the Coulomb gauge:

$$\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0 \tag{11.21}$$

is more convenient for certain problems. It is probably worth reminding y'all that the Lorentz gauge condition itself is really just one out of a whole family of choices.

Recall that (or more properly, observe that in its role in these wave equations)

$$\mu\epsilon = \frac{1}{v^2} \tag{11.22}$$

where v is the speed of light in the medium. For the time being, let's just simplify life a bit and agree to work in a vacuum:

$$\mu_0\epsilon_0 = \frac{1}{c^2} \tag{11.23}$$

so that:

$$\nabla^2\Phi - \frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2} = -\frac{\rho}{\epsilon_0} \tag{11.24}$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu_0\boldsymbol{J} \tag{11.25}$$

If/when we look at wave sources embedded in a dielectric medium, we can always change back as the general formalism will not be any different.

11.2 Green's Functions for the Wave Equation

As by now you should fully understand from working with the Poisson equation, one very general way to solve inhomogeneous partial differential equations (PDEs) is to build a Green's function[1] and write the solution as an integral equation.

Let's very quickly review the general concept (for a further discussion don't forget WIYF ,MWIYF). Suppose $\mathcal{D}$ is a general (second order) linear partial differential operator on e.g. $\mathbb{R}^3$ and one wishes to solve the inhomogeneous equation:

$$\mathcal{D}f(\boldsymbol{x}) = \rho(\boldsymbol{x}) \tag{11.26}$$

for f.

If one can find a solution $G(\boldsymbol{x}-\boldsymbol{x}_0)$ to the associated differential equation for a *point* source function[2] :

$$\mathcal{D}G(\boldsymbol{x},\boldsymbol{x}_0) = \delta(\boldsymbol{x}-\boldsymbol{x}_0) \tag{11.27}$$

[1]Note that this expression stands for: "The generalized point source potential/field developed by Green." A number of people criticize the various ways of referring to it – Green function (what color was that again? what shade of Green?), Greens function (a function made of lettuce and spinach and kale?), "a" Green's function (a singular representative of a plural class referenced as a singular object). All have problems. I tend to go with the latter of these as it seems least odd to me.

[2] Note well that both the Green's "function" and the associated Dirac delta "function" are not *functions* – they are defined in terms of *limits of a distribution* in such a way that the interchange of limits and values of the integrals above make sense. This is necessary as both of the objects are *singular* in the limit and hence are meaningless without the limiting process. However, we'll get into *real* trouble if we have to write "The limit of the distribution defined by Green that is the solution of an inhomogeneous PDE with a

then (subject to various conditions, such as the ability to interchange the differential operator and the integration) to solution to this problem is a *Fredholm Integral Equation* (a convolution of the Green's function with the source terms):

$$f(\boldsymbol{x}) = \chi(\boldsymbol{x}) + \int_{\mathbb{R}^3} G(\boldsymbol{x}, \boldsymbol{x}_0)\rho(\boldsymbol{x}_0)d^3x_0 \tag{11.28}$$

where $\chi(\boldsymbol{x})$ is an *arbitrary* solution to the associated homogeneous PDE:

$$\mathcal{D}\left[\chi(\boldsymbol{x})\right] = 0 \tag{11.29}$$

This solution can easily be verified:

$$f(\boldsymbol{x}) = \chi(\boldsymbol{x}) + \int_{\mathbb{R}^3} G(\boldsymbol{x}, \boldsymbol{x}_0)\rho(\boldsymbol{x}_0)d^3x_0 \tag{11.30}$$

$$\mathcal{D}f(\boldsymbol{x}) = \mathcal{D}\left[\chi(\boldsymbol{x})\right] + \mathcal{D}\int_{\mathbb{R}^3} G(\boldsymbol{x}, \boldsymbol{x}_0)\rho(\boldsymbol{x}_0)d^3x_0 \tag{11.31}$$

$$\rho(\boldsymbol{x}_0)d^3x_0 \tag{11.32}$$

$$\mathcal{D}f(\boldsymbol{x}) = 0 + \int_{\mathbb{R}^3} \mathcal{D}G(\boldsymbol{x}, \boldsymbol{x}_0)\rho(\boldsymbol{x}_0)d^3x_0 \tag{11.33}$$

$$\mathcal{D}f(\boldsymbol{x}) = 0 + \int_{\mathbb{R}^3} \delta(\boldsymbol{x} - \boldsymbol{x}_0)\rho(\boldsymbol{x}_0)d^3x_0 \tag{11.34}$$

$$\mathcal{D}f(\boldsymbol{x}) = \rho(\boldsymbol{x}) \tag{11.35}$$

It seems, therefore, that we should *thoroughly understand* the ways of building Green's functions in general for various important PDEs. I'm uncertain of how much of this to do within these notes, however. This isn't really "Electrodynamics", it is mathematical physics, one of the fundamental toolsets you need to do Electrodynamics, quantum mechanics, classical mechanics, and more. So check out Arfken, Wyld, WIYF , MWIYFand we'll content ourselves with a very quick review of the principle ones we need:

11.2.1 Poisson Equation

The Green's function for the Poisson (inhomogeneous Laplace) equation:

$$\nabla^2\phi = -\frac{\rho}{\epsilon_0} \tag{11.36}$$

source distribution that in the same limit approaches a unit source supported at a single point" instead of just "Green's function". So we won't.

is the solution to:

$$\nabla^2 G(\boldsymbol{x}, \boldsymbol{x}_0) = \delta(\boldsymbol{x} - \boldsymbol{x}_0) \tag{11.37}$$

Thus $G(\boldsymbol{x}, \boldsymbol{x}_0)$ satisfies the homogeneous Laplace PDE everywhere but at the single point $\boldsymbol{x}_0$. The solution to the Laplace equation that has the right degree of singularity is the "potential of a unit point charge":

$$G(\boldsymbol{x}, \boldsymbol{x}_0) = \frac{-1}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} \tag{11.38}$$

located at $\boldsymbol{x}_0$. Hence:

$$\phi(\boldsymbol{x}) = \chi_0(\boldsymbol{x}) + \frac{1}{4\pi\epsilon_0}\int_V \frac{\rho(\boldsymbol{x}_0)}{|\boldsymbol{x} - \boldsymbol{x}_0|} d^3x_0 \tag{11.39}$$

which is just exactly correct.

Note well that the *inhomogeneous term* $\chi_0(\boldsymbol{x})$ solves the *homogeneous* Laplace equation and has various interpretations. It can be viewed as a "boundary term" (surface integral on $S = \partial V$, the surface S bounding the volume V (Green's Theorem) or, as we shall see, as the potential of all the charges in the volume exterior to V, or as a gauge transformation of the potential. All are true, but the "best" way to view it is as the potential of exterior charges as that is *what it is* in nature even when it is expressed, via integration by parts, as a surface integral, for a very sensible choice of asymptotic behavior of the potential.

Note *equally* well that the Green's function itself has precisely the same gauge freedom, and can be written in its most general form as:

$$G(\boldsymbol{x}, \boldsymbol{x}_0) = F(\boldsymbol{x}, \boldsymbol{x}_0) + \frac{-1}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} \tag{11.40}$$

where $\nabla^2 F(\boldsymbol{x}, \boldsymbol{x}_0) = \nabla_0^2 F(\boldsymbol{x}, \boldsymbol{x}_0) = 0$ is any bilinear (symmetric in both coordinates) solution to the Laplace equation! However, we will *not* proceed this way in this part of the course as it is in a sense *unphysical* to express the PDEs this way even though it does upon occasion facilitate the solution algebraically.

11.2.2 Green's Function for the Helmholtz Equation

If we fourier transform the wave equation, or alternatively attempt to find solutions with a specified harmonic behavior in time $e^{-i\omega t}$, we convert it into the following *spatial* form:

$$\left(\nabla^2 + k^2\right)\phi(\boldsymbol{x}) = -\frac{\rho_\omega}{\epsilon_0} \tag{11.41}$$

(for example, from the wave equation above, where $\rho(\boldsymbol{x},t) = \rho_\omega(\boldsymbol{x})e^{-i\omega t}$, $\phi(\boldsymbol{x},t) = \phi_\omega(\boldsymbol{x})e^{-i\omega t}$, and $k^2c^2 = \omega^2$ by assumption). This is called the *inhomogeneous Helmholtz equation* (IHE).

The Green's function therefore has to solve the PDE:

$$\left(\nabla^2 + k^2\right)G(\boldsymbol{x},\boldsymbol{x}_0) = \delta(\boldsymbol{x}-\boldsymbol{x}_0) \tag{11.42}$$

Once again, the Green's function satisfies the *homogeneous* Helmholtz equation (HHE). Furthermore, clearly the Poisson equation is the $k \to 0$ limit of the Helmholtz equation. It is straightforward to show that there are *several* functions that are good candidates for G. They are:

$$\begin{aligned}
G_0(\boldsymbol{x},\boldsymbol{x}_0) &= \frac{-\cos(k|\boldsymbol{x}-\boldsymbol{x}_0|)}{4\pi|\boldsymbol{x}-\boldsymbol{x}_0|} && (11.43)\\
G_+(\boldsymbol{x},\boldsymbol{x}_0) &= \frac{-e^{+ik|\boldsymbol{x}-\boldsymbol{x}_0|}}{4\pi|\boldsymbol{x}-\boldsymbol{x}_0|} && (11.44)\\
G_-(\boldsymbol{x},\boldsymbol{x}_0) &= \frac{-e^{-ik|\boldsymbol{x}-\boldsymbol{x}_0|}}{4\pi|\boldsymbol{x}-\boldsymbol{x}_0|} && (11.45)
\end{aligned}$$

As before, one can add arbitrary bilinear solutions to the HHE, $(\nabla^2 + k^2)F(\boldsymbol{x},\boldsymbol{x}_0) = (\nabla_0^2 + k^2)F(\boldsymbol{x},\boldsymbol{x}_0) = 0$ to any of these and the result is still a Green's function. In fact, these forms are related by this sort of transformation and superposition:

$$G_0(\boldsymbol{x},\boldsymbol{x}_0) = \frac{1}{2}(G_+(\boldsymbol{x},\boldsymbol{x}_0) + G_-(\boldsymbol{x},\boldsymbol{x}_0)) \tag{11.46}$$

or

$$\begin{aligned}
G_+(\boldsymbol{x},\boldsymbol{x}_0) &= F(\boldsymbol{x},\boldsymbol{x}_0) + G_0(\boldsymbol{x},\boldsymbol{x}_0) && (11.47)\\
&= \frac{-i\sin(k|\boldsymbol{x}-\boldsymbol{x}_0|)}{4\pi|\boldsymbol{x}-\boldsymbol{x}_0|} + G_0(\boldsymbol{x},\boldsymbol{x}_0) && (11.48)
\end{aligned}$$

etc.

In terms of any of these:

$$\begin{aligned}
\phi(\boldsymbol{x}) &= \chi_0(\boldsymbol{x}) - \frac{1}{\epsilon_0}\int_V \rho(\boldsymbol{x}_0)G(\boldsymbol{x},\boldsymbol{x}_0)d^3x_0 && (11.49)\\
&= \chi_0(\boldsymbol{x}) + \frac{1}{4\pi\epsilon_0}\int_V \frac{\rho(\boldsymbol{x}_0)e^{ik|\boldsymbol{x}-\boldsymbol{x}_0|}}{|\boldsymbol{x}-\boldsymbol{x}_0|}d^3x_0 && (11.50)
\end{aligned}$$

where $(\nabla^2 + k^2)\chi_0(\boldsymbol{x}) = 0$ as usual.

We name these three basic Green's functions according to their *asymptotic time dependence* far away from the volume V. In this region we expect to see a time dependence emerge from the integral of e.g.

$$\phi(\boldsymbol{x}, t) \sim e^{ikr - i\omega t} \tag{11.51}$$

where $r = |\boldsymbol{x}|$. This is an *outgoing spherical wave.* Consequently the Green's functions above are usually called the *stationary wave*, *outgoing wave* and *incoming wave* Green's functions.

It is essential to note, however, that *any* solution to the IHE can be constructed from *any* of these Green's functions! This is because the *form* of the solutions always differ by a homogeneous solution (as do the Green's functions) themselves. The main reason to use one or the other is to keep the form of the solution *simple and intuitive*! For example, if we are looking for a $\phi(\boldsymbol{x}, t)$ that is supposed to describe the *radiation* of an electromagnetic field from a source, we are likely to use an outgoing wave Green's function where if we are trying to describe the *absorption* of an electromagnetic field by a source, we are likely to use the incoming wave Green's function, while if we are looking for stationary (standing) waves in some sort of large spherical cavity coupled to a source near the middle then (you guessed it) the stationary wave Green's function is just perfect.

[As a parenthetical aside, you will often see people get carried away in the literature and connect the outgoing wave Green's function for the IHE to the retarded Green's function for the Wave Equation (fairly done – they are related by a contour integral as we shall see momentarily) and argue for a causal interpretation of the related integral equation solutions. However, as you can *clearly* see above, not only is there no breaking of time symmetry, the resulting descriptions are all just different ways of viewing the same solution! This isn't completely a surprise – the process of taking the Fourier transform symmetrically samples all of the past and all of the future when doing the time integral.

As we will see when discussing radiation reaction and causality at the very end of the semester, if anything one gets into trouble when one assumes that it is *always* correct to use an outgoing wave or retarded Green's function, as the actual field at any point in space at any point in time is time reversal invariant in classical electrodynamics – absorption and emission are mirror processes and *both* are simultaneously occurring when a charged particle is being accelerated by an electromagnetic field.]

11.2.3 Green's Function for the Wave Equation

This time we are interested in solving the *inhomogeneous wave equation* (IWE)

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\phi(\boldsymbol{x}, t) = -\frac{\rho(\boldsymbol{x}, t)}{\epsilon_0} \tag{11.52}$$

(for example) directly, without doing the Fourier transform(s) we did to convert it into an IHE.

Proceeding as before, we seek a Green's function that satisfies:

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)G(\boldsymbol{x}, t, \boldsymbol{x}_0, t_0) = \delta(\boldsymbol{x} - \boldsymbol{x}')\delta(t - t'). \tag{11.53}$$

The primary differences between this and the previous cases are a) the PDE is hyperbolic, not elliptical, if you have any clue as to what that means; b) it is now four dimensional – the "point source" is one that exists only at a single point in space for a single instant in time.

Of course this mathematical description leaves us with a bit of an existential dilemna, as physicists. We generally have little trouble with the idea of gradually restricting the support of a distribution to a single point in space by a limiting process. We just squeeze it down, mentally. However, in a supposedly conservative Universe, it is hard for us to imagine one of those squeezed down distributions of charge just "popping into existence" and then popping right out. We can't even do it via a limiting process, as it is a bit bothersome to create/destroy charge out of nothingness even gradually! We are left with the uncomfortable feeling that this particular definition is nonphysical in that it can describe *no* actual physical sources – it is by far the most "mathematical" or "formal" of the constructs we must use. It also leaves us with something to understand.

One way we can proceed is to view the Green's functions for the IHE as being the *Fourier transform of the desired Green's function here!* That is, we can exploit the fact that:

$$\delta(t - t_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega(t-t_0)}d\omega \tag{11.54}$$

to create a Fourier transform of the PDE for the Green's function:

$$\left(\nabla^2 + k^2\right) G(\boldsymbol{x}, \boldsymbol{x}_0, \omega) = \delta(\boldsymbol{x} - \boldsymbol{x}_0) e^{i\omega t_0} \tag{11.55}$$

(where I'm indicating the explicit ω dependence for the moment).

From the previous section we already know these solutions:

$$G_0(\boldsymbol{x}, \boldsymbol{x}_0, \omega) = \frac{-\cos(k|\boldsymbol{x} - \boldsymbol{x}_0|)}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} e^{i\omega t_0} \tag{11.56}$$

$$G_+(\boldsymbol{x}, \boldsymbol{x}_0, \omega) = \frac{-e^{+ik|\boldsymbol{x} - \boldsymbol{x}_0|}}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} e^{i\omega t_0} \tag{11.57}$$

$$G_-(\boldsymbol{x}, \boldsymbol{x}_0, \omega) = \frac{-e^{-ik|\boldsymbol{x} - \boldsymbol{x}_0|}}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} e^{i\omega t_0} \tag{11.58}$$

At this point in time [3] the only thing left to do is to Fourier transform back – to this point in time:

$$G_+(\boldsymbol{x}, t, \boldsymbol{x}_0, t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{-e^{+ik|\boldsymbol{x} - \boldsymbol{x}_0|}}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} e^{-i\omega(t - t_0)} d\omega \tag{11.59}$$

$$= \frac{1}{2\pi} \frac{-1}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} \int_{-\infty}^{\infty} -e^{+i\frac{\omega}{c}|\boldsymbol{x} - \boldsymbol{x}_0|} e^{-i\omega(t - t_0)} d\omega \tag{11.60}$$

$$= \frac{-1}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} \times \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} -\exp\left(-i\omega\left[(t - t_0) - \frac{|\boldsymbol{x} - \boldsymbol{x}_0|}{c}\right]\right) d\omega \right\} \tag{11.61}$$

$$= \frac{-\delta\left((t - t_0) - \frac{|\boldsymbol{x} - \boldsymbol{x}_0|}{c}\right)}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} \tag{11.62}$$

so that:

$$G_\pm(\boldsymbol{x}, t, \boldsymbol{x}_0, t_0) = \frac{-\delta\left((t - t_0) \mp \frac{|\boldsymbol{x} - \boldsymbol{x}_0|}{c}\right)}{4\pi|\boldsymbol{x} - \boldsymbol{x}_0|} \tag{11.63}$$

$$G_0(\boldsymbol{x}, t, \boldsymbol{x}_0, t_0) = \frac{1}{2}\left(G_+(\boldsymbol{x}, t, \boldsymbol{x}_0, t_0) + G_-(\boldsymbol{x}, t, \boldsymbol{x}_0, t_0)\right) \tag{11.64}$$

Note that when we set $k = \omega/c$, we basically asserted that the solution is being defined without dispersion! If there *is* dispersion, the Fourier transform will no longer neatly line up and yield a delta function, because the different Fourier components will not travel at the same speed. In that case one

[3] Heh, heh, heh...:-)

might still expect a peaked distribution, but not an *infinitely sharp* peaked distribution.

The first pair are generally rearranged (using the symmetry of the delta function) and presented as:

$$G^{(\pm)}(\boldsymbol{x}, t; \boldsymbol{x}', t') = \frac{\delta(t' - \left[t \mp \frac{|\boldsymbol{x}-\boldsymbol{x}'|}{c}\right]}{|\ vx - \boldsymbol{x}'\ |} \tag{11.65}$$

and are called the **retarded (+)** and **advanced (-)** Green's functions for the wave equation.

The second form is a very interesting beast. It is obviously a Green's function by construction, but it is a symmetric combination of advanced and retarded. Its use "means" that a field at any given point in space-time $(\boldsymbol{x}, t)$ consists of two pieces – one half of it is due to all the sources in space in the past such that the fields they emit are contracting precisely to the point $\boldsymbol{x}$ at the instant t and the other half is due to all of those *same* sources in space in the *future* such that the fields currently emerging from the point x at t precisely arrive at them. According to this view, the field at all points in space-time is as much due to the charges in the future as it is those same charges in the past.

Again it is worthwhile to note that *any actual field configuration* (solution to the wave equation) can be constructed from *any of these Green's functions* augmented by the addition of an *arbitrary bilinear solution* to the homogeneous wave equation (HWE) in primed and unprimed coordinates. We usually select the retarded Green's function as the "causal" one to simplify the way we think of an evaluate solutions as "initial value problems", not because they are any more or less causal than the others. Cause may precede effect in human perception, but as far as the equations of classical electrodynamics are concerned the concept of "cause" is better expressed as one of interaction via a suitable propagator (Green's function) that may well be time-symmetric or advanced.

A final note before moving on is that there are simply lovely papers (that we hope to have time to study) by Dirac and by Wheeler and Feynman that examine radiation reaction and the radiation field as constructed by advanced and retarded Green's functions in considerable detail. Dirac showed that the *difference* between the advanced and retarded Green's functions *at* the position of a charge was an important quantity, related to the *change* it made in the field presumably created by all the *other* charges in the Universe at that point in space and time. We have a lot to study here, in other words.

Using (say) the usual retarded Green's function, we could as usual write an integral equation for the solution to the general IWE above for e.g. $\boldsymbol{A}(\boldsymbol{x}, t)$:

$$\boldsymbol{A}(\boldsymbol{x}, t) = \chi_A(\boldsymbol{x}, t) - \mu_0 \int_V G_+(\boldsymbol{x}, t; \boldsymbol{x}', t)\boldsymbol{J}(\boldsymbol{x}', t')d^3x'dt' \tag{11.66}$$

where χ_A solves the HWE. This (with $\chi_A = 0$) is essentially equation (9.2), which is why I have reviewed this. Obviously we also have

$$\phi(\boldsymbol{x}, t) = \chi_\phi(\boldsymbol{x}, t) - \frac{1}{\epsilon_0} \int_V G_+(\boldsymbol{x}, t; \boldsymbol{x}', t)\rho(\boldsymbol{x}', t')d^3x'dt' \tag{11.67}$$

for $\phi(\boldsymbol{x}, t)$ (the minus signs are in the differential equations with the sources, note). You should formally verify that these solutions "work" given the definition of the Green's function above and the ability to reverse the order of differentiation and integration (bringing the differential operators, applied from the left, in underneath the integral sign).

Jackson proceeds from these equations by fourier transforming back into a k representation (eliminating time) and expanding the result to get to multipolar radiation at any given frequency. However, because of the way we proceeded above, we don't *have* to do this. We could just as easily start by working with the IHE instead of the IWE and use our HE Green's functions. Indeed, that's the plan, Stan...

11.3 Simple Radiating Systems

Let us start by writing the integral equation for the vector potential $\boldsymbol{A}(\boldsymbol{x})$ where we presume that we've already transformed the IWE into the IHE. We will choose to use the outgoing wave Green's function to make it clear that the field we are looking for is the one that the source is emitting, not one that it is absorbing.

$$\boldsymbol{A}(\boldsymbol{x}) = +\mu_0 \int \frac{e^{ik|\boldsymbol{x}-\boldsymbol{x}'|}}{4\pi|\boldsymbol{x}-\boldsymbol{x}'|}\boldsymbol{J}(\boldsymbol{x}')d^3x'. \tag{11.68}$$

There is no inhomogeneous term if there are no boundaries with *a priori* known boundary conditions.

Note that a more general solution would be one that allowed for absorption of incoming waves as well as the emission of outgoing waves, but that this would require knowing something about the sources outside the domain considered to be infinite. We will talk about this later (scattering theory and the optical theorem).

From $\boldsymbol{A}(\boldsymbol{x})$ we can easily find $\boldsymbol{B}$ or $\boldsymbol{H}$:

$$\boldsymbol{B} = \mu_0 \boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} \tag{11.69}$$

(by definition). Outside of the source, though (where the currents are all zero) Ampere's law tells us that:

$$\boldsymbol{\nabla} \times \boldsymbol{H} = -i\omega \boldsymbol{D} \tag{11.70}$$

or

$$\begin{aligned} \boldsymbol{\nabla} \times \boldsymbol{B} &= -i\omega\mu_0\epsilon_0 \boldsymbol{E} && (11.71) \\ \boldsymbol{\nabla} \times \boldsymbol{B} &= -i\frac{\omega}{c^2}\boldsymbol{E} = i\frac{k}{c}\boldsymbol{E} && (11.72) \end{aligned}$$

or

$$\boldsymbol{E} = i\frac{c}{k}\boldsymbol{\nabla} \times \boldsymbol{B} \tag{11.73}$$

Doing the integral above can be quite difficult in the general case. However, we'll find that for most *reasonable, physical situations* we will be able to employ certain approximations that will enable us to obtain a systematic hierarchy of descriptions that converge to the correct answer as accurately as you like, at the same time they increase our physical insight into the radiative processes.

11.3.1 The Zones

Suppose the source lives inside a region of maximum size $d \ll \lambda$ where $\lambda = 2\pi c/\omega$. By that I mean that a sphere of radius d (about the origin) completely contains all charge–current distributions. Then we can define three **zones** of approximation:

1. The near (static) zone $d << r << \lambda$
2. The intermediate (induction) zone $d << r \sim \lambda$
3. The far (radiation) zone $d << \lambda << r$

The field has very different properties in these zones. We will briefly discuss each of them.

- (Atomic and Molecular) sources all live inside their own near zone at optical frequencies. If the atoms are in a liquid or solid, there is a near field interaction (implicitly alluded to in chapter 4 and 7) that may be important in determining optical dispersion and other observable phenomena. Only for microwave frequencies and less does the near zone become relevant on a macroscopic scale. For rf waves it becomes extremely relevant, as it may extend a hundred meters or more.
- The induction zone is an annoying region where most of the simple approximations fail. It is distinguished by not being either of the other two zones. The wave changes character completely inside this zone. Because condensed matter theory invariably has objects interacting within this zone it is important there, although it can only be crudely treated. Without doing an obnoxious amount of work, that is.
- The far zone is where we all live, most of the time, with respect to the major sources of EM radiation. Our detectors are many wavelengths away from the atoms. Our radios are many wavelengths away from the transmitters. And, generally, the sources are smaller, if not much smaller, than a wavelength[4]. In the far zone, the emitted EM fields are characteristically transverse and fall off in amplitude as $1/r$ or faster, and often far *enough* away they look locally like plane waves! This is typical of radiation fields from compact sources. We will spend most of our time considering solutions in the far zone.

11.3.2 The Near Zone

Suppose that we are in the **near zone**. Then by definition

$$k|\boldsymbol{x} - \boldsymbol{x}'| << 1$$

[4]We will learn to treat certain exceptions, believe me.

and

$$e^{ik|\boldsymbol{x}-\boldsymbol{x}'|} \approx 1$$

This makes the integral equation into the "static" form already considered in chapter 5 (cf. equation (5.32)). We see that $-1/4\pi|\boldsymbol{x}-\boldsymbol{x}'|$ is just the Green's function for the good old Poisson equation in this approximation and can be expanded in harmonic functions just like in the good old days:

$$G_0(\boldsymbol{x},\boldsymbol{x}') = \sum_L \frac{-1}{2\ell+1}\frac{r'^{\ell}}{r^{\ell+1}} Y_L(\hat{r}) Y_L(\hat{r'})^*. \tag{11.74}$$

Note Well: I will use $L \equiv (\ell, m)$ freely and without warning in this course. The sum is over all ℓ, m. Hopefully, by now you know what they run over. If not, read the chapter in Wyld on spherical harmonics and review Jackson as well. **This is important!**

This means that (if you like)

$$\lim_{kr\to 0} \boldsymbol{A}(\boldsymbol{x}) = \sum_L \frac{1}{(2\ell+1)r^{\ell+1}} Y_L(\hat{r}) \int \boldsymbol{J}(\boldsymbol{x}') r'^{\ell} Y_L(\hat{r'})^* d^3r'. \tag{11.75}$$

We will use expressions like this (derived from the multipolar expansion of the Green's function) frequently in what follows. For that reason I suggest that you study it carefully and be sure you understand it.

Since (for fixed r outside the source)

$$\lim_{k\to 0} \to \lim_{kr\to 0}$$

we see that this limit is reached (among other times) when

$$k \to 0$$

(relative to the size of the source and point of measurement)! But then the IHE turns back into the Poisson equation (or inhomogeneous Laplace equation, ILE) as it should, come to think about it. The near fields oscillate harmonically in time, but are *spatially* identical to the fields produced by a "static" current with the given spatial distribution. That's why we also call the near zone the "static zone".

11.3.3 The Far Zone

Exactly the opposite is true in the **far zone**. Here $kr >> 1$ and the exponential oscillates *rapidly*. We can approximate the argument of the exponential as follows:

$$\begin{aligned} |\boldsymbol{x} - \boldsymbol{x}'| &= \sqrt{r^2 + r'^2 - 2r\mathbf{n}\cdot\boldsymbol{x}'} \\ &= r\left\{1 - \frac{2}{r}\mathbf{n}\cdot\boldsymbol{x}' + \frac{r'^2}{r^2}\right\}^{1/2} \\ &= r - \mathbf{n}\cdot\boldsymbol{x}' + \mathcal{O}\left(\frac{1}{r}\right) \end{aligned} \tag{11.76}$$

where we have assumed that $r'_{\rm max} < d << r$ and used a binomial expansion of the root sum. We neglect higher order terms. Note that this approximation is good *independent* of k and may be good even in the near zone.

Then

$$\lim_{(k)r\to\infty} vA(\boldsymbol{x}) = \frac{\mu_0 e^{ikr}}{4\pi r}\int \boldsymbol{J}(\boldsymbol{x}')e^{-ik\hat{\boldsymbol{n}}\cdot\boldsymbol{x}'}d^3x'. \tag{11.77}$$

In the far zone, the solution behaves like an outgoing spherical wave times an amplitude that depends on integral over the source that depends on angles in an intricate fashion.

At this point I could continue and extract

$$\lim_{(k)r\to\infty} vA(\boldsymbol{x}) = \frac{\mu_0 e^{ikr}}{4\pi r}\sum_n \frac{(-ik)^n}{n!}\int \boldsymbol{J}(\boldsymbol{x}')(\hat{\boldsymbol{n}}\cdot\boldsymbol{x}')^n d^3x' \tag{11.78}$$

(if the source is actually small enough to allow expansion of the exponential in a series[5]). This would give us a cheap introduction into multipoles. But it is so *sloppy!*

Instead we are going to do it right. We will begin by reviewing the solutions to the *homogeneous* Helmholtz equation (which should really be discussed before we sweat solving the *inhomogeneous* equation, don't you think?) and will construct the *multipolar expansion* for the outgoing and incoming (and stationary) wave Green's function. Using this, it will be a trivial matter to write down a formally exact *and* convergent solution to the integral equation *on all space* that we can chop up and approximate as we please. This will provide a much more natural (and accurate) path to multipolar radiation. So let's start.

[5]Taylor? Power? Laurent? Who can remember...

11.4 The Homogeneous Helmholtz Equation

Recall as you read this that WIYF and MWIYFin addition to the treatment of this available in Jackson, chapters 2, 3, 6, and 8 of Wyld, and doubtless Arfkin, Morse and Feshback, and probably six other sources if you look. Very important stuff, can't know it too well.

Recall from above the **Homogeneous Helmholtz Equation** (HHE):

$$(\nabla^2 + k^2)\chi(\boldsymbol{x}) = 0. \tag{11.79}$$

We assume that[6]:

$$\chi(\boldsymbol{x}) = \sum_L f_\ell(r) Y_L(\theta, \phi). \tag{11.80}$$

We reduce the HHE with this assumption to the **radial differential equation**

$$\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} + k^2 - \frac{\ell(\ell+1)}{r^2}\right] f_\ell(r) = 0. \tag{11.81}$$

If we substitute

$$f_\ell(r) = \frac{1}{r^{1/2}} u_\ell(r) \tag{11.82}$$

we transform this into an equation for $u_\ell(r)$,

$$\left[\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} + k^2 - \frac{(\ell+\frac{1}{2})^2}{r^2}\right] u_\ell(r) = 0. \tag{11.83}$$

The is *Bessel's differential equation.* See Wyld, (2-6) or Jackson in various places (see key on back inside cover) for more detail. Or your own favorite Math Physics book.

Two linearly independent solutions on $\mathbb{R}^3$ minus the origin to this radial DE are:

$$\begin{aligned} f_\ell(r) &= j_\ell(kr) \quad \text{and} & (11.84)\\ f_\ell(r) &= n_\ell(kr), & (11.85)\end{aligned}$$

[6]This really isn't an assumption. We could equally well write ∇^2 in spherical polar coordinates, separate variables, note that the angular ODEs have spherical harmonics as eigenstates ("quantized" by the requirement of single-valuedness on e.g. rotations of 2π in ϕ) and reconstruct the separated solution. But that's too much work and we already did it at least once in our lives, right? So we'll "assume".

the *spherical bessel function* and *spherical neumann functions* respectively. They are both real, and hence are stationary in time (why?). The $j_\ell(kr)$ are **regular** (finite) at the origin while the $n_\ell(kr)$ are **irregular** (infinite) at the origin. This is in exact analogy with the situation for the homogeneous Laplace equation (which is a special case of this solution).

The following is a **MINIMAL** table of their important properties. A better table can be found in Wyld between chps. 6 and 7 and in Morse and Feshbach (I can't remember which volume).

11.4.1 Properties of Spherical Bessel Functions

Recursion Relation

Let $z_\ell(x)$ be either solution or a linear combination of the two. x is a complex scalar independent variable (in practice, $x = kr$). Then

$$z_{\ell+1}(x) = \frac{2\ell+1}{x} z_\ell(x) - z_{\ell-1}(x). \tag{11.86}$$

This relation is stable for increasing ℓ for $z_\ell = n_\ell$. It is stable for decreasing ℓ for $z_\ell = j_\ell$. For that reason it is *un*stable in both directions for $h_\ell^\pm$ (defined below). How would you make it? See Abramowitz and Stegun, *Handbook of Mathmatical Functions* for discussion of recursive algorithm and definition of power series expansions.

The Lowest Few Functions

$$j_0(x) = \frac{\sin(x)}{x} \tag{11.87}$$

$$j_1(x) = \frac{\sin(x)}{x^2} - \frac{\cos(x)}{x} \tag{11.88}$$

$$\vdots$$

$$n_0(x) = -\frac{\cos(x)}{x} \tag{11.89}$$

$$n_1(x) = -\frac{\cos(x)}{x^2} - \frac{\sin(x)}{x} \tag{11.90}$$

$$\vdots$$

Asymptotic Forms

Small x:

$$\lim_{x \to 0} j_\ell(x) = \frac{2^\ell \ell!}{(2\ell+1)!} x^\ell \tag{11.91}$$

$$\lim_{x \to 0} n_\ell(x) = -\frac{(2\ell)!}{2^\ell \ell!} \frac{1}{x^{\ell+1}}. \tag{11.92}$$

Note that for small x ($r << k$) $j_\ell(kr)$ is proportional to r^ℓ and $n_\ell(kr)$ is proportional to $1/r^{\ell+1}$, which are the regular and irregular solutions to the separated Laplace equation. This is the correct way to obtain the static limit.

Large x:

$$\lim_{x \to \infty} j_\ell(x) = \frac{1}{x} \cos(x - (\ell+1)\frac{\pi}{2}) \tag{11.93}$$

$$\lim_{x \to \infty} n_\ell(x) = \frac{1}{x} \sin(x - (\ell+1)\frac{\pi}{2}). \tag{11.94}$$

Note that both solutions are regular (go to zero smoothly) at infinity and are the same (trig) function shifted by $\pi/2$ over x there. Note that they are *not* square integrable on $\mathbb{R}^3$ (for your quantum course) but are still better than plane waves in that regard. Something to think about ...

Hankel Functions

Examining the asymptotic forms, we see that two particular complex linear combinations of the stationary solution have the behavior, at infinity, of an outgoing or incoming spherical wave when the time dependence is restored:

$$h_\ell^+(x) = j_\ell(x) + i n_\ell(x) \quad (= h_\ell^1(x)) \tag{11.95}$$

$$h_\ell^-(x) = j_\ell(x) - i n_\ell(x) \quad (= h_\ell^2(x)) \tag{11.96}$$

the **spherical hankel functions** of the **first** (+) (outgoing) and **second** (−) (incoming) kinds. *Both* of these solutions are singular at the origin like $1/x^{\ell+1}$ (why?) and behave like

$$\lim_{x \to \infty} h_\ell^+(x) = (-i)^{\ell+1} \frac{e^{ix}}{x} \tag{11.97}$$

$$\lim_{x \to \infty} h_\ell^-(x) = (i)^{\ell+1} \frac{e^{-ix}}{x} \tag{11.98}$$

at infinity. Two particularly useful spherical hankel functions to know are the zeroth order ones:

$$h_0^+(x) = \frac{e^{ix}}{ix} \tag{11.99}$$

$$h_0^-(x) = \frac{e^{-ix}}{-ix} \tag{11.100}$$

Plane Wave Expansion

Plane waves and free spherical waves both form an (on–shell) complete orthnormal set on $\mathbb{R}^3$ (with or without the origin). That means that one must be able to expand one in terms of the other. Plane waves can be expanded in terms of free spherical waves by:

$$\begin{aligned} e^{i\mathbf{k}\cdot\mathbf{r}} &= e^{ikr\cos(\Theta)} \\ &= \sum_L 4\pi i^\ell Y_L(\hat{k}) j_\ell(kr) Y_L(\hat{r})^*. \end{aligned} \tag{11.101}$$

This is due to Lord Rayleigh and is sometimes called the Rayleigh expansion. Recall that Θ is the angle betwixt the $\vec{r}$ and the $\vec{k}$ and that $\cos(\Theta) = \cos(-\Theta)$.

There is similarly an (integral) expression for $j_\ell(kr)$ in terms of an integral over the $e^{i\mathbf{k}\cdot\mathbf{r}}$ but we will not use it here. It follows from the completeness relation on page 214 in Wyld, the Rayleigh expansion, and the completeness relation on page 212. Derive it for homework (or find it somewhere and copy it, but you shouldn't have to). Check your result by finding it somewhere. I think it might be somewhere in Wyld, but I know it is elsewhere. This will be handed in.

11.4.2 $J_L(\mathbf{r})$, $N_L(\mathbf{r})$, and $H_L^\pm(\mathbf{r})$

For convenience, we **define** the following:

$$J_L(\mathbf{r}) = j_\ell(kr) Y_L(\hat{r}) \tag{11.102}$$

$$N_L(\mathbf{r}) = n_\ell(kr) Y_L(\hat{r}) \tag{11.103}$$

$$H_L^\pm(\mathbf{r}) = h_\ell^\pm(kr) Y_L(\hat{r}) \tag{11.104}$$

These are the basic solutions to the HHE that are also eigenfunctions of L^2 and L_z. Clearly there is an **implicit** label of k (or k^2) for these solutions.

A general solution (on a suitable domain) can be constructed out of a linear combination of any two of them.

11.4.3 General Solutions to the HHE

On "spherical" domains (the interior and exterior of a sphere, or in a spherical shell) the completely general solution to the HHE can therefore be written in *stationary* form as:

$$\sum_L A_L J_L(\mathbf{r}) + B_L N_L(\mathbf{r}) \tag{11.105}$$

or (for scattering theory, mostly) in the *outgoing wave* form

$$\sum_L C_L J_L(\mathbf{r}) + S_L H_L^+(\mathbf{r}). \tag{11.106}$$

Inside a sphere, B_L and S_L must be zero. Outside a sphere, or in a spherical annulus, all the coefficients can be non–zero *unlike* the situation for the Laplace equation (why?).

[This should provoke deep thoughts about the fundamental significance of the Laplace equation. Are there any "really" stationary sources in the dynamical, covariant, universe? Do we expect to have a contribution to the zero frequency charge/current density distribution in any region of space? What would this correspond to?]

11.4.4 Green's Functions and Free Spherical Waves

We expect, for physical reasons[7] that the wave **emitted** by a time dependent source should behave like an outgoing wave far from the source. Note that inside the bounding sphere of the source that need not be true. Earlier in this chapter, we used an "outgoing wave Green's function" to construct the solution to the IHE with this asymptotic behavior. Well, lo and behold:

$$G_{\pm}(\boldsymbol{x}, \boldsymbol{x}') = \mp \frac{ik}{4\pi} h_0^{\pm}(k|\boldsymbol{x} - \boldsymbol{x}'|) \tag{11.107}$$

[7]A cop–out phrase if there ever was one. It translates as: because that's the way it turns out at the end.

For stationary waves (useful in quantum theory)

$$G_0(\boldsymbol{x}, \boldsymbol{x}') = \frac{k}{4\pi} n_0(k|\boldsymbol{x} - \boldsymbol{x}'|). \tag{11.108}$$

This extremely important relation forms the connection between free spherical waves (reviewed above) and the integral equation solutions we are interested in constructing.

This connection follows from the *addition theorems* or *multipolar expansions* of the free spherical waves defined above. For the special case of $L = (0, 0)$ these are:

$$N_0(\mathbf{r} - \mathbf{r}') = n_0(k|\mathbf{r} - \mathbf{r}')\frac{1}{\sqrt{4\pi}} = \sqrt{4\pi}\sum_L N_L(\mathbf{r}_>)J_L(\mathbf{r}_<)^* \tag{11.109}$$

and

$$H_0^{\pm}(\mathbf{r} - \mathbf{r}') = h_0^{\pm}(k|\mathbf{r} - \mathbf{r}')\frac{1}{\sqrt{4\pi}} = \sqrt{4\pi}\sum_L H_L^{\pm}(\mathbf{r}_>)J_L(\mathbf{r}_<)^*. \tag{11.110}$$

From this and the above, the expansion of the Green's functions in free spherical multipolar waves immediately follows:

$$G_0(\mathbf{r} - \mathbf{r}') = k\sum_L N_L(\mathbf{r}_>)J_L(\mathbf{r}_<)^* \tag{11.111}$$

and

$$G_{\pm}(\mathbf{r} - \mathbf{r}') = \mp ik\sum_L H_L^{\pm}(\mathbf{r}_>)J_L(\mathbf{r}_<)^*. \tag{11.112}$$

Note Well: The complex conjugation operation under the sum is applied to the *spherical harmonic* (only), *not* the Hankel function(s). This is because the only function of the product $Y_L(\hat{\boldsymbol{r}})Y_L(\hat{\boldsymbol{r}}')^*$ is to reconstruct the $P_\ell(\Theta)$ via the addition theorem for spherical harmonics. Study this point in Wyld carefully on your own.

These relations will allow us to expand the Helmholtz Green's functions *exactly like* we expanded the Green's function for the Laplace/Poisson equation. This, in turn, will allow us to precisely and beautifully reconstruct the multipolar expansion of the vector potential, and hence the EM fields in the various zones *exactly*[8].

This ends our brief mathematical review of free spherical waves and we return to the description of Radiation.

[8]Well, in a uniformly convergent expansion, which is kind of exact, in the limit of an infinite sum. In the mean time, it is a damn good approximation. Usually.

11.5 Electric Dipole Radiation

Now that we have that under our belts we can address the multipolar expansion of the vector potential intelligently. To begin with, we will write the **general solution** for the vector potential in terms of the multipolar expansion for the outgoing wave Green's function defined above:

$$\begin{aligned}\boldsymbol{A}(\boldsymbol{r}) &= ik\sum_L \Big[J_L(\boldsymbol{r})\int_r^\infty \mu_0 \boldsymbol{J}(\boldsymbol{r}')H_L(\boldsymbol{r}')^{(*)}d^3r' \\ &\quad + H_L^+(\boldsymbol{r})\int_0^r \mu_0 \boldsymbol{J}(\boldsymbol{r}')J_L(\boldsymbol{r}')^{(*)}d^3r'\Big] \end{aligned} \tag{11.113}$$

where, by convention, $(*)$ means that the $Y_L(\hat{r})$ is conjugated but the bessel/neumann/hankel function is *not*. This is because the only point of the conjugation is to construct $P_\ell(\Theta)$ from the m–sum for each ℓ via the addition theorem for spherical harmonics. We certainly *don't* want to change h^+ into h^-, which changes the time dependent behavior of the solution[9]. Note that the integral over all space is broken up in such a way that the Green's function expansions above always converge. This solution is exact everywhere in space *including inside the source itself!*

We can therefore simplify our notation by defining certain functions of the radial variable:

$$\boldsymbol{A}(\boldsymbol{r}) = \sum_L ik\left[C_L(r)J_(\boldsymbol{r}) + S_L(r)H_L^+(\boldsymbol{r})\right]. \tag{11.114}$$

In this equation,

$$C_L(r) = \int_r^\infty \mu_0 \boldsymbol{J}(\boldsymbol{r}')H_L(\boldsymbol{r}')^{(*)}d^3r' \tag{11.115}$$

$$S_L(r) = \int_0^r \mu_0 \boldsymbol{J}(\boldsymbol{r}')J_L(\boldsymbol{r}')^{(*)}d^3r'. \tag{11.116}$$

Clearly $S_L(0) = 0$ and for $r > d$, $C_L(r) = 0$. At the origin the solution is completely regular and *stationary*. Outside the bounding sphere of the source distribution the solution behaves like a linear combination of outgoing spherical multipolar waves. From now on we will concentrate on the latter case, since it is the one relevant to the zones.

[9]This suggests that there are some interesting connections between the conjugation symmetry and time reversal symmetry. Too bad we won't have time to explore them. You may on your own, though.

11.5.1 Radiation outside the source

Outside the bounding sphere of the source,

$$\boldsymbol{A}(\boldsymbol{r}) = ik \sum_L H_L^+(\boldsymbol{r}) \int_0^\infty \mu_0 \boldsymbol{J}(\boldsymbol{r}') J_L(\boldsymbol{r}')^{(*)} d^3r'. \tag{11.117}$$

At last we have made it to Jackson's equation 9.11, but *look how elegant our approach was.* Instead of a form that is only valid in the far zone, we can now see that this is a limiting form of a convergent solution that works in *all* zones, including inside the source itself! The integrals that go into the $C_L(r)$ and $S_L(r)$ may well be daunting to a person armed with pen and paper (depending on how nasty $\boldsymbol{J}(\boldsymbol{x}')$ is) but they are very definitely computable with a computer!

Now, we must use several interesting observations. First of all, $J_L(\boldsymbol{r})$ gets small rapidly inside d as ℓ increases (beyond kd). This is the **angular momentum cut–off** in disguise and you should remember it. This means that if $\boldsymbol{J}(\mathbf{r})$ is sensibly bounded, the integral on the right (which is cut off at $r' = d$) will get small for "large" ℓ. In most cases of physical interest, $kd << 1$ by hypothesis and we need only keep the first few terms (!). In practically all of these cases, the lowest order term ($\ell = 0$) will yield an excellent approximation. This term produces the **electric dipole** radiation field.

11.5.2 Dipole Radiation

Let us evaluate this term. It is (c. f. J9.13):

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu_0 e^{ikr}}{4\pi r} \int_0^r \boldsymbol{J}(\boldsymbol{r}') d^3r' \tag{11.118}$$

(note: $Y_{00}(\hat{r}) = Y_{00}(\hat{r})^* = 1/\sqrt{4\pi}$). If we integrate this term by parts (a surprisingly difficult chore that will be an exercise) and use the continuity equation and the fact that the source is harmonic we get:

$$\boldsymbol{A}(\boldsymbol{r}) = -\frac{i\mu_0\omega}{4\pi} \boldsymbol{p} \frac{e^{ikr}}{r} \tag{11.119}$$

where

$$\boldsymbol{p} = \int \boldsymbol{r}' \rho(\boldsymbol{r}') d^3r' \tag{11.120}$$

is the **electric dipole moment** (see J4.8). Note that if we define $\rho(\boldsymbol{r})$ to be a "probability density" for the electrons during a transition this expression is still valid.

This is wonderfully simple. If only we could quit with the vector potential. Alas, no. We must reconstruct the *electromagnetic field* being radiated away from the source from the expressions previously given

$$\mathbf{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$$

and

$$\mathbf{E} = \frac{ic}{k}\boldsymbol{\nabla} \times \mathbf{B}.$$

After a tremendous amount of straightforward but nonetheless difficult algebra that you will do and hand in next week (see problems) you will obtain:

$$\mathbf{H} = \frac{ck^2}{4\pi}(\mathbf{n} \times \boldsymbol{p})\frac{e^{ikr}}{r}\left(1 - \frac{1}{ikr}\right) \tag{11.121}$$

and

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\left\{k^2(\mathbf{n} \times \boldsymbol{p}) \times \mathbf{n}\frac{e^{ikr}}{r} + [3\mathbf{n}(\mathbf{n} \cdot \boldsymbol{p}) - \boldsymbol{p}]\left(\frac{1}{r^3} - \frac{ik}{r^2}\right)e^{ikr}\right\} \tag{11.122}$$

The magnetic field is always transverse to the radial vector. **Electric dipole** radiation is therefore also called **transverse magnetic** radiation. The electric field is transverse in the far zone, but in the near zone it will have a component (in the $\boldsymbol{p}$ direction) that is not generally perpendicular to $\mathbf{n}$.

Asymptotic properties in the Zones

In the **near zone** we get:

$$\boldsymbol{B} = \mu_0\boldsymbol{H} = \frac{i\omega\mu_0}{4\pi}(\hat{\boldsymbol{n}} \times \boldsymbol{p})\frac{1}{r^2} \tag{11.123}$$

$$\boldsymbol{E} = \frac{1}{4\pi\epsilon_0}[3\hat{\boldsymbol{n}}(\hat{\boldsymbol{n}} \cdot \boldsymbol{p}) - \boldsymbol{p}]\frac{1}{r^3} \tag{11.124}$$

and can usually neglect the magnetic field relative to the electric field (it is smaller by a factor of $kr << 1$). The electric field is that of a "static" dipole (J4.13) oscillating harmonically.

In the **far zone** we get:

$$\begin{aligned} \boldsymbol{B} &= \mu_0 \boldsymbol{H} = \frac{ck^2\mu_0}{4\pi}\left(\hat{\boldsymbol{n}} \times \boldsymbol{p}\right)\frac{e^{ikr}}{r} && (11.125) \\ \boldsymbol{E} &= \frac{ic}{k}\boldsymbol{\nabla} \times \boldsymbol{B} = c\left(\boldsymbol{B} \times \hat{\boldsymbol{n}}\right). && (11.126) \end{aligned}$$

This is transverse EM radiation. Expanded about any point, it looks just like a plane wave (which is how "plane waves" are born!). We are most interested, as you know, in the radiation zone and so we will focus on it for a moment.

Energy radiated by the dipole

Recall our old buddy the **complex Poynting vector** for harmonic fields (J6.132):

$$\boldsymbol{S} = \frac{1}{2}\mathrm{Re}\left\{\boldsymbol{E} \times \boldsymbol{H}^*\right\}. \tag{11.127}$$

The factor of 1/2 comes from time averaging the fields. This is the energy per unit area per unit time that passes a point in space. To find the time average power per solid angle, we must relate the normal area through which the energy flux passes to the solid angle:

$$dA_n = r^2 d\Omega \tag{11.128}$$

and project out the appropriate piece of $\mathbf{S}$, i. e. — $\mathbf{n}\cdot\mathbf{S}$. We get (with $\mu = 1$)

$$\frac{dP}{d\Omega} = \frac{1}{2}\mathrm{Re}[r^2\mathbf{n}\cdot(\mathbf{E}\times\mathbf{H}^*)]. \tag{11.129}$$

where we must plug in $\mathbf{E}$ and $\mathbf{H}$ from the expressions above for the far field.

After a bunch of algebra that I'm sure you will enjoy doing, you will obtain:

$$\frac{dP}{d\Omega} = \frac{c^2}{32\pi^2}\sqrt{\frac{\mu_0}{\epsilon_0}}k^4 \mid (\mathbf{n} \times \boldsymbol{p}) \times \mathbf{n} \mid^2. \tag{11.130}$$

The polarization of the radiation is determined by the vector inside the absolute value signs. By this one means that one can project out each component of $\boldsymbol{p}$ (and hence the radiation) before evaluating the square independently, if so desired. Note that the different components of $\boldsymbol{p}$ need not have the same phase (elliptical polarization, etc.).

If all the components of $\boldsymbol{p}$ (in some coordinate system) have the same phase, then $\boldsymbol{p}$ necessarily lies along a line and the typical angular distribution is that of (linearly polarized) **dipole radiation**:

$$\frac{dP}{d\Omega} = \frac{c^2}{32\pi^2}\sqrt{\frac{\mu_0}{\epsilon_0}}k^4 \mid \boldsymbol{p} \mid^2 \sin^2\theta \tag{11.131}$$

where θ is measured between $\mathbf{p}$ and $\mathbf{n}$. When you integrate over the entire solid angle (as part of your assignment) you obtain the total power radiated:

$$P = \frac{c^2k^4}{12\pi}\sqrt{\frac{\mu_0}{\epsilon_0}} \mid \boldsymbol{p} \mid^2 \tag{11.132}$$

The most important feature of this is the k^4 dependence which is, after all, why the sky is blue (as we shall see, never fear).

Example: A centerfed, linear antenna

In this antenna, $d << \lambda$ and

$$I(z,t) = I_0\left(1 - \frac{2 \mid z \mid}{d}\right)e^{-i\omega t}. \tag{11.133}$$

From the continuity equation (and a little subtle geometry),

$$\boldsymbol{\nabla}\cdot\boldsymbol{J} = \frac{dI}{dz} = -\frac{\partial\rho'(z)e^{-i\omega t}}{\partial t} = i\omega\rho'(z) \tag{11.134}$$

and we find that the linear charge density (participating in the oscillation, with a presumed neutral background) is independent of z:

$$\rho'(z) = \pm\frac{2iI_0}{\omega d} \tag{11.135}$$

where the $+/-$ sign indicates the upper/lower branch of the antenna and the $'$ means that we are really treating $\rho/(dxdy)$ (which cancels the related terms in the volume integral below). We can then evaluate the dipole moment of the entire antenna for this frequency:

$$p_z = \int_{-d/2}^{d/2} z\rho'(z)dz = \frac{iI_0 d}{2\omega}. \tag{11.136}$$

The electric and magnetic fields for $r > d$ in the electric dipole approximation are now given by the previously derived expressions. The angular distribution of radiated power is

$$\frac{dP}{d\Omega} = \frac{I_0^2}{128\pi^2}\sqrt{\frac{\mu_0}{\epsilon_0}}(kd)^2 \sin^2\theta \tag{11.137}$$

and the total radiated power is

$$P = \frac{I_0^2(kd)^2}{48\pi}\sqrt{\frac{\mu_0}{\epsilon_0}}. \tag{11.138}$$

Remarks. For fixed current the power radiated increases as the square of the frequency (at least when $kd << 1$, i. e. – long wavelengths relative to the size of the antenna). The total power radiated by the antenna appears as a "loss" in "Ohm's Law" for the antenna. Factoring out $I_0^2/2$, the remainder must have the units of resistance and is called the **radiation resistance** of the antenna:

$$R_{\mathrm{rad}} = \frac{2P}{I_0^2} = \frac{(kd)^2}{24\pi}\sqrt{\frac{\mu_0}{\epsilon_0}} \approx 5(kd)^2 \quad \text{ohms}) \tag{11.139}$$

where we do the latter multiplication to convert the resulting units to ohms. Note that this resistance is there for harmonic currents even if the conductivity of the metal is perfect. Note further that by hypothesis this expression will only be valid for small values of R_{rad}.

Good golly, this is wonderful. We hopefully really understand electric dipole radiation at this point. It would be truly sublime if all radiators were dipole radiators. Physics would be so easy. But (alas) sometimes the current distribution has **no $\ell = 0$ moment** and there is therefore **no dipole term**! In that case we must look at the next term or so in the multipolar expansions

Lest you think that this is a wholly unlikely occurrance, please note that a humble loop carrying a current that varies harmonically is one such system. So let us proceed to:

11.6 Magnetic Dipole and Electric Quadrupole Radiation Fields

The next term in the multipolar expansion is the $\ell = 1$ term:

$$\mathbf{A}(\boldsymbol{x}) = ik\mu_0 h_1^+(kr) \sum_{m=-1}^{1} Y_{1,m}(\hat{r}) \int_0^\infty \boldsymbol{J}(\mathbf{x}') j_1(kr') Y_{1,m}(\hat{r'})^* d^3x' \qquad (11.140)$$

When you (for homework, of course)

1. m–sum the product of the $Y_{\ell,m}$'s
2. use the small kr expansion for $j_1(kr')$ in the integral and combine it with the explicit form for the resulting $P_1(\theta)$ to form a dot product
3. cancel the $2\ell + 1$'s
4. explicitly write out the hankel function in exponential form

you will get equation (J9.30, for – recall – distributions with compact support):

$$\mathbf{A}(\boldsymbol{x}) = \frac{\mu_0}{4\pi}\frac{e^{ikr}}{r}\left(\frac{1}{r} - ik\right)\int_0^\infty \boldsymbol{J}(\mathbf{x}')(\mathbf{n}\cdot\mathbf{x}')d^3x'. \qquad (11.141)$$

Of course, you can get it directly from J9.9 (to a lower approximation) as well, but that does *not* show you what to do if the small kr approximation is not valid (in step 2 above) and it neglects part of the outgoing wave!

There are two important and independent pieces in this expression. One of the two pieces is symmetric in $\boldsymbol{J}$ and $\boldsymbol{x}'$ and the other is antisymmetric (get a minus sign when the coordinate system is inverted). Any vector quantity can be decomposed in this manner so this is a very general step:

$$\boldsymbol{J}(\mathbf{n}\cdot\mathbf{x}') = \frac{1}{2}[(\mathbf{n}\cdot\mathbf{x}')\boldsymbol{J} + (\mathbf{n}\cdot\boldsymbol{J})\mathbf{x}'] + \frac{1}{2}(\mathbf{x}'\times\boldsymbol{J})\times\mathbf{n}. \qquad (11.142)$$

11.6.1 Magnetic Dipole Radiation

Let's look at the antisymmetric bit first, as it is somewhat simpler and we can leverage our existing results. The second term is the magnetization (density) due to the current $\boldsymbol{J}$:

$$\vec{\mathcal{M}} = \frac{1}{2}(\boldsymbol{x} \times \boldsymbol{J}) \tag{11.143}$$

(see J5.53, 5.54) so that

$$\boldsymbol{m} = \int \vec{\mathcal{M}}(\mathbf{x}')d^3x' \tag{11.144}$$

where $\boldsymbol{m}$ is the magnetic dipole moment of the (fourier component of) the current.

Considering only this antisymmetric term, we see that:

$$\boldsymbol{A}_{\mathrm{M1}}(\boldsymbol{x}) = \frac{ik\mu_0}{4\pi}(\mathbf{n} \times \mathbf{m})\frac{e^{ikr}}{r}\left(1 - \frac{1}{ikr}\right). \tag{11.145}$$

HMMMMMMM, (you had better say)! This looks "just like" the expression for the magnetic field that resulted from the electric dipole vector potential. Sure enough, when you (for homework) crank out the algebra, you will show that

$$\boldsymbol{B} = \frac{\mu_0}{4\pi}\left\{k^2(\mathbf{n} \times \mathbf{m}) \times \mathbf{n}\frac{e^{ikr}}{r} + [3\mathbf{n}(\mathbf{n} \cdot \mathbf{m}) - \mathbf{m}]\left(\frac{1}{r^3} - \frac{ik}{r^2}\right)e^{ikr}\right\} \tag{11.146}$$

and

$$\boldsymbol{E} = -\frac{1}{4\pi}\sqrt{\frac{\mu_0}{\epsilon_0}}k^2(\mathbf{n} \times \mathbf{m})\frac{e^{ikr}}{r}\left(1 - \frac{1}{ikr}\right). \tag{11.147}$$

Clearly, we don't need to discuss the behavior of the fields in the zones since they are completely analogous. The electric field is always transverse, and the total field arises from a harmonic magnetic dipole. For this reason, this kind of radiation is called either **magnetic dipole** (M1) radiation or **transverse electric** radiation. For what it's worth, electric dipole radiation is also called (E1) radiation.

However, this is only ONE part of the contribution from $\ell = 1$ terms in the Green's function expansion. What about the other (symmetric) piece? Oooo, ouch.

11.6.2 Electric Quadrupole Radiation

Now let's to untangle the first (symmetric) piece. This will turn out to be a remarkably unpleasant job. In fact it is my nefarious and sadistic plan that it be *so* unpleasant that it properly motivates a change in approach to one that handles this nasty tensor stuff "naturally".

We have to evaluate the integral of the symmetric piece. We get:

$$\frac{1}{2}\int [(\hat{\boldsymbol{n}} \cdot \boldsymbol{x}')\boldsymbol{J} + (\hat{\boldsymbol{n}} \cdot \boldsymbol{J})\boldsymbol{x}']d^3x' = -\frac{i\omega}{2}\int \boldsymbol{x}'(\hat{\boldsymbol{n}} \cdot \boldsymbol{x}')\rho(\boldsymbol{x}')d^3x' \tag{11.148}$$

The steps involved are:

1. integrate by parts (working to obtain divergences of $\boldsymbol{J}$).
2. changing $\boldsymbol{\nabla} \cdot \boldsymbol{J}$ into a ρ times whatever from the continuity equation (for a harmonic source).
3. rearranging and recombining.

Don't forget the boundary condition at infinity ($\boldsymbol{J}$ and ρ have compact support)! You'll love doing this one...

The vector potential is thus:

$$\boldsymbol{A}_{\mathrm{E2}}(\boldsymbol{x}) = -\frac{\mu_0 ck^2}{8\pi}\frac{e^{ikr}}{r}\left(1 - \frac{1}{ikr}\right)\int \boldsymbol{x}'(\hat{\boldsymbol{n}} \cdot \boldsymbol{x}')\rho(\boldsymbol{x}')d^3x'. \tag{11.149}$$

Note that $\boldsymbol{x}'$ appears *twice* under the integral, and that its vector character similarly appears twice: once in $\boldsymbol{x}'$ itself and once in its projection on $\hat{\boldsymbol{n}}$. The integral is the **electric quadrupole moment** of the oscillating charge density distribution and the resulting radiation field is called an **electric quadrupole (radiation) field** or an **E2 radiation field** (for short).

To get the fields from this expression by taking its curl, and then the curl of its curl, is – ahem – most unpleasant. *Jackson* wimps out! Actually, taking the curls is no more difficult than it was for the magnetic term, but untangling the integrals with the result is, because of the tensor forms that appear. Consequently we too will wimp out (in the comforting knowledge that we will shortly do this *right* and *not* wimp out to arbitrary order in a precise decomposition) and will restrict our attention to the far zone.

There we need only consider the lowest order surviving term, which always comes from the curl of the exponential times the rest:

$$\boldsymbol{B} = ik(\hat{\boldsymbol{n}} \times \boldsymbol{A}) \tag{11.150}$$

$$\boldsymbol{E} = ik\sqrt{\frac{\mu_0}{\epsilon_0}}(\hat{\boldsymbol{n}} \times \boldsymbol{A}) \times \hat{\boldsymbol{n}}. \tag{11.151}$$

If we keep only the lowest order terms of *this* we get

$$\boldsymbol{B} = -\frac{ick^2\mu_0}{8\pi}\frac{e^{ikr}}{r}\int(\hat{\boldsymbol{n}} \times \boldsymbol{x}')(\hat{\boldsymbol{n}} \cdot \boldsymbol{x}')\rho(\boldsymbol{x}')d^3x'. \tag{11.152}$$

If we recall (from the beginning of Chapter 4) the discussion and definition of multipole moments, in particular the **quadrupole moment tensor**

$$Q_{\alpha\beta} = \int(3x'_\alpha x'_\beta - r'^2\delta_{\alpha\beta})\rho(\boldsymbol{x}')d^3x' \tag{11.153}$$

whose various components can be related to the five spherical harmonics with $\ell = 2$ (!) we can simplify matters. We can write the one messy integral in terms of another:

$$\hat{\boldsymbol{n}} \times \int \boldsymbol{x}'(\hat{\boldsymbol{n}} \cdot \boldsymbol{x}')\rho(\boldsymbol{x}')d^3x' = \frac{1}{3}\hat{\boldsymbol{n}} \times \boldsymbol{Q}(\hat{\boldsymbol{n}}) \tag{11.154}$$

where

$$\boldsymbol{Q}(\hat{\boldsymbol{n}}) = \sum_\beta Q_{\alpha\beta}n_\beta\hat{\boldsymbol{x}}_\beta. \tag{11.155}$$

Note that the "vector" $\boldsymbol{Q}(\hat{\boldsymbol{n}})$ (and hence the fields) depends in both the magnitude and direction on the direction to the point of observation $\mathbf{n}$ as well as the properties of the source. With these definitions,

$$\boldsymbol{B} = -\frac{ick^3\mu_0}{24\pi}\frac{e^{ikr}}{r}(\hat{\boldsymbol{n}} \times \boldsymbol{Q}(\hat{\boldsymbol{n}})) \tag{11.156}$$

which looks (except for the peculiar form of $\mathbf{Q}$) much like the E1 magnetic field. It is transverse. The electric field is obtained by appending $\times\mathbf{n}$ and is also transverse. Following exactly the same algebraic procedure as before, we find from

$$\boldsymbol{S} = \frac{1}{2}\mathrm{Re}\{\boldsymbol{E} \times \boldsymbol{H}^*\} \tag{11.157}$$

and computing the flux of the Poynting vector through a sphere of radius r as a function of angle that the angular power distribution is:

$$\frac{dP}{d\Omega} = \frac{c^2}{1152\pi^2}\sqrt{\frac{\mu_0}{\epsilon_0}}k^6 \mid (\hat{\boldsymbol{n}} \times \boldsymbol{Q}(\hat{\boldsymbol{n}})) \times \hat{\boldsymbol{n}} \mid^2 \tag{11.158}$$

The angular distribution is too complicated to play with further unless you need to calculate it, in which case you will have to work it out. The total power can be calculated in a "straightforward" way (to quote Jackson). First one changes the cross product to dot products using the second relation on the front cover and squares it. One then writes out the result in tensor components. One can then perform the angular integrals of the products of the components of the **n** (which is straightforward). Finally one term in the resulting expression goes away because $Q_{\alpha\beta}$ is traceless. The result is

$$P = \frac{c^2 k^6}{1440\pi}\sqrt{\frac{\mu_0}{\epsilon_0}}\sum_{\alpha,\beta} \mid Q_{\alpha\beta} \mid^2 \tag{11.159}$$

(note k^6 frequency dependence). For the numerologists among you, note that there is almost certainly some sort of cosmic significance in the 1440 in the denominator as this is the number of seconds in a day.

Just kidding.

For certain symmetric distributions of charge the general quadrupole moment tensor simplifies still further. A typical case of this occurs when there is an additional, e. g. azimuthal symmetry such as an oscillating spheroidal distribution of charge. In this case, the off–diagonal components of $Q_{\alpha\beta}$ vanish and only two of the remaining three are independent. We can write

$$Q_{33} = Q_0, \quad Q_{11} = Q_{22} = -\frac{1}{2}Q_0 \tag{11.160}$$

and the angular distribution of radiated power is

$$\frac{dp}{d\Omega} = \frac{c^2 k^6}{512\pi^2}\sqrt{\frac{\mu_0}{\epsilon_0}}Q_0^2 \sin^2\theta\cos^2\theta \tag{11.161}$$

which is a four–lobed radiation pattern characteristic of azimuthally symmetric sources. In this case it really is straightforward to integrate over the entire solid angle (or do the sum in the expression above) and show that:

$$P = \frac{c^2 k^6}{960\pi}\sqrt{\frac{\mu_0}{\epsilon_0}}Q_0^2. \tag{11.162}$$

At this point it should be clear that we are off on the wrong track. To quote Jackson:

> The labor involved in manipulating higher terms in (the multipolar expansion of $\mathbf{A}(\boldsymbol{x})$) becomes increasingly prohibitive as the expansion is extended beyond the electric quadrupole terms.

Some would say that we should have quit after the electric dipole or magnetic dipole.

The problem has several roots. First, in the second and all succeeding terms in the expansion as written, the magnetic and electric terms are all mixed up and of different tensorial character. This means that we have to project out the particular parts we want, which is not all that easy even in the simplest cases. Second, this approach is useful only when the wavelength is long relative to the source ($kd << 1$) which is not (always) physical for radio antennae. Third, what we have done is algebraically inefficient; we keep having to do the same algebra over and over again and it gets no easier.

Understanding the problem points out the way to solve it. We must start again at the level of the Green's function expansion, but this time we must construct a generalized *tensorial* multipolar expansion to use in the integral equation. After that, we must do "once and for all" the necessary curl and divergence algebra, and classify the resulting parts according to their formal transformation properties. Finally, we will reassemble the solution in the new **vector** multipoles and glory in its formal simplicity. Of course, the catch is that it is a lot of work at first. The payoff is that it is *general* and *systematically extendable* to all orders.

As we do this, I'm leaving you to work out the various example problems in Jackson (e.g. section J9.4, 9.5) on your own. We've already covered most of J9.6 but we have to do a bit more review of the angular part of the Laplace operator, which we largely skipped before. This will turn out to be key as we develop Multipolar Radiation Fields *properly*.

11.7 Radiation Assignment

1. Derive the integral expression for spherical bessel functions in terms of plane waves at the same wavenumber.

2. The addition theorems:

$$N_0(\mathbf{r}-\mathbf{r}') = n_0(k \mid \mathbf{r}-\mathbf{r}')\frac{1}{\sqrt{4\pi}} = \sqrt{4\pi}\sum_L N_L(\mathbf{r}_>)J_L(\mathbf{r}_<)^* \quad (11.163)$$

   and

$$H_0^\pm(\mathbf{r}-\mathbf{r}') = h_0^\pm(k \mid \mathbf{r}-\mathbf{r}')\frac{1}{\sqrt{4\pi}} = \sqrt{4\pi}\sum_L H_L^\pm(\mathbf{r}_>)J_L(\mathbf{r}_<)^*. \quad (11.164)$$

   are derived someplace, for both this special case and for the general case. Find at least one such place (for $L = 0, 0$), copy the derivation (with acknowledgement), and hand it in. If you work in a group, see how many places you can find it and compare. LEARN what you can from the process, that is, read the text accompanying the derivation(s) you find and try to understand it. Work it out. For extra credit, find in the literature the original paper that derives the general addition theorem. Hints: JMP, Danos and Maximon. Study it.

3. Derive the Green's function for the Helmholtz equation in free space (zero boundary conditions at infinity). Do not use the addition theorem, since you do not (in principle) know its form yet and so do not know that it is a Neumann or Hankel function. Naturally, you can follow Wyld or Jackson or Arfken, but acknowledge your source and show the entire derivation.

4. Make a neat sheet with Everything You Never Wanted To Know About Spherical Bessel/Neumann/Hankel Functions but were Afraid Not To Ask on it. Don't hand it in, this will be your guide through life (for at least a few weeks). Do NOT simply photocopy my notes. Do it by hand. Pool your sheets with those of your friends — put together everything to make a "best" sheet and then photocopy it. I use the term "sheet" loosely. I expect it will fill several (it did in my notes).

5. Using the addition theorem derived above (in the form of the Green's function) and the asymptotic relations on your worksheet, derive the static result for the vector potential **A** we previously obtained for the near field zone (my equation 66). Find the *lowest order correction* to this expression. This will, of course, involve finding more out about spherical waves than I have so far told you! item Using the same addition theorem and the other asymptotic relations, derive an expression

for the v.p. **A** in the *far zone*. Is there a correspondance of some sort with our previous result (Jackson 9.9)?

6. Show that

$$\mathbf{A}(\boldsymbol{x}) = ikh_1^+(kr)\sum_{m=-1}^{1} Y_{1,m}(\hat{r})\int \boldsymbol{J}(\mathbf{x}')j_1(kr')Y_{1,m}(\hat{r'})^* d^3x'$$

is equivalent to

$$\mathbf{A}(\boldsymbol{x}) = \frac{e^{ikr}}{4\pi r}\left(\frac{1}{r} - ik\right)\int \boldsymbol{J}(\mathbf{x}')(\mathbf{n}\cdot\mathbf{x}')d^3x'$$

for $kd << 1$.

7. Any vector quantity can be decomposed in a symmetric and an antisymmetric piece. Prove that, in the case of the $\ell = 1$ term derived above, the current term can be decomposed into

$$\boldsymbol{J}(\mathbf{n}\cdot\mathbf{x}') = \frac{1}{2}[(\mathbf{n}\cdot\mathbf{x}')\boldsymbol{J} + (\mathbf{n}\cdot\boldsymbol{J})\mathbf{x}'] + \frac{1}{2}(\mathbf{x}'\times\boldsymbol{J})\times\mathbf{n}$$

8. Evaluate the antisymmetric piece. Show (from the magnetic dipole vector potential) that

$$\boldsymbol{B} = \frac{\mu_0}{4\pi}\left\{k^2(\mathbf{n}\times\mathbf{m})\times\mathbf{n}\frac{e^{ikr}}{r} + [3\mathbf{n}(\mathbf{n}\cdot\mathbf{m}) - \mathbf{m}]\left(\frac{1}{r^3} - \frac{ik}{r^2}\right)e^{ikr}\right\}$$

and

$$\boldsymbol{E} = -\frac{1}{4\pi}\sqrt{\frac{\mu_0}{\epsilon_0}}k^2(\mathbf{n}\times\mathbf{m})\frac{e^{ikr}}{r}\left(1 - \frac{1}{ikr}\right).$$

Remark upon the similarities and differences between this result and the electric dipole result.

9. Next start to evaluate the integral of the symmetric piece. Show that you get:

$$\frac{1}{2}\int[(\mathbf{n}\cdot\mathbf{x}')\boldsymbol{J} + (\mathbf{n}\cdot\mathbf{J})\mathbf{x}']d^3x' = -\frac{i\omega}{2}\int \mathbf{x}'(\mathbf{n}\cdot\mathbf{x}')\rho(\mathbf{x}')d^3x'$$

The steps involved are:

(a) integrate by parts (working to obtain divergences of $\boldsymbol{J}$).

(b) changing $\boldsymbol{\nabla} \cdot \boldsymbol{J}$ into a ρ times whatever from the continuity equation (for a harmonic source).

(c) rearranging and recombining.

Don't forget the boundary condition at infinity!

10. Homemade tables, part II. What you did for spherical bessel functions, do for spherical harmonics. In particular, derive the commutation rules for the raising and lowering operators from the cartesian commutation relations for $\mathbf{L}$. From the commutation rules and $L_z Y_{\ell m} = m Y_{\ell m}$ derive the (normalized) action of $L_\pm$ on $Y_{\ell,m}$.

11. Jackson, problems 9.2, 9.3, 9.4

Chapter 12

Vector Multipoles

As I noted just above, we're already half way through J9.6, which is mostly the review of spherical bessel, neumann, and hankel functions that we have just had. The remainder is a lightning review of scalar spherical harmonics. Since we're about to *generalize* that concept, we'll quickly go over the high parts.

12.1 Angular momentum and spherical harmonics

The angular part of the Laplace operator ∇^2 can be written:

$$\frac{1}{r^2}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right\}=-\frac{L^2}{r^2} \tag{12.1}$$

Eliminating $-r^2$ (to solve for the L^2 differential equation) one needs to solve an eigenvalue problem:

$$L^2\psi = e\psi \tag{12.2}$$

where e are the eigenvalues, subject to the condition that the solution be single valued on $\phi \in [0, 2\pi)$ and $\theta \in [0, \pi]$.

This equation easily separates in θ, ϕ. The ϕ equation is trivial – solutions periodic in ϕ are indexed with integer m. The θ equation one has to work at a bit – there are constraints on the solutions that can be obtained for any given m – but there are many ways to solve it and at this point you should know that its solutions are *associated Legendre polynomials* $P_{\ell,m}(x)$ where $x = \cos\theta$. Thus the eigensolution becomes:

$$L^2 Y_{\ell m} = \ell(\ell+1) Y_{\ell m} \tag{12.3}$$

where $\ell = 0, 1, 2...$ and $m = -\ell, -\ell+1, ..., \ell-1, \ell$ and is typically orthonormal(ized) on the solid angle 4π.

The angular part of the Laplacian is related to the *angular momentum* of a wave in quantum theory. In units where $\hbar = 1$, the angular momentum operator is:

$$\boldsymbol{L} = \frac{1}{i}(\boldsymbol{x} \times \boldsymbol{\nabla}) \tag{12.4}$$

and

$$L^2 = L_x^2 + L_y^2 + L_z^2 \tag{12.5}$$

Note that in all of these expressions $\mathbf{L}$, L^2, L_z, etc. are all *operators*. This means that they are *applied* to the functions on their right (by convention). When you see them appearing by themselves, remember that they only mean something when they are applied, so $\boldsymbol{\nabla}$'s out by themselves on the right are ok.

The z component of L is:

$$L_z = -i\frac{\partial}{\partial\phi} \tag{12.6}$$

and we see that in fact Ylm satisfies the *two* eigenvalue equations:

$$L^2 Y_{\ell m} = \ell(\ell+1) Y_{\ell m} \tag{12.7}$$

and

$$L_z Y_{\ell m} = m Y_{\ell m} \tag{12.8}$$

The Ylm's cannot be eigensolutions of more than one of the components of $\boldsymbol{L}$ at once. However, we can write the cartesian components of $\mathbf{L}$ so that they form an *first rank tensor algebra* of operators that transform the $Y_{\ell m}$, for a given ℓ, among themselves (they cannot change ℓ, only mix m). This is the hopefully familiar set of equations:

$$\begin{aligned} L_+ &= L_x + iL_y & (12.9) \\ L_- &= L_x - iL_y & (12.10) \\ L_0 &= L_z & (12.11) \end{aligned}$$

The Cartesian components of $\boldsymbol{L}$ do not commute. In fact, they form a nice antisymmetric set:

$$[L_i, L_j] = i\epsilon_{ijk}L_k \tag{12.12}$$

which can be written in the shorthand notation

$$\mathbf{L} \times \mathbf{L} = i\mathbf{L}. \tag{12.13}$$

Consequently, the components expressed as a first rank tensor also do not commute among themselves:

$$[L_+, L_-] = 2L_z \tag{12.14}$$

and

$$[L_\pm, L_z] = \mp L_\pm \tag{12.15}$$

but *all* these ways of arranging the components of $\boldsymbol{L}$ commute with L^2:

$$[L_i, L^2] = 0 \tag{12.16}$$

and therefore with the Laplacian itself:

$$[\nabla^2, L_i] = 0 \tag{12.17}$$

which can be written in terms of L^2 as:

$$\nabla^2 = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\ ) - \frac{L^2}{r^2} \tag{12.18}$$

As one can easily show either by considering the explict action of the actual differential forms on the actual eigensolutions $Y_{\ell m}$ or more subtly by considering the action of L_z on $L_\pm Y_{\ell\ell}$ (and showing that they behave like raising and lower operators for m and preserving normalization) one obtains:

$$\begin{aligned} L_+Y_{\ell m} &= \sqrt{(\ell - m)(\ell + m + 1)}\, Y_{\ell,m+1} & (12.19) \\ L_-Y_{\ell m} &= \sqrt{(\ell + m)(\ell - m + 1)}\, Y_{\ell,m-1} & (12.20) \\ L_zY_{\ell m} &= mY_{\ell m} & (12.21) \end{aligned}$$

Finally, note that $\mathbf{L}$ is always orthogonal to $\mathbf{r}$ where both are considered as *operators* and $\mathbf{r}$ acts from the left:

$$\mathbf{r} \cdot \mathbf{L} = 0. \tag{12.22}$$

You will see many cases where identities such as this have to be written down in a particular order.

Before we go on to do a more leisurely tour of *vector* spherical harmonics, we pause to motivate the construction.

12.2 Magnetic and Electric Multipoles Revisited

As we have now seen repeatedly from Chapter J6 on, in a source free region of space, harmonic electromagnetic fields are divergenceless and have curls given by:

$$\begin{aligned} \boldsymbol{\nabla} \times \boldsymbol{E} &= i\omega \boldsymbol{B} = ikc\boldsymbol{B} && (12.23) \\ \boldsymbol{\nabla} \times \boldsymbol{B} &= -i\frac{k}{c}\boldsymbol{E}. && (12.24) \end{aligned}$$

By massaging these a little bit (recall $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{X}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{X}) - \nabla^2 \boldsymbol{X}$ and $\boldsymbol{\nabla} \cdot \boldsymbol{X} = 0$ for $\boldsymbol{X} = \boldsymbol{E}, \boldsymbol{B}$) we can easily show that both $\boldsymbol{E}$ and $\boldsymbol{B}$ must be divergenceless solutions to the HHE:

$$(\nabla^2 + k^2)\boldsymbol{X} = 0 \tag{12.25}$$

If we know a solution to this equation for $\boldsymbol{X} = \boldsymbol{E}$ we can obtain $\boldsymbol{B}$ from its curl from the equation above:

$$\boldsymbol{B} = -\frac{i}{\omega}\boldsymbol{\nabla} \times \boldsymbol{E} \tag{12.26}$$

and vice versa. However, this is annoying to treat directly, because of the vector charactor of $\boldsymbol{E}$ and $\boldsymbol{B}$ which complicate the description (as we have seen – transverse electric fields are related to magnetic multipoles and vice versa). Let's eliminate it.

By considering the action of the Laplacian on the scalar product of $\boldsymbol{r}$ with a well–behaved vector field $\boldsymbol{X}$,

$$\nabla^2(\boldsymbol{r} \cdot \boldsymbol{X}) = \boldsymbol{r} \cdot (\nabla^2 \boldsymbol{X}) + 2\boldsymbol{\nabla} \cdot \boldsymbol{X} \tag{12.27}$$

and using the divergenceless of $\boldsymbol{E}$ and $\boldsymbol{B}$, we see that the scalars $(\boldsymbol{r}\cdot\boldsymbol{E})$ and $(\boldsymbol{r}\cdot\boldsymbol{B})$ *also* satisfy the HHE:

$$(\nabla^2 + k^2)(\boldsymbol{r}\cdot\boldsymbol{E}) = 0 \tag{12.28}$$

$$(\nabla^2 + k^2)(\boldsymbol{r}\cdot\boldsymbol{B}) = 0 \tag{12.29}$$

We already know how to write a general solution to either of these equations in terms of the spherical bessel, neumann, and hankel functions times spherical harmonics.

Recall, that when we played around with multipole fields, I kept emphasizing that electric n-pole fields were transverse magnetic and vice versa? Well, transverse electric fields have $(\boldsymbol{r}\cdot\boldsymbol{E}) = 0$ by definition, right? So now we *define* a **magnetic multipole field of order L** by

$$\boldsymbol{r}\cdot\boldsymbol{B}_L^{(M)} = \frac{\ell(\ell+1)}{k} g_\ell(kr) Y_L(\hat{r}) \tag{12.30}$$

$$\boldsymbol{r}\cdot\boldsymbol{E}_L^{(M)} \;=\; 0. \tag{12.31}$$

Similarly, a **electric multipole field of order L** (which must be transverse magnetic) is any solution such that

$$\boldsymbol{r}\cdot\boldsymbol{E}_L^{(E)} = -\frac{\ell(\ell+1)}{k} f_\ell(kr) Y_L(\hat{r}) \tag{12.32}$$

$$\boldsymbol{r}\cdot\boldsymbol{B}_L^{(E)} \;=\; 0. \tag{12.33}$$

In these two definitions, $g_\ell(kr)$ and $f_\ell(kr)$ are arbitrary linear combinations of spherical bessel functions[1], two at a time. Jackson uses the two hankel functions in (J9.113)k, but this is not necessary.

Now, a little trickery. Using the curl equation for $\boldsymbol{B}$ we get:

$$k(\boldsymbol{r}\cdot\boldsymbol{B}_L^{(M)}) = \frac{1}{i}\boldsymbol{r}\cdot(\boldsymbol{\nabla}\times\boldsymbol{E}_L^{(M)}) = \frac{1}{i}(\boldsymbol{r}\times\boldsymbol{\nabla})\cdot\boldsymbol{E}_L^{(M)} = \mathbf{L}\cdot\boldsymbol{E}_L^{(M)} \tag{12.34}$$

so that $\mathbf{L}\cdot\boldsymbol{E}_L^{(}M)$ is a scalar solution to the HHE for magnetic multipolar fields. Ditto for $\mathbf{L}\cdot\boldsymbol{B}_L^{(E)}$ in the case of electric multipolar fields. Thus,

$$\mathbf{L}\cdot\boldsymbol{E}_L^{(M)} = \ell(\ell+1) g_\ell(kr) Y_L(\hat{r}) \tag{12.35}$$

etc. for $\mathbf{L}\cdot\boldsymbol{B}_L^{(E)}$.

[1] From now on, this term is generic unless clearly otherwise in context.

Now we get *really* clever. Remember that $\boldsymbol{r} \cdot \mathbf{L} = 0$. Also, $L^2 = \mathbf{L} \cdot \mathbf{L}$. We have arranged things just so that if we write:

$$\boldsymbol{E}_L^{(M)} = g_\ell(kr)\mathbf{L}Y_L(\hat{r}) \tag{12.36}$$

$$\boldsymbol{B}_L^{(M)} = -\frac{i}{\omega}\boldsymbol{\nabla} \times \boldsymbol{E}_L^{(M)} \tag{12.37}$$

we exactly reconstruct the solutions above. Neato! This gives us a completely general TE, MM EMF. A TM, EM EMF follows similarly with $g \to f$ and $\boldsymbol{E} \leftrightarrow \boldsymbol{B}$ (and a minus sign in the second equation).

This is good news and bad news. The good news is that this is a hell of a lot simpler than screwing around with symmetric and antisymmetric vector decompositions and integrations by parts *ad nauseam*. The radial part of the solutions is straightforward, and the angular part is written in a concise notation. The bad news is we have never *seen* that notation, good or bad, ever before. We have two choices. Either we can laboriously crank out the operator products and curls for each problem as we need to (which is really just as bad as what we have been doing) or we have to work out the algebra of these new objects once and for all so we can plug and chug out the most difficult of answers with comparative ease.

Guess which one we're about to do.

12.3 Vector Spherical Harmonics and Multipoles

Recall that

$$\mathbf{L} = -i\boldsymbol{r} \times \boldsymbol{\nabla}. \tag{12.38}$$

This is an "orbital" rotation operator. In systems with spin it is more convenient in many cases to define a "total" rotation operator that adds the orbital rotation operator to a "spin" rotation operator (defined below). Since total angular momentum (as opposed to orbital angular momentum) is a relativistically invariant quantity that appears "naturally" in covariant kinematics, we are inspired to find a representation that is

1. A vector function of its coordinates.

2. Simultaneous eigenfunctions of J^2, L^2, and J_z.

3. Possessed of certain desirable properties we will derive.

Actually, figuring out something like this the first time is not quite so easy; it is full of false starts and exploring alternatives. After the fact, however, it is clear that this is the correct choice. It is also extremely useful in quantum theory.

The total rotation operator is

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \tag{12.39}$$

where

$$\mathbf{S} = i\mathbf{I}\times \tag{12.40}$$

is the "spin" operator.

Aside: The Spin Operator

$\mathbf{S}$ in this expression is a *tensor operator*. It (like all operators) has no meaning by itself. It is, however, quite different from the scalar operators you are used to. Among other things, when $\mathbf{S}$ operates on a vector $\mathbf{A}$, it generates a *new* vector that points in a *different direction*. Let us see this.

In the definition of $\mathbf{S}$, $\mathbf{I}$ is the identity tensor (unit diagonal matrix) and it is crossed into whatever sits on its right. To understand its action, let us evaluate its cartesian components acting on some vector $\mathbf{A}$:

$$S_x\mathbf{A} = i\mathbf{I}_x \times \mathbf{A} = i\hat{x}\times\mathbf{A} \tag{12.41}$$
$$S_y\mathbf{A} = i\hat{y}\times\mathbf{A} \tag{12.42}$$
$$S_x\mathbf{A} = i\hat{z}\times\mathbf{A} \tag{12.43}$$

or (e. g.)

$$S_z\mathbf{A} = i(A_x\hat{y} - A_y\hat{x}). \tag{12.44}$$

Note that the action of a component of $\mathbf{S}$ on a vector $\mathbf{A}$ shifts the direction of $\mathbf{A}$ to a direction perpendicular to both $\mathbf{S}$ and the component. Only by considering the action of all the components can the total vector action of $\mathbf{S}$ on $\mathbf{A}$ in a given direction be evaluated.

There are several important properties of $\mathbf{S}$. The first is to note that it has the form of an angular momentum operator with a special action on vectors. If we form S^2 and evaluate its action on $\mathbf{A}$:

$$\begin{aligned}
S^2\mathbf{A} &= -\Big\{\hat{x}\times(\hat{x}\times\mathbf{A})+\hat{y}\times(\hat{y}\times\mathbf{A})+\hat{z}\times(\hat{z}\times\mathbf{A})\Big\} \\
&= -\{\mathbf{A}-3\mathbf{A}\} \\
&= 2\mathbf{A} = s(s+1)\mathbf{A} \qquad (12.45)
\end{aligned}$$

for $s = 1$. S^2 acting on *any vector* produces 2 times the same vector, telling us that a vector has "spin angular momentum" of 1. Note that this connection is universal. In field theory a "vector boson" has spin 1. In electrodynamics (quantum or classical) the "vector radiation field" has spin one.

The spin operator thus formed is more general, because its action can be extended to higher rank tensors. (2nd rank tensor) gravitational fields have spin 2. Scalar (0th rank tensor) fields have spin 0. To treat more general cases, however, we have to work with tensor indices explicitly and you'll see enough of that in the section on relativity. Feel free to study this matter further. Louck and Biedenharn's book (Encycl. of Math Phys., see me for ref.) contains a much deeper discussion of this entire subject.

It may seem that with such a peculiar structure, S_z can have no eigenvectors. This is not the case. You should verify that

$$\chi_1^1 = -\frac{1}{\sqrt{2}}(\hat{x}+i\hat{y}) \qquad (12.46)$$

$$\chi_1^0 = \hat{z} \qquad (12.47)$$

$$\chi_1^{-1} = \frac{1}{\sqrt{2}}(\hat{x}-i\hat{y}) \qquad (12.48)$$

are eigenvectors such that

$$S_z\chi_1^{m_s} = m_s\chi_1^{m_s} \qquad (12.49)$$

for $m_s = -1, 0, 1$ and

$$S^2\chi_1^{m_s} = s(s+1)\chi_1^{m_s} \qquad (12.50)$$

for $s = 1$. You should also verify the commutation relations for the components of $\mathbf{S}$, that is, show that

$$\mathbf{S}\times\mathbf{S} = i\mathbf{S} \qquad (12.51)$$

making it a "true" rotation/angular momentum operator.

In addition, we will need to use the operators

$$J^2 = J_x J_x + J_y J_y + J_z J_z, \tag{12.52}$$

$$J_z = L_z + S_z \tag{12.53}$$

(etc.) and

$$L^2 = L_x L_x + L_y L_y + L_z L_z \tag{12.54}$$

so that

$$J^2 = L^2 + 2 + 2i\mathbf{L}\times \tag{12.55}$$

which can be proven as follows.

Consider its action on $\mathbf{A}$ (as usual):

$$\begin{aligned} J^2\mathbf{A} &= \{L^2 + S^2 + 2\mathbf{L}\cdot\mathbf{S}\}\mathbf{A} \\ &= \{L^2 + 2 + 2i[L_x(\hat{x}\times\quad) + L_y(\hat{y}\times\quad) + L_z(\hat{z}\times\quad)]\}\mathbf{A} \\ &= \{L^2 + S^2 + 2i(\mathbf{L}\times\quad)\}\mathbf{A} \end{aligned} \tag{12.56}$$

where the meaning of the latter expression is hopefully now clear.

Then we define the **vector spherical harmonics $\mathbf{Y}^m_{j,\ell}$** by:

$$J^2\boldsymbol{Y}^m_{j,\ell} = j(j+1)\boldsymbol{Y}^m_{j,\ell} \tag{12.57}$$

$$L^2\boldsymbol{Y}^m_{j,\ell} = \ell(\ell+1)\boldsymbol{Y}^m_{j,\ell} \tag{12.58}$$

$$J_z\boldsymbol{Y}^m_{j,\ell} = m\boldsymbol{Y}^m_{j,\ell}. \tag{12.59}$$

Note that in order for the latter expression to be true, we might reasonably expect the vector spherical harmonics to be constructed out of sums of products of spherical harmonics and the eigenvectors of the operator S_z defined above. This is the vector analogue of constructing a spinor wavefunction in quantum theory.

In addition, we normalize these orthogonal functions so that they are *orthonormal* as a *dot product*. This will allow us to use them to construct projections.

$$\int \boldsymbol{Y}^{m*}_{j,\ell}(\theta,\phi)\cdot\boldsymbol{Y}^{m'}_{j',\ell'}(\theta,\phi)d\Omega = \delta_{jj'}\delta_{\ell\ell'}\delta mm' \tag{12.60}$$

We now need to derive the properties of these functions. We begin by applying J^2 to $\boldsymbol{Y}^m_{j,\ell}$

$$J^2\boldsymbol{Y}^m_{j,\ell} = \left\{L^2 + 2 + 2i\mathbf{L}\times\right\}\boldsymbol{Y}^m_{j,\ell} \tag{12.61}$$

so that we get

$$2i\mathbf{L} \times \boldsymbol{Y}^m_{j,\ell} = \{j(j+1) - \ell(\ell+1) - 2\}\, \boldsymbol{Y}^m_{j,\ell}. \tag{12.62}$$

Most of the later results will be based on this one, so understand it completely.

If we take $\mathbf{L}\cdot$ of both sides of (12.62), use a vector identity and recall that $\mathbf{L} \times \mathbf{L} = i\mathbf{L}$ we get:

$$[j(j+1) - \ell(\ell+1)]\mathbf{L} \cdot \boldsymbol{Y}^m_{j,\ell} = 0. \tag{12.63}$$

Similarly, we form the vector product of $\mathbf{L}$ with both sides of (12.62):

$$\{j(j+1) - \ell(\ell+1) - 2\}\, \mathbf{L} \times \boldsymbol{Y}^m_{j,\ell} = 2i\mathbf{L} \times (\mathbf{L} \times \boldsymbol{Y}^m_{j,\ell}). \tag{12.64}$$

To reduce this further, we must use the operator vector identity (which you should prove)

$$\mathbf{L} \times (\mathbf{L} \times \mathbf{V}) = \mathbf{L}(\mathbf{L} \cdot \mathbf{V}) + i\mathbf{L} \times \mathbf{V} - L^2\mathbf{V} \tag{12.65}$$

and eliminate the $\mathbf{L} \times \mathbf{Y}$ using (12.62). One gets:

$$\begin{aligned} [j(j+1) - \ell(\ell+1)]\, [j(j+1) - \ell(\ell+1) - 2]\, \boldsymbol{Y}^m_{j,\ell} &= \\ & 4\ell(\ell+1)\boldsymbol{Y}^m_{j,\ell} - 4\mathbf{L}(\mathbf{L} \cdot \boldsymbol{Y}^m_{j,\ell}) \end{aligned} \tag{12.66}$$

If we eliminate the $\mathbf{L}\cdot\mathbf{Y}$ (using the result above) we get the *characteristic equation* that is a constraint on the possible values of j and ℓ:

$$x^3 - 2x^2 - 4\ell(\ell+1)x = 0 \tag{12.67}$$

where

$$x = j(j+1) - \ell(\ell+1) \tag{12.68}$$

by definition. The solutions to this factorizable cubic are:

$$j = \ell, \ell+1, \ell-1, -\ell-1, -\ell-2, -\ell.$$

We only need to consider the solutions with positive j in this problem as the others are not independent in this case. Since $\ell \geq 0$ we only need consider the first three possibilities.

Solutions with $j = \ell$

Then $x = 0$ and

$$j(j+1)\boldsymbol{Y}_{jj}^{m} = \mathbf{L}(\mathbf{L}\cdot\boldsymbol{Y}_{jj}^{m}) \tag{12.69}$$

from the third equation above. If we take the dot product of $\mathbf{L}$ with this relation, we get

$$L^2(\mathbf{L}\cdot\boldsymbol{Y}_{jj}^{m}) = j(j+1)(\mathbf{L}\cdot\boldsymbol{Y}_{jj}^{m}) \tag{12.70}$$

and we thus see that $\mathbf{L}\cdot\boldsymbol{Y}_{jj}^{m} \propto Y_{j,m}$ and so:

$$\boldsymbol{Y}_{jj}^{m} = \frac{1}{\sqrt{j(j+1)}}\mathbf{L}Y_{j,m} \tag{12.71}$$

(!) where we have normalized the result.

We have at last found something recognizable. This is precisely the combination of spherical harmonics and $\mathbf{L}$ we found in our brief excursion into multipoles! We see that we could have written the (e. g.) magnetic solution as

$$\boldsymbol{E}_{L}^{(M)} = g_\ell(kr)\sqrt{\ell(\ell+1)}\boldsymbol{Y}_{\ell\ell}^{m} \tag{12.72}$$

$$\boldsymbol{B}_{L}^{(M)} = -\frac{i}{\omega}\boldsymbol{\nabla}\times\boldsymbol{E}_{L}^{(M)}. \tag{12.73}$$

With just a little more work (later) we will be able to obtain the curl part as a general result, which will *really* simplify life for us. It is a trivial exercise (left for the reader) to verify that

$$J_z\boldsymbol{Y}_{jj}^{m} = m\boldsymbol{Y}_{jj}^{m}. \tag{12.74}$$

One simply plugs in the explicit form of J_z and commutes the resultant L_z with $\mathbf{L}$ to cancel the "spin" part.

Solutions with $j \neq \ell$

If $j \neq \ell$, we see from the equation after (12.62) that $\mathbf{L}\cdot\mathbf{Y} = 0$. To go further we have to go back to (12.62) and follow a different line. If we multiply both sides by $\hat{r}\cdot$ and $\hat{r}\times$,

$$[j(j+1) - \ell(\ell+1) - 2]\,\hat{r}\cdot\boldsymbol{Y}_{j\ell}^{m} = 2i\hat{r}\cdot\mathbf{L}\times\boldsymbol{Y}_{j\ell}^{m} \tag{12.75}$$

and

$$[j(j+1) - \ell(\ell+1) - 2]\,\hat{r}\times\boldsymbol{Y}_{j\ell}^{m} = 2i\hat{r}\times(\mathbf{L}\times\boldsymbol{Y}_{j\ell}^{m}) \tag{12.76}$$

We can reduce these with the vector identities

$$\hat{r} \cdot (\mathbf{L} \times \mathbf{A}) = 2i\hat{r} \cdot \mathbf{A} - \mathbf{L} \cdot (\hat{r} \times \mathbf{A}) \tag{12.77}$$

and

$$\hat{r} \times (\mathbf{L} \times \mathbf{A}) = \mathbf{L}(\hat{r} \cdot \mathbf{A}) + i\hat{r} \times \mathbf{A}. \tag{12.78}$$

You should get

$$[j(j+1) - \ell(\ell+1) + 2]\, \hat{r} \cdot \boldsymbol{Y}_{j\ell}^{m} = -2i\mathbf{L} \cdot (\hat{r} \times \boldsymbol{Y}_{j\ell}^{m}) \tag{12.79}$$

and

$$[j(j+1) - \ell(\ell+1)]\, \hat{r} \times \boldsymbol{Y}_{j\ell}^{m} = 2i\mathbf{L}(\hat{r} \cdot \boldsymbol{Y}_{j\ell}^{m}). \tag{12.80}$$

Finally, if we plug the second of these into the first and eliminate the cross product, we get the scalar equation:

$$\frac{1}{4}\, [j(j+1) - \ell(\ell+1)]\, [j(j+1) - \ell(\ell+1) + 2]\, (\hat{r} \cdot \boldsymbol{Y}_{j\ell}^{m}) = L^2(\hat{r} \cdot \boldsymbol{Y}_{j\ell}^{m}). \tag{12.81}$$

This implies that $(\hat{r} \cdot \boldsymbol{Y}_{j\ell}^{m})$ is a spherical harmonic: that is a constant $\times Y_{k,m}$. What? This is not obvious to you? Well, just this once:

$$\left[\frac{j(j+1) - \ell(\ell+1)}{2}\right] \left[\frac{j(j+1) - \ell(\ell+1)}{2} + 1\right] = k(k+1) \tag{12.82}$$

This has the solutions

1. $k = \left[\frac{j(j+1)-\ell(\ell+1)}{2}\right]$
2. $k = \left[\frac{j(j+1)-\ell(\ell+1)}{2}\right] - 1.$

Since we already know that $j = \ell \pm 1$, we can investigate these two cases explicitly. The positive solutions (in both cases) are easily seen to be $k = j$. We can then construct the complete solutions, since

$$\boldsymbol{Y}_{j,\ell}^{m} = \hat{r}(\hat{r} \cdot \boldsymbol{Y}_{j,\ell}^{m}) - \hat{r} \times (\hat{r} \times \boldsymbol{Y}_{j,\ell}^{m}) \tag{12.83}$$

is an identity (related to the symmetric/antisymmetric decomposition and hence worth proving) and since we have already shown that

$$\hat{r} \times \boldsymbol{Y}_{j,\ell}^{m} = 2i\, [j(j+1) - \ell(\ell+1)]^{-1}\, \mathbf{L}(\hat{r} \cdot \boldsymbol{Y}_{j,\ell}^{m}) \tag{12.84}$$

with $(\hat{r} \cdot \boldsymbol{Y}^m_{j,\ell})$ a constant times $Y_{\ell,m}$. We get:

$$\boldsymbol{Y}^m_{j,\ell} = (constant) \left\{ \hat{r} - 2i[j(j+1) - \ell(\ell+1)]^{-1}(\hat{r} \times \mathbf{L}) \right\} Y_{\ell,m}. \qquad (12.85)$$

An exercise will be to verify the normalization of the final solutions:

$$\begin{aligned} \boldsymbol{Y}^m_{j,j-1} &= -\frac{1}{\sqrt{j(2j+1)}}[-j\hat{r} + i\hat{r} \times \mathbf{L}]Y_{\ell,m} & (12.86) \\ \boldsymbol{Y}^m_{j,j+1} &= -\frac{1}{\sqrt{(j+1)(2j+1)}}[(j+1)\hat{r} + i\hat{r} \times \mathbf{L}]Y_{\ell,m}. & (12.87) \end{aligned}$$

You must also verify that they satisfy the equation for J_z.

Finally, you are probably wondering why we have bothered to learn all of this about the $j \neq \ell$ cases in the first place. It is because

$$\begin{aligned} i\boldsymbol{\nabla} \times (\boldsymbol{Y}^m_{jj} f(r)) &= \sqrt{\frac{j+1}{2j+1}}\Big[(j+1)\frac{f}{r} + \frac{df}{dr}\Big]\mathbf{Y}^m_{j,j-1} \\ &+ \sqrt{\frac{j}{2j+1}}\Big[-j\frac{f}{r} + \frac{df}{dr}\Big]\mathbf{Y}^m_{j,j+1}. \qquad (12.88) \end{aligned}$$

The action of the curl *mixes* the vector spherical harmonics. In fact, it acts to shift j by one in any permitted direction (see handout sheet). Therefore, in order to evaluate the entire EM field and express it compactly, one must use the notation of the vector spherical harmonics. You should prove this, and at leat one of the divergence equations for homework. You will need to get the components of the v.s.h. along and transverse to $\hat{r}$ in order to do the vector algebra.

This is not too bad, but (as we shall see) it is not the best we can do. By carefully defining a particular set of multipolar solutions, we can make our notation itself do almost all the work of doing the curls, etc. so that all we have to do at either end is translate a paticular problem into and out of the notation with the formal solution in hand. Next time we will do just that as we develop the **Hansen Multipolar Solutions**.

Chapter 13

The Hansen Multipoles

We have already seen how if we let $\mathbf{E}$ or $\mathbf{B}$ be given by

$$\mathbf{E} \;\text{ or }\; \mathbf{B} = \frac{1}{\sqrt{\ell(\ell+1)}} f_\ell(kr)\mathbf{L}Y_L(\hat{r}) \tag{13.1}$$

then

1. Both the fields given above and their partner fields (given by the curl) have zero divergence.
2. The fields given above are completely transverse, since $\hat{r} \cdot \mathbf{L} = 0$ (operator).
3. The partner fields given by the curl are not purely transverse.
4. In order to be consistent, the fields above are also the curls of the partner fields. In fact, this follows from vector identities for divergenceless fields.

It is therefore sensible to define, once and for all, a set of multipoles that embody these properties. In addition, anticipating a need to treat longitudinal fields as well as transverse fields, we will define a third kind of multipoles with zero curl but non–zero divergence. These will necessarily be "connected" to sources (why?). We will call these "pre-computed" combinations of bessel functions, vector spherical harmonics, and their curls the *Hansen* Multipoles (following unpublished notes from L. C. Biedenharn as I have been unable to determine his original reference):

13.1 The Hansen Multipoles

13.1.1 The Basic Solutions

The Hansen solutions to the *vector* HHE (that can expand the free space solutions for the vector potential or vector fields) are as follows. $\boldsymbol{M}_L$ is the (normalized) elementary solution consisting of a bessel function times $\boldsymbol{L}Y_L = \boldsymbol{Y}_{ll}^m$. It is (by construction) purely transverse: $\hat{\boldsymbol{r}} \cdot \boldsymbol{M}_L = 0$. $\boldsymbol{N}_L$ is the solution constructed by the taking the curl of $\boldsymbol{M}_L$. $\boldsymbol{L}_L$ is the "longitudinal" solution constructed by taking the gradient of the scalar solution – it is left as an exercise to show that this still satisfies the HHE. The three of these pieces span the range of possible solutions and reconstruct an identity tensor that can be used to construct a vector harmonic Green's function expansion.

This is summarized, with correction for factors of k introduced by the derivatives, here:

$$\boldsymbol{M}_L = \frac{1}{\sqrt{\ell(\ell+1)}}\boldsymbol{L}\left(f_\ell(kr)Y_L(\hat{\boldsymbol{r}})\right) = \frac{1}{\sqrt{\ell(\ell+1)}} f_\ell(kr)\boldsymbol{Y}_{ll}^m(\hat{\boldsymbol{r}}) \quad (13.2)$$

$$\mathbf{N}_L = \frac{i}{k}\boldsymbol{\nabla} \times \mathbf{M}_L \quad (13.3)$$

$$\mathbf{L}_L = -\frac{i}{k}\boldsymbol{\nabla}(f_\ell(kr)Y_L(\hat{r})) \quad (13.4)$$

13.1.2 Their Significant Properties

The virtue of the Hansen solutions is that they "automatically" work to decompose field components into parts that are mutual curls (as required by Faraday/Ampere's laws for the fields) or divergences (as required by Gauss's laws for the fields):

$$\boldsymbol{\nabla} \cdot \boldsymbol{M}_L = 0 \quad (13.5)$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{N}_L = 0 \quad (13.6)$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{L}_L = ikf_\ell(kr)Y_L(\hat{\boldsymbol{r}}) \quad (13.7)$$

Hence $\boldsymbol{M}_L$ and $\boldsymbol{N}_L$ are divergenceless, while the divergence of $\boldsymbol{L}_L$ is a scalar solution to the HHE! $\boldsymbol{L}_L$ is related to the scalar field and the gauge invariance of the theory in an interesting way we will develop. Also:

$$\begin{aligned}
\boldsymbol{\nabla} \times \boldsymbol{M}_L &= -ik\boldsymbol{N}_L & (13.8)\\
\boldsymbol{\nabla} \times \boldsymbol{N}_L &= ik\boldsymbol{M}_L & (13.9)\\
\boldsymbol{\nabla} \times \boldsymbol{L}_L &= 0 & (13.10)
\end{aligned}$$

which shows how $\boldsymbol{M}_L$ and $\boldsymbol{N}_L$ are now ideally suited to form the components of electric and magnetic multipole fields mutually linked by Ampere's and Faraday's law.

13.1.3 Explicit Forms

The beauty of the definitions above is that they permit us to do algebra that *initially* skips the following fully expanded forms in terms of the vector spherical harmonics. However ultimately one has to do computations, of course – there are no free lunches. The following results come from actually working out the gradients, divergences, and curls in the definitions:

$$\begin{aligned}
\boldsymbol{M}_L &= f_\ell(kr)\boldsymbol{Y}_{\ell\ell}^m & (13.11)\\
\boldsymbol{N}_L &= \sqrt{\frac{\ell+1}{2\ell+1}}f_{\ell-1}(kr)\boldsymbol{Y}_{\ell,\ell-1}^m - \sqrt{\frac{\ell}{2\ell+1}}f_{\ell+1}(kr)\boldsymbol{Y}_{\ell,\ell+1}^m & (13.12)\\
\boldsymbol{L}_L &= \sqrt{\frac{\ell}{2\ell+1}}f_{\ell-1}(kr)\boldsymbol{Y}_{\ell,\ell-1}^m + \sqrt{\frac{\ell+1}{2\ell+1}}f_{\ell+1}(kr)\boldsymbol{Y}_{\ell,\ell+1}^m & (13.13)
\end{aligned}$$

or (in differential form)

$$\begin{aligned}
\boldsymbol{M}_L &= f_\ell(kr)\boldsymbol{Y}_{\ell\ell}^m & (13.14)\\
\boldsymbol{N}_L &= \frac{1}{kr}\left\{\frac{d}{d(kr)}(krf_\ell)(i\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m) - \hat{\boldsymbol{r}}\sqrt{\ell(\ell+1)}f_\ell Y_L\right\} & (13.15)\\
\boldsymbol{L}_L &= \sqrt{\ell(\ell+1)}\frac{1}{kr}(i\hat{\boldsymbol{r}} \times f_\ell\boldsymbol{Y}_{\ell\ell}^m) - \hat{\boldsymbol{r}}\left[\frac{d}{d(kr)}f_\ell\right]Y_L & (13.16)
\end{aligned}$$

As we will see, these relations allow us to construct the *completely general* solution to the EM field equations in a way that is intuitive, reasonable, and mathematically and numerically tractible. In other words, we're (mostly) done with the grunt work and can begin to reap the rewards.

What grunt work remains, you might ask? Well, there are a *slew* of identities and evaluations and relations developed from the definitions of the spherical harmonics themselves, the spherical bessel/neumann/hankel functions *them*selves, and the vector spherical harmonics and Hansen solutions that can be worked out and assembled in a table of sorts to simplify the actual process of doing algebra or computations using them.

Such a table is presented at the end of this chapter, and proving relations on that table constitute most of the homework related to the chapter, since once this work is done doing actual computations for specific charge/current densities is reduced to quadratures (another way of saying "expressible as a bunch of definite integrals" that can either be done analytically if they are relatively simple or numerically if not).

Those rewards are most readily apparent when we construct the vector Green's function for the vector IHE.

13.2 Green's Functions for the Vector Helmholtz Equation

The correct form for the Green's function for the vector Helmholtz equation is

$$\overleftrightarrow{\mathbf{G}}_{\pm}(\mathbf{r},\mathbf{r}') = \overleftrightarrow{\mathbf{I}}\, G_{\pm}(\mathbf{r},\mathbf{r}') \tag{13.17}$$

(where $G_{\pm}(\mathbf{r},\mathbf{r}')$ is a Green's function for the scalar IHE, that is:

$$G_{\pm}(\mathbf{r},\mathbf{r}') = -\frac{e^{\pm ikR}}{4\pi R} \tag{13.18}$$

for $R = |\,\mathbf{r}-\mathbf{r}'\,|$. The identity tensor transforms a vector on the right into the same vector, so this seems like a trivial definition. However, the point is that we can now expand the identity tensor times the scalar Green's function in **vector spherical harmonics** or **Hansen functions** *directly!*

We get:

$$\begin{aligned}
\overleftrightarrow{\mathbf{G}}_{\pm}(\mathbf{r},\mathbf{r}') &= \mp ik \sum_{j,\ell,m} h_\ell^{\pm}(kr_>) j_\ell(kr_<) \boldsymbol{Y}_{j\ell}^{m}(\hat{r}) \boldsymbol{Y}_{j\ell}^{m*}(\hat{r}') \\
&= \mp ik \sum_{L} \Big\{ \mathbf{M}_L^{+}(\mathbf{r}_>) \mathbf{M}_L^{0\,*}(\mathbf{r}_<) + \mathbf{N}_L^{+}(\mathbf{r}_>) \mathbf{N}_L^{0\,*}(\mathbf{r}_<) +
\end{aligned}$$

$$\mathbf{L}_L^+(\mathbf{r}_>)\mathbf{L}_L^{0\,*}(\mathbf{r}_<)\Big\} \tag{13.19}$$

In all cases the "*"s are to be considered sliding, able to apply to the $\boldsymbol{Y}_{jl}^m(\hat{\boldsymbol{r}})$ only of either term under an integral.

I do not intend to prove a key element of this assertion (that the products of the $\boldsymbol{Y}_{jl}^m(\hat{\boldsymbol{r}})$ involved reduce to Legendre polynomials in the angle between the arguments times the identity tensor) in class. Instead, I leave it as an exercise. To get you started, consider how similar completeness/addition theorems are proven for the spherical harmonics themselves from the given orthonormality relation.

With these relations in hand, we end our mathematical digression into vector spherical harmonics and the Hansen solutions and return to the land of multipolar radiation.

13.3 Multipolar Radiation, revisited

We will now, at long last, study the *complete* radiation field including the scalar, longitudinal, and transverse parts. Recall that we wish to solve the two equations (in the Lorentz gauge):

$$\{\nabla^2 + k^2\}\Phi(\boldsymbol{x}) = -\frac{\rho}{\epsilon_0}(\mathbf{r}) \tag{13.20}$$

$$\{\nabla^2 + k^2\}\boldsymbol{A}(\boldsymbol{x}) = -\mu_0 \boldsymbol{J}(\boldsymbol{x}) \tag{13.21}$$

with the Lorentz condition:

$$\boldsymbol{\nabla}\cdot\boldsymbol{A} + \frac{1}{c^2}\frac{\partial\Phi}{\partial t} = 0 \tag{13.22}$$

which is connected (as we shall see) to the continuity equation for charge and current.

E and **B** are now (as usual) determined from the vector potential by the full relations, i. e. – we make no assumption that we are outside the region of sources:

$$\boldsymbol{E} = -\boldsymbol{\nabla}\Phi - \frac{\partial \boldsymbol{A}}{\partial t} \tag{13.23}$$

$$\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}, \tag{13.24}$$

Using the methods discussed before (writing the solution as an integral equation, breaking the integral up into the interior and exterior of the sphere of radius r, and using the correct order of the multipolar expansion of the Green's function in the interior and exterior regions) we can easily show that the general solution to the IHE's above is:

$$\Phi(\boldsymbol{r}) = ik \sum_L \left\{ p_L^{\rm ext}(r) J_L(\boldsymbol{r}) + p_L^{\rm int}(r) H_L^+(\boldsymbol{r}) \right\} \tag{13.25}$$

where

$$\begin{aligned} p_L^{\rm ext}(r) &= \int_r^\infty h_\ell^+(kr') Y_L^*(\hat{\boldsymbol{r}}') \rho(\boldsymbol{r}') d^3r' & (13.26) \\ p_L^{\rm int}(r) &= \int_0^r j_\ell(kr') Y_L^*(\hat{\boldsymbol{r}}') \rho(\boldsymbol{r}') d^3r' & (13.27) \end{aligned}$$

Outside the (bounding sphere of the) source, the exterior coefficient is zero and the interior coefficient is the scalar multipole moment $p_L = p_L^{\rm int}(\infty)$ of the charge source distribution, so that:

$$\Phi(\boldsymbol{r}) = \frac{ik}{\epsilon_0} \sum_L p_L H_L^+(\boldsymbol{r}) \tag{13.28}$$

This is an important relation and will play an significant role in the implementation of the gauge condition below.

Similarly we can write the interior and exterior multipolar moments of the current in terms of integrals over the various Hansen functions to obtain a completely general expression for the vector potential $\boldsymbol{A}(\boldsymbol{r})$. To simplify matters, I am going to only write down the solution obtained outside the current density distribution, although the integration volume can easily be split into $r_<$ and $r_>$ pieces as above and an exact solution obtained on all space *including inside the charge distribution*. It is:

$$\boldsymbol{A}(\boldsymbol{r}) = ik\mu_0 \sum_L \left\{ m_L \boldsymbol{M}_L^+(\boldsymbol{r}) + n_L \boldsymbol{N}_L^+(\boldsymbol{r}) + l_L \boldsymbol{L}_L(\boldsymbol{r}) \right\} \tag{13.29}$$

where

$$\begin{aligned} m_L &= \int \boldsymbol{J}(\boldsymbol{r}') \cdot \boldsymbol{M}_L^0(\boldsymbol{r}')^* d^3r' & (13.30) \\ n_L &= \int \boldsymbol{J}(\boldsymbol{r}') \cdot \boldsymbol{N}_L^0(\boldsymbol{r}')^* d^3r' & (13.31) \\ l_L &= \int \boldsymbol{J}(\boldsymbol{r}') \cdot \boldsymbol{L}_L^0(\boldsymbol{r}')^* d^3r' & (13.32) \end{aligned}$$

Note well that the action of the dot product within the dyadic form for the Green's function (expanded in Hansen solutions) reduces the dyadic tensor to a vector again.

It turns out that these *four* sets of numbers: p_L, m_L, n_L, l_L are not independent. They are related by the requirement that the solutions satisfy the *Lorentz gauge condition*, which is a *constraint* on the admissible solutions. If we substitute these forms into the gauge condition itself and use the differential relations given above for the Hansen functions to simplify the results, we obtain:

$$\begin{aligned}
\boldsymbol{\nabla}\cdot\boldsymbol{A}+\frac{1}{c^2}\frac{\partial\Phi}{\partial t} &= 0\\
ik\sum_L\left\{\mu_0 l_L\boldsymbol{\nabla}\cdot\boldsymbol{L}_L^+-\frac{i\omega}{c^2\epsilon_0}p_L H_L^+\right\} &= 0\\
ik\sum_L\left\{l_L\boldsymbol{\nabla}\cdot\boldsymbol{L}_L^+-ikcp_L H_L^+\right\} &= 0\\
-k^2\sum_L\left\{l_L-cp_L\right\}H_L^+ &= 0
\end{aligned}\tag{13.33}$$

where we used $\boldsymbol{\nabla}\cdot\mathbf{L}_L^+=ikH_L^+$ in the last step. If we multiply from the left by $Y^*_{\ell',m'}$ and use the fact that the Y_L form a complete orthonormal set, we find the relation:

$$l_L-cp_L=0\tag{13.34}$$

or

$$l_L=cp_L\tag{13.35}$$

This tells us that the effect of the scalar moments and the longitudinal moments are connected by the gauge condition. Instead of four relevant moments we have at most three. In fact, as we will see below, we have only two!

Recall that the *potentials* are not unique – they can and do vary according to the gauge chosen. The *fields*, however, *must* be unique or we'd get different experimental results in different gauges. This would obviously be a problem!

Let us therefore calculate the fields. There are two ways to proceed. We can compute $\boldsymbol{B}$ directly from vA:

$$\begin{aligned}
\boldsymbol{B} &= \boldsymbol{\nabla}\times\boldsymbol{A}\\
&= ik\mu_0\sum_L\left\{m_l(\boldsymbol{\nabla}\times\boldsymbol{M}_L^+)+n_l(\boldsymbol{\nabla}\times\boldsymbol{N}_L^+)+l_l(\boldsymbol{\nabla}\times\boldsymbol{L}_L^+)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= ik\mu_0 \sum_L \left\{ m_l(-ik\boldsymbol{N}_L^+) + n_l(ik\boldsymbol{M}_L^+) \right\} \\
&= k^2\mu_0 \sum_L \left\{ m_L\boldsymbol{N}_L^+ - n_L\boldsymbol{M}_L^+ \right\} \qquad (13.36)
\end{aligned}
$$

and use Ampere's Law, $\boldsymbol{\nabla} \times \boldsymbol{B} = \mu_0\epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} = -i\omega\mu_0\epsilon_0 E$ to find $\boldsymbol{E}$:

$$
\begin{aligned}
\boldsymbol{E} &= \frac{ic^2}{kc}\boldsymbol{\nabla} \times \boldsymbol{B} \\
&= ikc\mu_0 \sum_L \left\{ m_L(\boldsymbol{\nabla} \times \boldsymbol{N}_L^+) - n_L(\boldsymbol{\nabla} \times \boldsymbol{M}_L^+) \right\} \\
&= ik\sqrt{\frac{1}{\mu_0\epsilon_0}}\mu_0 \sum_L \left\{ m_L(ik\boldsymbol{M}_L^+) - n_L(-ik\boldsymbol{N}_L^+) \right\} \\
&= -k^2\sqrt{\frac{\mu_0}{\epsilon_0}} \sum_L \left\{ m_L\boldsymbol{M}_L^+ + n_L\boldsymbol{N}_L^+ \right\} \\
&= -k^2 Z_0 \sum_L \left\{ m_L\boldsymbol{M}_L^+ + n_L\boldsymbol{N}_L^+ \right\}. \qquad (13.37)
\end{aligned}
$$

where $Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}$ is the usual impedance of free space, around 377 ohms.

Wow! Recall that the $\mathbf{M}$ waves are transverse, so the m_L and n_L are the magnetic (transverse electric) and electric (transverse magnetic) multipole moments respectively. The field outside of the source is a *pure* expansion in elementary transverse multipoles. (Later we will show that the (approximate) definitions we have used to date as "multipoles" are the limiting forms of these exact definitions.)

Note well that the actual fields require only *two* of the basic hansen solutions – the two that are mutually transverse. Something happened to the longitudinal part and the dependence of the field on the scalar potential. To see just what, let us re-evaluate the electric field from:

$$
\begin{aligned}
\boldsymbol{E} &= -\boldsymbol{\nabla}\Phi - \frac{\partial \mathbf{A}}{\partial t} \\
&= -\boldsymbol{\nabla}\left(\frac{ik}{\epsilon_0} \sum_L p_L H_L^+ \right) + i\omega\left(ik\mu_0 \sum_L \left\{ m_L\boldsymbol{M}_L^+(\boldsymbol{r}) + n_L\boldsymbol{N}_L^+(\boldsymbol{r}) + l_L\boldsymbol{L}_L(\boldsymbol{r}) \right\} \right) \\
&= -\frac{ik}{\epsilon_0} \sum_L \left\{ p_L(\boldsymbol{\nabla} H_L^+) - ikc\mu_0\epsilon_0 \sum_L l_L\boldsymbol{L}_L^+ \right\} - k^2\mu_0 c \sum_L \left\{ m_l\boldsymbol{M}_L^+ + n_L\boldsymbol{N}_L^+ \right\} \\
&= \frac{k^2}{\epsilon_0} \sum_L \left\{ p_L - \frac{1}{c} l_L \right\} \boldsymbol{L}_L^+ - k^2 Z_0 \sum_L \left\{ m_l\boldsymbol{M}_L^+ + n_L\boldsymbol{N}_L^+ \right\} \qquad (13.38)
\end{aligned}
$$

(Note that we used $\omega = kc$ and $\boldsymbol{\nabla} H_L^+ = ik\mathbf{L}_L^+$.) From this we see that *if* the gauge condition:

$$l_L = cp_L \tag{13.39}$$

is satisfied, *the scalar and longitudinal vector parts of the electric field cancel exactly!* All that survives are the transverse parts:

$$\boldsymbol{E} = -k^2 Z_0 \sum_L \left\{ m_L \boldsymbol{M}_L^+ + n_L \boldsymbol{N}_L^+ \right\} \tag{13.40}$$

as before. The Lorentz gauge condition is thus intimately connected to the vanishing of a scalar or longitudinal contribution to the $\boldsymbol{E}$ field! Also note that the magnitude of $\boldsymbol{E}$ is greater than that of $\boldsymbol{B}$ by c, the velocity of light.

Now, we are interested (as usual) mostly in obtaining the fields in the far zone, where this already simple expression attains a clean asymptotic form. Using the $kr \to \infty$ form of the hankel function,

$$\lim_{kr\to\infty} h_\ell^+(kr) = \frac{e^{ikr-(\ell+1)\frac{i\pi}{2}}}{kr} \tag{13.41}$$

we obtain the limiting forms (for $kr \to \infty$):

$$\boldsymbol{M}_L^+ \sim \frac{e^{ikr-(\ell+1)\frac{i\pi}{2}}}{kr} \boldsymbol{Y}_{\ell\ell}^m = (-i)^{\ell+1} \frac{e^{ikr}}{kr} \boldsymbol{Y}_{\ell\ell}^m \tag{13.42}$$

$$\boldsymbol{N}_L^+ \sim \frac{e^{ikr-\ell\frac{i\pi}{2}}}{kr} \left[\sqrt{\frac{\ell+1}{2\ell+1}} \boldsymbol{Y}_{\ell,\ell-1}^m + \sqrt{\frac{\ell}{2\ell+1}} \boldsymbol{Y}_{\ell,\ell+1}^m \right] \tag{13.43}$$

The bracket in the second equation can be simplified, using the results of the table I handed out previously. Note that

$$\left[\sqrt{\frac{\ell+1}{2\ell+1}} \boldsymbol{Y}_{\ell,\ell-1}^m + \sqrt{\frac{\ell}{2\ell+1}} \boldsymbol{Y}_{\ell,\ell+1}^m \right] = i(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m) = -e^{-i\frac{\pi}{2}} (\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m) \tag{13.44}$$

so that (still in the far zone limit)

$$\boldsymbol{N}_L^+ \sim -\frac{e^{ikr-(\ell+1)\frac{i\pi}{2}}}{kr} (\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m) = -(-i)^{\ell+1} \frac{e^{ikr}}{kr} (\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m). \tag{13.45}$$

Let us pause to admire this result before moseying on. This is just

$$\boldsymbol{B} = -k^2 \mu_0 \frac{e^{ikr}}{kr} \sum_L (-i)^{\ell+1} \left\{ m_L (\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m) + n_L \boldsymbol{Y}_{\ell\ell}^m \right\}$$

$$
\begin{aligned}
&= -k\mu_0 \frac{e^{ikr}}{r} \sum_L (-i)^{\ell+1} \left\{ m_L \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m \right) + n_L \boldsymbol{Y}_{\ell\ell}^m \right\} && (13.46) \\
\boldsymbol{E} &= -k^2 Z_0 \frac{e^{ikr}}{kr} \sum_L (-i)^{\ell+1} \left\{ m_L \boldsymbol{Y}_{\ell\ell}^m - n_L \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m \right) \right\} \\
&= -k Z_0 \frac{e^{ikr}}{r} \sum_L (-i)^{\ell+1} \left\{ m_L \boldsymbol{Y}_{\ell\ell}^m - n_L \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m \right) \right\}. && (13.47)
\end{aligned}
$$

If I have made a small error at this point, forgive me. Correct me, too. This is a purely transverse outgoing spherical wave whose vector character is finally translucent, if not transparent.

The power flux in the outgoing wave is still not too easy to express, but it is a damn sight easier than it was before. At least we have the satisfaction of knowing that we can express it as a general result. Recalling (as usual)

$$
\boldsymbol{S} = \frac{1}{2} \mathrm{Re}(\boldsymbol{E} \times \boldsymbol{H}^*) \tag{13.48}
$$

and that the power distribution is related to the flux of the Poynting vector through a surface at distance r in a differential solid angle $d\Omega$:

$$
\frac{dP}{d\Omega} = \frac{1}{2} \mathrm{Re}[r^2 \hat{\boldsymbol{n}} \cdot (\boldsymbol{E} \times \boldsymbol{H}^*)] \tag{13.49}
$$

we get

$$
\begin{aligned}
\boldsymbol{S} &= \frac{k^2}{2r^2} Z_0 \mathrm{Re} \left[\sum_L \sum_{L'} i^{\ell'-\ell} \left\{ m_L \boldsymbol{Y}_{\ell\ell}^m - n_L \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m \right) \right\} \right. \\
&\quad \times \left. \left\{ m_{L'}^* \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell'\ell'}^{m'*} \right) + n_{L'}^* \boldsymbol{Y}_{\ell'\ell'}^{m'*} \right\} \right] && (13.50)
\end{aligned}
$$

(Note: Units here need to be rechecked, but they appear to be consistent at first glance).

This is an extremely complicated result, but it has to be, since it expresses the *most general possible angular distribution of radiation* (in the far zone). The power distribution follows trivially. We can, however, evaluate the total power radiated, which is a very useful number. This will be an exercise. You will need the results

$$
\begin{aligned}
\int d^2\Omega \hat{r} \cdot \boldsymbol{Y}_{\ell\ell}^m \times \left(\hat{r} \times \boldsymbol{Y}_{\ell'\ell'}^{m'*} \right) &= \int d^2\Omega \boldsymbol{Y}_{\ell\ell}^m \cdot \boldsymbol{Y}_{\ell'\ell'}^{m'*} \\
&= \delta_{\ell\ell'} \delta_{mm'} && (13.51)
\end{aligned}
$$

and

$$\int d^2\Omega \hat{r} \cdot \left(\boldsymbol{Y}_{\ell\ell}^{m} \times \boldsymbol{Y}_{\ell'\ell'}^{m'*} \right) = 0 \tag{13.52}$$

to evaluate typical terms. Using these relations, it is not too difficult to show that

$$P = \frac{k^2}{2} Z_0 \sum_L \left\{ | \ m_L \ |^2 + | \ n_L \ |^2 \right\} \tag{13.53}$$

which is the sum of the power emitted from all the individual multipoles (there is no interference between multipoles!).

Let us examine e.g. the electric multipolar moment n_L to see how it compares to the usual static results. Static results are obtained in the $k \to 0$ (long wavelength) limit. In this limit e.g. $j_\ell(kr) \sim k^\ell r^\ell$ and:

$$n_L \approx ic \sqrt{\frac{\ell+1}{\ell}} \frac{k^\ell}{(2\ell+1)!} \int \rho(\boldsymbol{r}) r^\ell Y_{\ell,m}(\hat{\boldsymbol{r}}) d^3r \tag{13.54}$$

The dipole term comes from $\ell = 1$. For a simple dipole:

$$\begin{aligned} n_{1,m} &\approx ic\frac{\sqrt{2}}{3} k \int \rho r Y_{1,m} d^3r \\ &\approx i\frac{kc\sqrt{2}}{3}\sqrt{\frac{3}{4\pi}} e < r > \\ &\approx i\frac{kc\sqrt{6}}{36\pi} e < r > \\ &\approx -\frac{ie}{\sqrt{6\pi}\omega} < \ddot{r} > \end{aligned} \tag{13.55}$$

where we use $< \ddot{r} >= -\omega^2 < r >$.

In terms of this the average power radiated by a single electron dipole is:

$$P = \frac{1}{2}\left(\frac{e^2}{6\pi\epsilon_0 c^3}\right) |\ddot{r}|^2 \tag{13.56}$$

which compares well with the *Larmor Formula*:

$$P = \frac{2}{3}\left(\frac{e^2}{4\pi\epsilon_0 c^3}\right) |\ddot{r}|^2 \tag{13.57}$$

The latter is the formula for the instantaneous power radiated from a point charge as it is accelerated. Either flavor is the death knell of classical mechanics – it is very difficult to build a model for a stable atom based on classical trajectories of an electron around a nucleus that does not involve acceleration of the electron in question.

While it is not easy to see, the results above are essentially those obtained in Jackson (J9.155) except that (comparing e.g. J9.119, J9.122, and J91.165 to related results above) Jackson's $a_{E,M}(\ell, m)$ moments differ from the Hansen multipolar moments by factors of several powers of k. If one works hard enough, though, one can show that the results are identical, and even though Jackson's algebra is more than a bit Evil it is worthwhile to do this if only to validate the results above (where recall there has been a unit conversion and hence they do need validation).

Another useful exercise is to recover our old friends, the dipole and quadrupole radiation terms of J9 from the exact definition of their respective moments. One must make the long wavelength approximation under the integral in the definition of the multipole moments, integrate by parts liberally, and use the continuity equation. This is quite difficult, as it turns out, unless you have seen it before, so let us look at an example. Let us apply the methods we have developed above to obtain the radiation pattern of a dipole antenna, this time *without* assuming that it's length is small w.r.t. a wavelength. Jackson solves more or less the same problem in his section 9.12, so this will permit the direct comparison of the coefficients and constants in the final expressions for total radiated power or the angular distribution of power.

13.4 A Linear Center-Fed Half-Wave Antenna

Suppose we are given a center-fed dipole antenna with length $\lambda/2$ (half-wave antenna). We will assume further that the antenna is aligned with the z axis and centered on the origin, with a current given by:

$$I = I_0 \cos(\omega t) \cos\left(\frac{2\pi z}{\lambda}\right) \tag{13.58}$$

Note that in "real life" it is not easy to arrange for a given current because the current instantaneously depends on the "resistance" which is a function of the radiation field itself. The current itself thus comes out of the solution of an extremely complicated **boundary value problem**. For atomic or nuclear radiation, however, the "currents" are generally matrix elements associated with transitions and hence are known.

In any event, the current density corresponding to *this* current is

$$\boldsymbol{J} = \hat{\boldsymbol{z}} I_0 \cos\left(\frac{2\pi r}{\lambda}\right) \frac{\delta(1 - |\cos\theta|)}{2\pi r^2 \sin\theta} \tag{13.59}$$

for $r \leq \lambda/4$ and

$$\boldsymbol{J} = 0 \tag{13.60}$$

for $r > \lambda/4$.

When we use the Hansen multipoles, there is little incentive to convert this into a form where we integrate against the charge density in the antenna. Instead we can easily and directly calculate the multipole moments. The magnetic moment is

$$\begin{aligned} m_L &= \int \boldsymbol{J} \cdot \boldsymbol{M}_L^{0*} d^3 r \\ &= \frac{I_0}{2\pi} \int_0^{2\pi} \int_0^{\lambda/4} \cos(kr) j_\ell(kr) \left\{ \hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell\ell}^{m*}(0,\phi) + \hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell\ell}^{m*}(\pi,\phi) \right\} d\phi dr \end{aligned} \tag{13.61}$$

(where we have done the integral over θ). Now,

$$\hat{z} \cdot \boldsymbol{Y}_{\ell\ell}^m = \frac{1}{\sqrt{\ell(\ell+1)}} m Y_L \tag{13.62}$$

(Why? Consider $(\hat{\boldsymbol{z}} \cdot \boldsymbol{L}) Y_L$)...) and yet

$$Y_L(0,\phi) = \delta_{m0} \left[\frac{2\ell+1}{4\pi}\right]^{1/2} \tag{13.63}$$

$$Y_L(\pi,\phi) = (-1)^\ell \delta_{m0} \left[\frac{2\ell+1}{4\pi}\right]^{1/2}. \tag{13.64}$$

Consequently, we can conclude ($m\delta_{m0} = 0$) that

$$m_L = 0. \tag{13.65}$$

All magnetic multipole moments of this linear dipole vanish. Since the magnetic multipoles should be connected to the rotational part of the current density (which is zero for linear flow) this should not surprise you.

The electric moments are

$$\begin{aligned} n_L &= \int \boldsymbol{J} \cdot \boldsymbol{N}_L^{0*} d^3r \\ &= \frac{I_0}{2\pi} \int_0^{2\pi} \int_0^{\lambda/4} \cos(kr) \Bigg\{ \sqrt{\frac{\ell+1}{2\ell+1}} j_{\ell-1}(kr) \left[\hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell,\ell-1}^{m*}(0,\phi) + \hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell,\ell+1}^{m*}(\pi,\phi) \right] d\phi \, dr \\ &\quad - \sqrt{\frac{\ell}{2\ell+1}} j_{\ell+1}(kr) \left[\hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell,\ell+1}^{m*}(0,\phi) + \hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell,\ell-1}^{m*}(\pi,\phi) \right] \Bigg\}. \end{aligned} \tag{13.66}$$

If we look up the definition of the v.s.h.'s on the handout table, the z components are given by:

$$\hat{z} \cdot \boldsymbol{Y}_{\ell,\ell-1}^{m*}(0,\phi) = \delta_{m0} \sqrt{\frac{\ell}{4\pi}} \tag{13.67}$$

$$\hat{z} \cdot \boldsymbol{Y}_{\ell,\ell-1}^{m*}(\pi,\phi) = (-1)^{\ell-1} \delta_{m0} \sqrt{\frac{\ell}{4\pi}} \tag{13.68}$$

$$\hat{z} \cdot \boldsymbol{Y}_{\ell,\ell+1}^{m*}(0,\phi) = -\delta_{m0} \sqrt{\frac{\ell+1}{4\pi}} \tag{13.69}$$

$$\hat{z} \cdot \boldsymbol{Y}_{\ell,\ell+1}^{m*}(\pi,\phi) = -(-1)^{\ell-1} \delta_{m0} \sqrt{\frac{\ell+1}{4\pi}} \tag{13.70}$$

so the electric multipole moments vanish for $m \neq 0$, and

$$n_{\ell,0} = I_0 \delta_{m0} \sqrt{\frac{\ell(\ell+1)}{4\pi(2\ell+1)}} \left(1 + (-1)^{\ell+1}\right) \int_0^{\lambda/4} \cos(kr) \Big[j_{\ell-1}(kr) + j_{\ell+1}(kr) \Big] dr. \tag{13.71}$$

Examining this equation, we see that all the **even** ℓ terms vanish! However, all the odd ℓ, $m = 0$ terms do *not* vanish, so we can't quit yet. We use the following relations:

$$j_{\ell-1} + j_{\ell+1} = \frac{2\ell+1}{kr} j_\ell \tag{13.72}$$

(the fundamental recursion relation),

$$n_0(kr) = -\frac{\cos(kr)}{kr} \tag{13.73}$$

(true fact) and

$$\int dz \, f_\ell(z) g_{\ell'}(z) = \frac{z^2}{[\ell'(\ell'+1) - \ell(\ell+1)]} \left(f_\ell' g_{\ell'} - f_\ell g_{\ell'}' \right) \tag{13.74}$$

for any two spherical bessel type functions (a valuable thing to know that follows from integration by parts and the recursion relation). From these we get

$$n_{\ell,0} = \frac{\pi I_0}{2k}\delta_{m0}\sqrt{\frac{2\ell+1}{4\pi\ell(\ell+1)}}\left(1+(-1)^{\ell+1}\right)j_\ell(\pi/2). \tag{13.75}$$

Naturally, there is a wee tad of algebra involved here that I have skipped. You shouldn't. Now, let's figure out the power radiated from this source. Recall from above that:

$$\begin{aligned} P &= \frac{k^2}{2}\sqrt{\frac{\mu_0}{\epsilon_0}}\sum_L\left\{\mid m_L\mid^2 + \mid n_L\mid^2\right\} \\ &= \frac{k^2}{2}\sqrt{\frac{\mu_0}{\epsilon_0}}\sum_{\ell \text{ odd}}|n_{\ell,0}|^2 \\ &= \frac{\pi I_0^2}{8}\sqrt{\frac{\mu_0}{\epsilon_0}}\sum_{\ell \text{ odd}}\left(\frac{2\ell+1}{\ell(\ell+1)}\right)\left[j_\ell(\pi/2)\right]^2 \end{aligned} \tag{13.76}$$

Now this also equals (recall) $\frac{1}{2}I_0^2 R_{\text{rad}}$, from which we can find the radiation resistance of the half wave antenna:

$$R_{\text{rad}} = \frac{\pi}{4}\sqrt{\frac{\mu_0}{\epsilon_0}}\sum_{\ell \text{ odd}}\left(\frac{2\ell+1}{\ell(\ell+1)}\right)\left[j_\ell(\pi/2)\right]^2. \tag{13.77}$$

We are blessed by this having manifest units of resistance, as we recognize our old friend $Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 377\Omega$ (the impedance of free space) and a bunch of dimensionless numbers! In terms of this:

$$R_{\text{rad}} = Z_0\left(\frac{\pi}{4}\sum_{\ell \text{ odd}}\left(\frac{2\ell+1}{\ell(\ell+1)}\right)\left[j_\ell(\pi/2)\right]^2\right). \tag{13.78}$$

We can obtain a good estimate of the magnitude by evaluating the first few terms. Noting that

$$j_1(\pi/2) = \left(\frac{2}{\pi}\right)^2 \tag{13.79}$$

$$j_3(\pi/2) = \left(\frac{2}{\pi}\right)^2\left[\frac{60}{\pi^2} - 6\right] \tag{13.80}$$

and doing some arithmetic, you should be able to show that $R_{\text{rad}} = 73.1\Omega$.

Note that the ratio of the first (dipole) term to the third (octupole) term is

$$\begin{aligned}\left|\frac{n_3}{n_1}\right|^2 &= \frac{7}{12}\frac{2}{3}\left[\frac{60}{\pi^2}-6\right]^2 \\ &= \frac{7}{18}\left[\frac{60}{\pi^2}-6\right]^2 \approx 0.00244\end{aligned}$$

That means that this is likely to be a *good* approximation (the answer is very nearly unchanged by the inclusion of the extra term). Even if the length of the antenna is on the order of λ, the multipole expansion is an extremely accurate and rapidly converging approximation. That is, after all, why we use it so much in all kinds of localized source wave theory.

However, if we plug in the "long wavelength" approximation we previously obtained for a **short** dipole antenna (with $d = \lambda/2$) we get:

$$R_{\mathrm{rad}} = \frac{(kd)^2}{24\pi}\sqrt{\frac{\mu_0}{\epsilon_0}} \approx 48\Omega \tag{13.81}$$

which is off by close to a factor of 50%. This is not such a good result. Using this formula with a long wavelength approximation for the dipole moment (only) of

$$n_{1,0} \approx \frac{I_0}{k}\sqrt{\frac{2}{3\pi}} \tag{13.82}$$

yields $R_{\mathrm{rad}} \approx 80\Omega$, still off by 11%.

13.5 Connection to Old (Approximate) Multipole Moments

To conclude our discussion of multipole fields, let us relate the multipole moments defined and used above (which are *exact*) to the "usual" static, long wavelength moments we deduced in our earlier studies. Well,

$$n_L = \int \mathbf{J}\cdot\mathbf{N}_L^{0*} d^3r \tag{13.83}$$

and

$$\boldsymbol{N}_L = \frac{1}{\sqrt{\ell(\ell+1)}}\frac{1}{k}\boldsymbol{\nabla}\times(\boldsymbol{r}\times\boldsymbol{\nabla})(f_\ell(kr)Y_L(\hat{\boldsymbol{r}}))$$

$$= \frac{1}{\sqrt{\ell(\ell+1)}}\frac{1}{k}\left[\boldsymbol{r}\nabla^2 - \boldsymbol{\nabla}\left(r\frac{\partial}{\partial r}+1\right)\right](f_\ell(kr)Y_L(\hat{\boldsymbol{r}})) \quad (13.84)$$

(using the vector identity

$$\boldsymbol{\nabla}\times\boldsymbol{L} = i\left[\boldsymbol{r}\nabla^2 - \boldsymbol{\nabla}\left(r\frac{\partial}{\partial r}+1\right)\right] \quad (13.85)$$

to simplify). Then

$$\begin{aligned} n_L &= \frac{-1}{k\sqrt{\ell(\ell+1)}}\left\{k^2\int(\boldsymbol{r}\cdot\boldsymbol{J})j_\ell(kr)Y_L^*(\hat{\boldsymbol{r}})d^3r + \right. \\ &\qquad \left. \int(\boldsymbol{J}\cdot\boldsymbol{\nabla})\left[Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\right]d^3r\right\} \end{aligned} \quad (13.86)$$

Now, (from the continuity equation)

$$\boldsymbol{\nabla}\cdot\boldsymbol{J} = i\omega\rho \quad (13.87)$$

so when we (sigh) integrate the second term by parts, (by using

$$\boldsymbol{\nabla}\cdot(a\boldsymbol{B}) = \boldsymbol{B}\cdot\boldsymbol{\nabla}a + a\boldsymbol{\nabla}\cdot\boldsymbol{B} \quad (13.88)$$

so that

$$(\boldsymbol{J}\cdot\boldsymbol{\nabla})\left[Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\right] = \boldsymbol{\nabla}\cdot\left[\boldsymbol{J}Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\right] - Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\left[\boldsymbol{\nabla}\cdot\boldsymbol{J}\right] \quad (13.89)$$

and the divergence theorem on the first term,

$$\begin{aligned} \int_V \boldsymbol{\nabla}\cdot\left[\boldsymbol{J}Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\right]dV &= \int_{\partial V\to\infty}\hat{n}\cdot\left[\boldsymbol{J}Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\right]dA \\ &= 0 \end{aligned} \quad (13.90)$$

for sources with compact support to do the integration) we get

$$\begin{aligned} n_L &= \frac{-1}{k\sqrt{\ell(\ell+1)}}\left\{k^2\int(\boldsymbol{r}\cdot\boldsymbol{J})j_\ell(kr)Y_L^*(\hat{\boldsymbol{r}})d^3r - \right. \\ &\qquad \left. \int(i\omega\rho(\boldsymbol{r}))\left[Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\right]d^3r\right\} \\ &= \frac{ic}{\sqrt{\ell(\ell+1)}}\int\rho(\boldsymbol{r})\left[Y_L^*(\hat{\boldsymbol{r}})\frac{\partial}{\partial r}(rj_\ell(kr))\right]d^3r \\ &\qquad -\frac{k}{\sqrt{\ell(\ell+1)}}\int(\boldsymbol{r}\cdot\boldsymbol{J})j_\ell(kr)Y_L^*(\hat{\boldsymbol{r}})d^3r \end{aligned} \quad (13.91)$$

The electric multipole moment thus consists of two terms. The first term appears to arise from oscillations of the charge density itself, and might be expected to correspond to our usual definition. The second term is the contribution to the radiation from the **radial** oscillation of the current density. (Note that it is the **axial** or **transverse** current density oscillations that give rise to the magnetic multipoles.)

Only if the wavelength is much larger than the source is the second term of lesser order (by a factor of $\frac{ik}{c}$). In that case we can write

$$n_L \approx \frac{ic}{\sqrt{\ell(\ell+1)}} \int \rho Y_L^* \frac{\partial}{\partial r}(r j_\ell(kr)) d^3r. \tag{13.92}$$

Finally, using the long wavelength approximation on the bessel functions,

$$\begin{aligned} n_L &\approx \frac{ic}{(2\ell+1)!!}\sqrt{\frac{\ell+1}{\ell}} k^\ell \int \rho r^\ell Y_L^* d^3r & (13.93) \\ &\approx \frac{ic}{(2\ell+1)!!}\sqrt{\frac{\ell+1}{\ell}} k^\ell q_{\ell,m} & (13.94) \end{aligned}$$

and the connection with the static electric multipole moments $q_{\ell,m}$ is complete. In a similar manner one can establish the long wavelength connection between the m_L and the magnetic moments of earlier chapters. Also note well that the relationship is **not equality**. The "approximate" multipoles need to be renormalized in order to fit together properly with the Hansen functions to reconstruct the EM field.

13.6 Angular Momentum Flux

Let us consider the angular momentum radiated away with the electromagnetic field. The angular momentum flux density is basically vr crossed into the momentum density $\boldsymbol{S}/c$ or:

$$\boldsymbol{\mathcal{L}} = \frac{1}{2}\mathrm{Re}\left\{ \frac{\boldsymbol{r} \times (\boldsymbol{E} \times \boldsymbol{H}^*)}{c} \right\} \tag{13.95}$$

Into this expression we must substitute our expressions for $\boldsymbol{E}$ and $\boldsymbol{H}$:

$$\begin{aligned} \boldsymbol{E} &= -k^2 Z_0 \sum_L \left\{ m_L \boldsymbol{M}_L^+ + n_L \boldsymbol{N}_L^+ \right\} & (13.96) \\ \boldsymbol{H} &= k^2 \sum_L \left\{ m_L \boldsymbol{N}_L^+ - n_L \boldsymbol{M}_L^+ \right\}. & (13.97) \end{aligned}$$

If we try to use the asymptotic far field results:

$$\boldsymbol{E} = -kZ_0 \frac{e^{ikr}}{r} \sum_L (-i)^{\ell+1} \left\{ m_L \boldsymbol{Y}_{\ell\ell}^m - n_L \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m \right) \right\} \tag{13.98}$$

$$\boldsymbol{H} = -k \frac{e^{ikr}}{r} \sum_L (-i)^{\ell+1} \left\{ m_L \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m \right) + n_L \boldsymbol{Y}_{\ell\ell}^m \right\} \tag{13.99}$$

we get:

$$\begin{aligned}
\boldsymbol{E} \times \boldsymbol{H}^* &= \frac{k^2 Z_0}{r^2} \sum_L \sum_{L'} i^{\ell-\ell'} \left\{ m_L \boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) - n_L \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \right) \right\} \times \\
&\qquad \left\{ m_{L'}^* \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}}) \right) + n_{L'}^* \boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}}) \right\} \\
&= \frac{k^2 Z_0}{r^2} \sum_L \sum_{L'} i^{\ell-\ell'} \left\{ m_L m_{L'}^* \boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \times \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}}) \right) \right. \\
&\qquad + m_L n_{L'}^* \boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \times \boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}}) \\
&\qquad - n_L m_{L'}^* \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \right) \times \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}}) \right) \\
&\qquad \left. - n_L n_{L'}^* \left(\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \right) \times \boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}}) \right\}.
\end{aligned} \tag{13.100}$$

With some effort this can be shown to be a *radial* result – the Poynting vector points directly away from the source in the far field to leading order. Consequently, this leading order behavior contributes *nothing* to the angular momentum flux. We must keep at least the leading correction term to the asymptotic result.

It is convenient to use a radial/tangential decomposition of the Hansen solutions. The $\boldsymbol{M}_L$ are completely tangential (recall $\boldsymbol{r} \cdot \boldsymbol{M}_L = 0$). For the $\boldsymbol{N}_L$ we have:

$$\boldsymbol{N}_L(\boldsymbol{r}) = \frac{1}{kr} \frac{d}{dr} \left(r f_\ell(kr) \right) \left(i \hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \right) - \hat{\boldsymbol{r}} \frac{\sqrt{\ell(\ell+1)}}{kr} f_e ll(kr) Y_L(\hat{\boldsymbol{r}}) \tag{13.101}$$

Using our *full* expressions for $\boldsymbol{E}$ and $\boldsymbol{H}^*$:

$$\boldsymbol{E} = -k^2 Z_0 \sum_L \left\{ m_L \boldsymbol{M}_L^+ + n_L \boldsymbol{N}_L^+ \right\} \tag{13.102}$$

$$\boldsymbol{H} = k^2 \sum_L \left\{ m_L \boldsymbol{N}_L^+ - n_L \boldsymbol{M}_L^+ \right\} \tag{13.103}$$

with this form substituted for $\boldsymbol{N}_L$ and the usual form for $\boldsymbol{M}_L$ we get:

$$\boldsymbol{\mathcal{L}} = \frac{1}{2} \mathrm{Re} \left\{ \frac{\boldsymbol{r} \times (\boldsymbol{E} \times \boldsymbol{H}^*)}{c} \right\}$$

$$
\begin{aligned}
&= -\frac{k^4 Z_0}{2c}\mathrm{Re}\sum_L\sum_{L'} \boldsymbol{r}\times\Big\{m_L h_\ell^+(kr)\boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \\
&\quad +n_L\left[\frac{1}{kr}\frac{d(rh_\ell^+(kr))}{dr}(i\hat{\boldsymbol{r}}\times\boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}})) - \hat{\boldsymbol{r}}\sqrt{\ell(\ell+1)}\frac{h_\ell^+(kr)}{kr}Y_L(\hat{\boldsymbol{r}})\right]\Big\} \\
&\quad \times\Big\{m_{L'}^*\left[\frac{1}{kr}\frac{d(rh_{\ell'}^-(kr))}{dr}(-i\hat{\boldsymbol{r}}\times\boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}})) - \hat{\boldsymbol{r}}\sqrt{\ell(\ell+1)}\frac{h_{\ell'}^-(kr)}{kr}Y_{L'}^*(\hat{\boldsymbol{r}})\right] \\
&\quad +n_{L'}^* h_{\ell'}^-(kr)\boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}})\Big\}
\end{aligned}
\tag{13.104}
$$

All the purely radial terms in the outermost $\Big\{\Big\}$ under the sum do not contribute to the angular momentum flux density. The surviving terms are:

$$
\begin{aligned}
\boldsymbol{\mathcal{L}} &= -\frac{k^4 Z_0}{2c}\mathrm{Re}\sum_L\sum_{L'} \boldsymbol{r}\times\Big\{m_L m_{L'}^* h_\ell^+(kr)\sqrt{\ell'(\ell'+1)}\frac{h_{\ell'}^-(kr)}{kr}(\boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}})\times\hat{\boldsymbol{r}})Y_{L'}^*(\hat{\boldsymbol{r}}) \\
&\quad +n_L m_{L'}^*\frac{1}{kr}\frac{d(rh_\ell^+(kr))}{dr}\left((i\hat{\boldsymbol{r}}\times\boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}})\times\hat{\boldsymbol{r}}\right)\sqrt{\ell'(\ell'+1)}\frac{h_{\ell'}^-(kr)}{kr}Y_{L'}^*(\hat{\boldsymbol{r}}) \\
&\quad -n_L m_{L'}^*\frac{h_\ell^+(kr)}{kr}\sqrt{\ell(\ell+1)}\frac{1}{kr}\frac{d(rh_{\ell'}^-(kr))}{dr}\left(i\hat{\boldsymbol{r}}\times(\hat{\boldsymbol{r}}\times\boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}}))\right) \\
&\quad -n_L n_{L'}^*\frac{h_\ell^+(kr)}{kr}\sqrt{\ell(\ell+1)}h_{\ell'}^-(kr)(\hat{\boldsymbol{r}}\times\boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}})\Big\}
\end{aligned}
\tag{13.105}
$$

The lowest order term in the asymptotic form for the spherical bessel functions makes a contribution in the above expressions. After untangling the cross products and substituting the asymptotic forms, we get:

$$
\begin{aligned}
\boldsymbol{\mathcal{L}} &= \frac{k\mu_0}{2r^2}\mathrm{Re}\sum_L\sum_{L'}\Big\{m_L m_{L'}\sqrt{\ell'(\ell'+1)}i^{\ell'-\ell}Y_{L'}^*(\hat{\boldsymbol{r}})\boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}}) \\
&\quad -n_L m_{L'}^*\sqrt{\ell(\ell+1)}i^{\ell'-\ell}Y_{L'}^*(\hat{\boldsymbol{r}})\left(\hat{\boldsymbol{r}}\times\boldsymbol{Y}_{\ell\ell}^m(\hat{\boldsymbol{r}})\right) \\
&\quad +n_L m_{L'}^*\sqrt{\ell(\ell+1)}i^{\ell'-\ell}Y_L(\hat{\boldsymbol{r}})\left(\hat{\boldsymbol{r}}\times\boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}})\right) \\
&\quad +n_L n_{L'}^*\sqrt{\ell(\ell+1)}i^{\ell'-\ell}Y_L(\hat{\boldsymbol{r}})\boldsymbol{Y}_{\ell'\ell'}^{m'*}(\hat{\boldsymbol{r}})\Big\}
\end{aligned}
\tag{13.106}
$$

The angular momentum about a given axis emitted per unit time is obtained by selecting a particular component of this and integrating its flux through a distant spherical surface. For example, for the z-component we find (noting that r^2 cancels as it should):

$$
\frac{dL_z}{dt} = \frac{k\mu_0}{2}\mathrm{Re}\sum_L\sum_{L'}\int\hat{\boldsymbol{z}}\cdot\Big\{...\Big\}\sin(\theta)d\theta d\phi \tag{13.107}
$$

where the brackets indicate the expression above. We look up the components of the vector harmonics to let us do the dot product and find:

$$\begin{aligned}
\hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell\ell}^{m} &= \frac{m}{\sqrt{\ell(\ell+1)}} Y_{\ell,m} && (13.108)\\
\hat{\boldsymbol{z}} \cdot (\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^{m}) &= -i\left[\sqrt{\frac{\ell+1}{2\ell+1}}\hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell,\ell-1}^{m} + \sqrt{\frac{\ell}{2\ell+1}}\hat{\boldsymbol{z}} \cdot \boldsymbol{Y}_{\ell\ell+1}^{m}\right]\\
&= -i\left[\sqrt{\frac{(\ell+1)(\ell^2-m^2)}{\ell(2\ell-1)(2\ell+1)}} Y_{\ell-1,m} - \sqrt{\frac{[(\ell+1)^2-m^2]\ell}{(2\ell+1)(2\ell+3)(\ell+1)}} Y_{\ell+1,m}\right] && (13.109)
\end{aligned}$$

Doing the integral is now simple, using the orthonormality of the spherical harmonics. One obtains (after still more work, of course):

$$\frac{dL_z}{dt} = \frac{k\mu_0}{2}\sum_L m\left(|m_L|^2 + |n_L|^2\right) \qquad (13.110)$$

Compare this to:

$$P = \frac{k^2}{2} Z_0 \sum_L \left\{| \, m_L \, |^2 + | \, n_L \, |^2\right\} \qquad (13.111)$$

term by term. For example:

$$\begin{aligned}
\frac{dL_z(m_L)}{dt} &= \frac{k\mu_0 m}{2}\left\{\frac{2}{k^2\mu_0 c} P(m_L) = |m_L|^2\right\}\\
&= \frac{m}{\omega} P(m_L) && (13.112)
\end{aligned}$$

(where m in the fraction is the spherical harmonic m, not the multipole m_L). In other words, for a pure multipole the rate of angular momentum about any given axis transferred is m/ω times the rate of energy transferred, where m is the angular momentum aligned with that axis. (Note that if we chose some other axis we could, with enough work, find an answer, but the *algebra* is only simple along the z-axis as the multipoles were originally defined with their m-index referred to this axis. Alternatively we could rotate frames to align with the new direction and do the entire computation over.)

This is quite profound. If we insist, for example, that energy be transferred in units of $\hbar\omega$, then angular momentum is *also* transferred in units of $m\hbar$!

13.7 Concluding Remarks About Multipoles

There are still many, many things we could study concerning multipoles and radiation. For example, we have not yet done a magnetic loop antenna, but doing one should now be straightforward (to obtain a magnetic dipole radiation field to leading order). Hmmm, sounds like a homework or exam problem to me...

Still, I hope that this has left you with enough fundamentals that you:

1. Understand bessel functions;
2. Understand spherical harmonics;
3. Understand at least something about vector spherical harmonics;
4. Know what a "multipolar expansion" is;
5. Know how to expand a variety of important Green's functions for vector and scalar Helmholtz equations (including the Poisson equation).
6. Know how to formulate an integral equation solution to these differential equations based on the Green's function, and at least formally solve it by partitioning the integral into domains of convergence.
7. Know how to describe the electromagnetic field at a variety of levels. These levels had better include the elementary description of the E1, E2, and M1 "static" levels as well as enough knowledge to be able to do it correctly for extended sources or sources where higher order moments are important, at least if your life or job or next paper depend on it.
8. Can pass prelims.

If you feel deficient in any of these areas, I recommend that you take the time to review and learn the material again, carefully. This has been the most important part of the course and is the one thing you should not fail to take out of here with you.

I hope you have enjoyed it.

13.8 Table of Properties of Vector Harmonics

1. Basic Definitions

$$
\begin{aligned}
\boldsymbol{Y}_{\ell\ell}^{m} &= \frac{1}{\sqrt{\ell(\ell+1)}}\boldsymbol{L}Y_{\ell,m} \\
\boldsymbol{Y}_{\ell\ell-1}^{m} &= -\frac{1}{\sqrt{\ell(2\ell+1)}}\left[-\ell\hat{\boldsymbol{r}} + i\hat{\boldsymbol{r}}\times\boldsymbol{L}\right]Y_{\ell,m} \\
\boldsymbol{Y}_{\ell\ell+1}^{m} &= -\frac{1}{\sqrt{(\ell+1)(2\ell+1)}}\left[(\ell+1)\hat{\boldsymbol{r}} + i\hat{\boldsymbol{r}}\times\boldsymbol{L}\right]Y_{\ell,m}
\end{aligned}
$$

2. Eigenvalues (j, ℓ, m are integral):

$$
\begin{aligned}
J^2\boldsymbol{Y}_{j\ell}^{m} &= j(j+1)\boldsymbol{Y}_{j\ell}^{m} \\
L^2\boldsymbol{Y}_{j\ell}^{m} &= \ell(\ell+1)\boldsymbol{Y}_{j\ell}^{m} \\
J_z\boldsymbol{Y}_{j\ell}^{m} &= m\boldsymbol{Y}_{j\ell}^{m}
\end{aligned}
$$

3. Projective Orthonormality:

$$
\int \boldsymbol{Y}_{j\ell}^{m}\cdot\boldsymbol{Y}_{j'\ell'}^{m'*}\,d\Omega = \delta_{jj'}\delta_{\ell\ell'}\delta_{mm'}
$$

4. Complex Conjugation:

$$
\boldsymbol{Y}_{j\ell}^{m*} = (-1)^{\ell+1-j}(-1)^{m}\boldsymbol{Y}_{j\ell}^{-m}
$$

5. Addition Theorem (LCB notes corrupt – this needs to be checked):

$$
\begin{aligned}
\boldsymbol{Y}_{j\ell}^{m*}\cdot\boldsymbol{Y}_{j'\ell'}^{m'} &= \sum_n (-1)^{m+1}\sqrt{\frac{(2\ell+1)(2\ell'+1)(2j'+1)(2j+1)}{4\pi(2n+1)}}\times \\
&\qquad C_{000}^{\ell\ell'n}C_{0,-m,m'}^{jj'n}W(j\ell j'\ell';n)Y_{n,(m'-m)}
\end{aligned}
$$

6. For F any function of r only:

$$
\begin{aligned}
\boldsymbol{\nabla}\cdot(\boldsymbol{Y}_{\ell\ell}^{m}F) &= 0 \\
\boldsymbol{\nabla}\cdot(\boldsymbol{Y}_{\ell\ell-1}^{m}F) &= \sqrt{\frac{\ell}{2\ell+1}}\left[(\ell-1)\frac{F}{r} - \frac{dF}{dr}\right]Y_{\ell,m} \\
\boldsymbol{\nabla}\cdot(\boldsymbol{Y}_{\ell\ell+1}^{m}F) &= \sqrt{\frac{\ell+1}{2\ell+1}}\left[(\ell+2)\frac{F}{r} - \frac{dF}{dr}\right]Y_{\ell,m}
\end{aligned}
$$

7. Ditto:

$$
\begin{aligned}
i\boldsymbol{\nabla} \times (\boldsymbol{Y}^m_{\ell\ell} F) &= \sqrt{\frac{\ell+1}{2\ell+1}} \left[(\ell+1)\frac{F}{r} + \frac{dF}{dr} \right] \boldsymbol{Y}^m_{\ell\ell-1} + \sqrt{\frac{\ell}{2\ell+1}} \left[-\ell\frac{F}{r} + \frac{dF}{dr} \right] \boldsymbol{Y}^m_{\ell\ell+1} \\
i\boldsymbol{\nabla} \times (\boldsymbol{Y}^m_{\ell\ell-1} F) &= -\sqrt{\frac{\ell+1}{2\ell+1}} \left[(\ell-1)\frac{F}{r} - \frac{dF}{dr} \right] \boldsymbol{Y}^m_{\ell\ell} \\
i\boldsymbol{\nabla} \times (\boldsymbol{Y}^m_{\ell\ell+1} F) &= \sqrt{\frac{\ell}{2\ell+1}} \left[(\ell+2)\frac{F}{r} - \frac{dF}{dr} \right] \boldsymbol{Y}^m_{\ell\ell}
\end{aligned}
$$

8. This puts the VSHs into vector form:

$$
\boldsymbol{Y}^m_{\ell\ell} = \begin{pmatrix} -\sqrt{\frac{(\ell+m)(\ell-m+1)}{2\ell(\ell+1)}} Y_{\ell,m-1} \\ \\ \frac{m}{\sqrt{\ell(\ell+1)}} Y_{\ell,m} \\ \\ \sqrt{\frac{(\ell-m)(\ell+m+1)}{2\ell(\ell+1)}} Y_{\ell,m+1} \end{pmatrix}
$$

$$
\boldsymbol{Y}^m_{\ell\ell-1} = \begin{pmatrix} \sqrt{\frac{(\ell+m-1)(\ell+m)}{2\ell(2\ell-1)}} Y_{\ell-1,m-1} \\ \\ \sqrt{\frac{(\ell-m)(\ell+m)}{\ell(2\ell-1)}} Y_{\ell-1,m} \\ \\ \sqrt{\frac{(\ell-m-1)(\ell-m)}{2\ell(2\ell-1)}} Y_{\ell-1,m+1} \end{pmatrix}
$$

$$
\boldsymbol{Y}^m_{\ell\ell+1} = \begin{pmatrix} \sqrt{\frac{(\ell-m+1)(\ell-m+2)}{2(\ell+1)(2\ell+3)}} Y_{\ell+1,m-1} \\ \\ \sqrt{\frac{(\ell-m+1)(\ell+m+1)}{(\ell+1)(2\ell+3)}} Y_{\ell+1,m} \\ \\ \sqrt{\frac{(\ell+m+2)(\ell+m+1)}{2(\ell+1)(2\ell+3)}} Y_{\ell+1,m+1} \end{pmatrix}
$$

9. Hansen Multipole Properties

$$
\begin{aligned}
\boldsymbol{\nabla} \cdot \boldsymbol{M}_L &= 0 \\
\boldsymbol{\nabla} \cdot \boldsymbol{N}_L &= 0 \\
\boldsymbol{\nabla} \cdot \boldsymbol{L}_L &= ikf_\ell(kr)Y_L(\hat{\boldsymbol{r}})
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\nabla} \times \boldsymbol{M}_L &= -ik\boldsymbol{N}_L \\
\boldsymbol{\nabla} \times \boldsymbol{N}_L &= ik\boldsymbol{M}_L \\
\boldsymbol{\nabla} \times \boldsymbol{L}_L &= 0
\end{aligned}
$$

10. Hansen Multipole Explicit Forms

$$
\begin{aligned}
\boldsymbol{M}_L &= f_\ell(kr)\boldsymbol{Y}_{\ell\ell}^m \\
\boldsymbol{N}_L &= \sqrt{\frac{\ell+1}{2\ell+1}} f_{\ell-1}(kr)\boldsymbol{Y}_{\ell,\ell-1}^m - \sqrt{\frac{\ell}{2\ell+1}} f_{\ell+1}(kr)\boldsymbol{Y}_{\ell,\ell+1}^m \\
\boldsymbol{L}_L &= \sqrt{\frac{\ell}{2\ell+1}} f_{\ell-1}(kr)\boldsymbol{Y}_{\ell,\ell-1}^m + \sqrt{\frac{\ell+1}{2\ell+1}} f_{\ell+1}(kr)\boldsymbol{Y}_{\ell,\ell+1}^m
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{M}_L &= f_\ell(kr)\boldsymbol{Y}_{\ell\ell}^m \\
\boldsymbol{N}_L &= \frac{1}{kr}\left\{\frac{d}{d(kr)}(krf_\ell)(i\hat{\boldsymbol{r}} \times \boldsymbol{Y}_{\ell\ell}^m) - \hat{\boldsymbol{r}}\sqrt{\ell(\ell+1)} f_\ell Y_L\right\} \\
\boldsymbol{L}_L &= \sqrt{\ell(\ell+1)}\frac{1}{kr}(i\hat{\boldsymbol{r}} \times f_\ell \boldsymbol{Y}_{\ell\ell}^m) - \hat{\boldsymbol{r}}\left[\frac{d}{d(kr)} f_\ell\right] Y_L
\end{aligned}
$$

Chapter 14

Optical Scattering

14.1 Radiation Reaction of a Polarizable Medium

Usually, when we consider optical scattering, we imagine that we have a **monochromatic plane wave** incident upon a **polarizable medium** embedded in (for the sake of argument) free space. The **target** we imagine is a "particle" of some shape and hence is mathematically a (simply) connected domain with compact support. The picture we must describe is thus

The incident wave (in the absence of the target) is thus a pure plane wave:

$$\begin{aligned} \boldsymbol{E}_{\text{inc}} &= \hat{\boldsymbol{\epsilon}}_0 E_0 e^{ik\hat{\boldsymbol{n}}_0\cdot\mathbf{r}} && (14.1)\\ \boldsymbol{H}_{\text{inc}} &= \hat{\boldsymbol{n}}_0 \times \boldsymbol{E}_{\text{inc}}/Z_0. && (14.2)\end{aligned}$$

The incident wave **induces** a time dependent polarization density into the medium. If we imagine (not unreasonably) that the target is a particle or atom much smaller than a wavelength, then we can describe the field radiated from its induced dipole moment in the **far zone** and **dipole approximation** (see e.g. 4.122):

$$\begin{aligned} \boldsymbol{E}_{\text{sc}} &= \frac{1}{4\pi\epsilon_0}k^2\frac{e^{ikr}}{r}\left\{(\hat{\boldsymbol{n}}\times\boldsymbol{p})\times\hat{\boldsymbol{n}} - \hat{\boldsymbol{n}}\times\boldsymbol{m}/c\right\} && (14.3)\\ \boldsymbol{H}_{\text{sc}} &= \hat{\boldsymbol{n}}\times\boldsymbol{E}_{\text{sc}}/Z_0. && (14.4)\end{aligned}$$

In these expressions, $\hat{\boldsymbol{n}}_0 = \frac{\boldsymbol{k}_0}{k_0}$ and $\hat{\boldsymbol{n}} = \frac{\boldsymbol{k}}{k}$, while $\hat{\boldsymbol{\epsilon}}_0, \hat{\boldsymbol{\epsilon}}$ are the polarization of the incident and scattered waves, respectively.

We are interested in the **relative power distribution** in the scattered field (which should be proportional to the incident field in a way that can be made independent of its magnitude in a linear response/susceptibility approximation). The power radiated in direction $\hat{\boldsymbol{n}}$ with polarization $\hat{\boldsymbol{\epsilon}}$ is needed per unit intensity in the incident wave with $\hat{\boldsymbol{n}}_0, \hat{\boldsymbol{\epsilon}}_0$. This quantity is expressed as

$$\frac{d\sigma}{d\Omega}(\hat{\boldsymbol{n}},\hat{\boldsymbol{\epsilon}},\hat{\boldsymbol{n}}_0,\hat{\boldsymbol{\epsilon}}_0) = r^2\frac{\frac{1}{2Z_0}\left|\hat{\boldsymbol{\epsilon}}^*\cdot\boldsymbol{E}_{\text{sc}}\right|^2}{\frac{1}{2Z_0}\left|\hat{\boldsymbol{\epsilon}}_0^*\cdot\boldsymbol{E}_{\text{inc}}\right|^2} \tag{14.5}$$

[One gets this by considering the power distribution:

$$\begin{aligned} \frac{dP}{d\Omega} &= \frac{1}{2}\text{Re}\left\{r^2\hat{\boldsymbol{n}}\cdot(\boldsymbol{E}\times\boldsymbol{H}^*)\right\}\\ &= \frac{1}{2Z_0}\left|\boldsymbol{E}\times(\hat{\boldsymbol{n}}\times\boldsymbol{E})\right|\\ &= \frac{1}{2Z_0}\left|\boldsymbol{E}\right|^2 && (14.6)\end{aligned}$$

as usual, where the latter relation steps hold for transverse EM fields 7.1 and 7.2 only and where we've projected out a single polarization from the incident and scattered waves so we can discuss polarization later.]

This quantity has the units of area (r^2) and is called the **differential cross–section**:

$$\frac{d\sigma}{d\Omega} = \frac{dP/d\Omega}{dP_0/dA} \propto \frac{dA}{d\Omega} \sim A. \tag{14.7}$$

In quantum theory a scattering cross–section one would substitute "intensity" (number of particles/second) for "power" in this definition but it still holds. Since the units of angles, solid or not, are dimensionless, a cross–section always has the units of area. If one integrates the cross–section around the 4π solid angle, the resulting area is the "effective" cross–sectional area of the scatterer, that is, the integrated are of its effective "shadow". This is the basis of the optical theorem, which I will mention but we will not study (derive) for lack of time.

The point in defining it is that it is generally a property of the scattering target that linearly determines the scattered power:

$$\frac{dP}{d\Omega} = \frac{d\sigma}{d\Omega} \times I_0 \tag{14.8}$$

where the last quantity is the intensity of the incident plane wave beam. The cross-section is independent (within reason) of the incident intensity and can be calculated or measured "once and for all" and then used to predict the power distribution for a given beam intensity.

We need to use the apparatus of chapter 7 to handle the vector polarization correctly. That is, technically we need to use the Stokes parameters or something similar to help us project out of $\mathbf{E}$ a particular polarization component. Then (as can easily be shown by meditating on:

$$\hat{\boldsymbol{\epsilon}}^* \cdot \boldsymbol{E}_{\rm sc} = \frac{1}{4\pi\epsilon_0} k^2 \frac{e^{ikr}}{r} \left\{ \hat{\boldsymbol{\epsilon}}^* \cdot \{ (\hat{\boldsymbol{n}} \times \boldsymbol{p}) \times \hat{\boldsymbol{n}} - \hat{\boldsymbol{n}} \times \boldsymbol{m}/c \} \right\} \tag{14.9}$$

for a transverse field):

$$\frac{d\sigma}{d\Omega} = r^2 \frac{\frac{1}{2Z_0} \left| \hat{\boldsymbol{\epsilon}}^* \cdot \boldsymbol{E}_{\rm sc} \right|^2}{\frac{1}{2Z_0} \left| \hat{\boldsymbol{\epsilon}}_0^* \cdot \boldsymbol{E}_{\rm inc} \right|^2} = \frac{k^4}{(4\pi\epsilon_0 E_0)^2} \left| \hat{\boldsymbol{\epsilon}}^* \cdot \boldsymbol{p} + (\hat{\boldsymbol{n}} \times \hat{\boldsymbol{\epsilon}}^*) \times \boldsymbol{m}/c \right|^2 . \tag{14.10}$$

To get this result, we had to evaluate (using vector identities)

$$\hat{\boldsymbol{\epsilon}}^* \cdot (\hat{\boldsymbol{n}} \times \boldsymbol{p}) \times \hat{\boldsymbol{n}} = \hat{\boldsymbol{\epsilon}}^* \cdot \boldsymbol{p} \tag{14.11}$$

and

$$\hat{\boldsymbol{\epsilon}}^* \cdot (\hat{\boldsymbol{n}} \times \boldsymbol{m}/c) = -\boldsymbol{m} \cdot (\hat{\boldsymbol{n}} \times \hat{\boldsymbol{\epsilon}}^*). \tag{14.12}$$

From this we immediately see one important result:

$$\frac{d\sigma}{d\Omega} \propto k^4 \propto \frac{1}{\lambda^4}. \tag{14.13}$$

This is called **Rayleigh's Law**; the scattering cross-section (and hence proportion of the power scattered from a given incident beam) by a polarizable medium is proportional to the *inverse fourth power of the wavelength*. Or, if you prefer, short wavelengths (still long with respect to the size of the scatterer and only if the dipole term in the scattering dominates) are scattered more strongly than long wavelengths. This is the original "blue sky" theory and probably the origin of the phrase!

To go further in our understanding, and to gain some useful practice against the day you have to use this theory or teach it to someone who might use it, we must consider some specific cases.

14.2 Scattering from a Small Dielectric Sphere

This is a relatively simple, and hence very standard problem.

Now, we have no desire to "reinvent the sphere"[1] but it is important that you understand where our results come from. First of all, let us introduce dimensionless, scaled versions of the *relative* permeability and permittivity (a step that Jackson apparently performs in J10 but does not document or

[1]Hyuk, hyuk, hyuk...

explain):

$$\begin{aligned} \epsilon_r &= \epsilon(\omega)/\epsilon_0 && (14.14) \\ \mu_r &= \mu(\omega)/\mu_0 \approx 1 && (14.15) \end{aligned}$$

where we assume that we are not at a resonance so that the spheres have normal dispersion and that these numbers are basically real. The latter is a good approximation for non-magnetic, non-conducting scatterers e.g. oxygen or nitrogen molecules.

If you refer back to J4.4, equation J4.56 and the surrounding text, you will see that the induced dipole moment in a dielectric sphere in terms of the relative permittivity is:

$$\boldsymbol{p} = 4\pi\epsilon_0 \left(\frac{\epsilon_r - 1}{\epsilon_r + 2}\right) a^3 \boldsymbol{E}_{\text{inc}} \tag{14.16}$$

To recapitulate the derivation (useful since this is a *common* question on qualifiers and the like) we note that the sphere has azimuthal symmetry around the direction of $\boldsymbol{E}$, so we can express the scalar potential inside and outside the sphere as

$$\begin{aligned} \phi_{\text{in}} &= \sum_\ell A_\ell r^\ell P_\ell(\cos\theta) && (14.17) \\ \phi_{\text{out}} &= \sum_\ell \left\{ B_\ell r^\ell + C_\ell \frac{1}{r^{\ell+1}} \right\} P_\ell(\cos\theta). && (14.18) \end{aligned}$$

We need to evaluate this. At infinity we know that the field should be (to lowest order) undisturbed, so the potential must asymptotically go over to

$$\lim_{r\to\infty} \phi_{\text{out}} = -E_0 z = -E_0 r\cos\theta = -E_0 r P_1(\cos\theta) \tag{14.19}$$

so we conclude that $B_1 = -E_0$ and all other $B_{\ell>1} = 0$. To proceed further, we must use the **matching conditions** of the tangential and normal fields at the surface of the sphere:

$$-\frac{1}{a}\left.\frac{\partial\phi_{\text{in}}}{\partial\theta}\right|_{r=a} = -\frac{1}{a}\left.\frac{\partial\phi_{\text{out}}}{\partial\theta}\right|_{r=a} \tag{14.20}$$

(tangential component) and

$$-\epsilon\left.\frac{\partial\phi_{\text{in}}}{\partial r}\right|_{r=a} = -\epsilon_0\left.\frac{\partial\phi_{\text{out}}}{\partial r}\right|_{r=a} \tag{14.21}$$

(normal $\boldsymbol{D}$ onto $\boldsymbol{E}$).

Since this is the surface of a sphere (!) we can project out each spherical component if we wish and cause these equations to be satisfied term by term. From the first (tangential) equation we just match ϕ itself:

$$\frac{1}{a}(A_\ell a^\ell) = \frac{1}{a}\left(B_\ell a^\ell + C_\ell \frac{1}{a^{\ell+1}}\right) \tag{14.22}$$

or (using our knowledge of B_ℓ)

$$\begin{aligned} A_1 &= -E_0 + \frac{C_1}{a^3} && \ell = 1 && (14.23) \\ A_\ell &= \frac{C_\ell}{a^{2\ell+1}} && \text{else} && (14.24) \end{aligned}$$

From the second (normal) equation we get

$$\begin{aligned} \epsilon_r A_1 &= -E_0 - 2\frac{C_1}{a^3} && \ell = 1 && (14.25) \\ \epsilon_r A_\ell &= -\frac{(\ell+1)C_\ell}{a^{2\ell+1}} && \text{else.} && (14.26) \end{aligned}$$

The second equation of each pair are incompatible and have only the trivial

$$A_\ell = C_\ell = 0 \qquad \ell \neq 1. \tag{14.27}$$

Only the $\ell = 1$ term survives. With a little work one can show that

$$\begin{aligned} A_1 &= -\frac{3E_0}{2+\epsilon_r} && (14.28) \\ C_1 &= \left(\frac{\epsilon_r - 1}{\epsilon_r + 2}\right) a^3 E_0 && (14.29) \end{aligned}$$

so that

$$\begin{aligned} \phi_{\text{in}} &= -\left(\frac{3}{\epsilon_r + 2}\right) E_0 r \cos\theta && (14.30) \\ \phi_{\text{out}} &= -E_0 r \cos\theta + \left(\frac{\epsilon_r - 1}{\epsilon_r + 2}\right) E_0 \frac{a^3}{r^2} \cos\theta. && (14.31) \end{aligned}$$

When we identify the second term of the external field with the dipole potential and compare with the expansion of the dipole potential

$$\phi(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0} \frac{\boldsymbol{p}\cdot\boldsymbol{r}}{r^3} \tag{14.32}$$

we conclude that the induced dipole moment is:

$$\boldsymbol{p} = 4\pi\epsilon_0 \left(\frac{\epsilon_r - 1}{\epsilon_r + 2}\right) a^3 E_0 \hat{\boldsymbol{z}}. \tag{14.33}$$

as given above.

There is no magnetic dipole moment, because $\mu_r = 1$ and therefore the sphere behaves like a "dipole antenna". Thus $\boldsymbol{m} = 0$ and there is no magnetic scattering of radiation from this system. This one equation, therefore, (together with our original definitions of the fields) is sufficient to determine the differential cross–section:

$$\frac{d\sigma}{d\Omega} = k^4 a^6 \left|\frac{\epsilon_r - 1}{\epsilon_r + 2}\right|^2 |\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 \tag{14.34}$$

where remember that $\epsilon_r(\omega)$ (for dispersion) and hopefully everybody notes the difference between dielectric ϵ and polarization $\hat{\boldsymbol{\epsilon}}$ (sigh – we need more symbols). This equation can be used to find the explicit differential cross–sections given $(\hat{\boldsymbol{n}}, \hat{\boldsymbol{n}}_0, \hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{\epsilon}}_0)$, as desired.

However, the light incident on the sphere will generally be unpolarized. Then the question naturally arises of whether the various independent polarizations of the incident light beam will be scattered identically. Or, to put it another way, what is the angular distribution function of radiation with a definite polarization? To answer this, we need to consider a suitable decomposition of the possible polarization directions.

This decomposition is apparent from considering the following picture of the general geometry:

Let $\hat{\boldsymbol{n}}, \hat{\boldsymbol{n}}_0$ define the plane of scattering. We have to fix $\hat{\boldsymbol{\epsilon}}^{(1)}$ and $\hat{\boldsymbol{\epsilon}}^{(2)}$ relative to this scattering plane and average over the polarizations in the incident light, $\hat{\boldsymbol{\epsilon}}_0^{(1)}$ and $\hat{\boldsymbol{\epsilon}}_0^{(2)}$ (also fixed relative to this plane). We can always choose the directions of polarization such that $\hat{\boldsymbol{\epsilon}}^{(2)} = \hat{\boldsymbol{\epsilon}}_0^{(2)}$ is perpendicular to the scattering plane and $\hat{\boldsymbol{\epsilon}}^{(1)} = \hat{\boldsymbol{\epsilon}}_0^{(1)}$ are in it, and perpendicular to the directions $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{n}}_0$ respectively. The dot products are thus

$$\begin{aligned} \hat{\boldsymbol{\epsilon}}^{(1)*} \cdot \hat{\boldsymbol{\epsilon}}_0^{(1)} &= \hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{n}}_0 = \cos\theta && (14.35) \\ \hat{\boldsymbol{\epsilon}}^{(2)*} \cdot \hat{\boldsymbol{\epsilon}}_0^{(2)} &= 1. && (14.36) \end{aligned}$$

We need the average of the squares of these quantities. This is essentially averaging $\sin^2\phi$ and $\cos^2\phi$ over $\phi \in (0, 2\pi)$. Alternatively, we can meditate upon symmetry and conclude that the average is just $\frac{1}{2}$. Thus (for the polarization in the plane ($\|$) and perpendicular to the plane ($\perp$) of scattering, respectively) we have:

$$\begin{aligned} \frac{d\sigma_\|}{d\Omega} &= k^4 a^6 \left|\frac{\epsilon_r - 1}{\epsilon_r + 2}\right|^2 \frac{\cos^2\theta}{2} && (14.37) \\ \frac{d\sigma_\perp}{d\Omega} &= k^4 a^6 \left|\frac{\epsilon_r - 1}{\epsilon_r + 2}\right|^2 \frac{1}{2} && (14.38) \end{aligned}$$

We see that light polarized perpendicular to the plane of scattering has no θ dependence, while light polarized in that plane is not scattered parallel to the direction of propagation at all (along $\theta = 0$ or π). We will invert this statement in a moment so that it makes more sense. See the diagram below.

Unfortunately, everything thus far is expressed with respect to the plane of scattering, which varies with the direction of the scattered light. If we define the **polarization** $\Pi(\theta)$ of the scattered radiation to be

$$\Pi(\theta) = \frac{\frac{d\sigma_\perp}{d\Omega} - \frac{d\sigma_\|}{d\Omega}}{\frac{d\sigma_\perp}{d\Omega} + \frac{\sigma_\|}{d\Omega}} = \frac{\sin^2\theta}{1 + \cos^2\theta} \tag{14.39}$$

then we obtain a quantity that is in accord with our intuition. $\Pi(\theta)$ is maximum at $\theta = \pi/2$. The radiation scattered through an angle of 90 degrees is completely polarized in a plane perpendicular to the plane of scattering.

Finally, we can add the two pieces of the differential cross–section together:

$$\frac{d\sigma}{d\Omega} = k^4 a^6 \left(\frac{\epsilon - 1}{\epsilon + 2}\right)^2 \frac{1}{2}(1 + \cos^2\theta) \tag{14.40}$$

which is strongly and symmetrically peaked forward and backward. Finally, this is easy to integrate to obtain the **total cross–section**:

$$\sigma = \frac{8\pi}{3} k^4 a^6 \left(\frac{\epsilon_r - 1}{\epsilon_r + 2}\right)^2 . \tag{14.41}$$

At last, we can put it all together. Molecules in the atmosphere behave, far from resonance, like itty–bitty dielectric spheres to a remarkable approximation. Since blue light is scattered more strongly than red, light seen away from its direction of incidence (the sky and not the sun) is shifted in color from white to blue. When Mr. Sun is examined directly through a thick layer of atmosphere (at sunset) the blue is all scattered out and the remaining light looks red. Finally, light from directly overhead at sunup or sundown is polarized in a north–south direction; at noon the light from the horizon is polarized parallel to the horizon (and hence is filtered by vertical transmission axis polarized sunglasses. You should verify this at your next opportunity outdoors with a pair of polarized sunglasses, as this *whole discussion* is taught in *elementary* terms in second semester introductory physics courses.

Don't say I never taught you anything[2].

The last remarks I would make concern the total cross–section. Note that if we factor out a $4\pi a^2$ we get the "area" of the sphere times a pure (dimensionless) number $(ka)^4$ associated with the relative size of the sphere radius and the wavelength and a second pure number involving only the dielectric properties of the medium:

$$\sigma = (4\pi a^2)(ka)^4 \left\{ \frac{2}{3} \left(\frac{\epsilon_r - 1}{\epsilon_r + 2}\right)^2 \right\} . \tag{14.42}$$

This expression isn't any more useful than the one above, but it does make the role of the different terms that contribute to the total scattering cross-section more clear.

[2]Even if it's true . . .

14.3 Scattering from a Small Conducting Sphere

Perfect conductors are not just dielectrics where the electric field is completely zero inside. The electric field *is* exactly cancelled on the interior by the induced surface charge. As we have seen, this cancellation occurs close to the surface (within a few times the skin depth). However, the induced currents *also* tend to expel the time dependent magnetic field. We therefore have two modification of our results from the previous section. The electric polarization will have a different form, and there will be a contribution from the induced *magnetic* moment of the sphere as well.

Recall (from J2.5) that the induced dipole moment on a conducting sphere is

$$\boldsymbol{p} = 4\pi\epsilon_0 a^3 \boldsymbol{E}_{\mathrm{inc}}. \tag{14.43}$$

This is indeed the generalization of the result for $\mathbf{p}$ last time, as you should be able to derive in a few minutes of work. Either review that section or solve the boundary value problem where $\boldsymbol{E}_\perp$ is discontinuous at the surface and $\boldsymbol{E}_{||} = 0$ on the surface to obtain:

$$\phi = -E_0 \left(r - \frac{a^3}{r^2} \right) \cos\theta \tag{14.44}$$

from which we can easily extract this $\boldsymbol{p}$.

But, the magnetic field is also varying, and it induces an EMF that runs in loops around the magnetic field lines and opposes the change in magnetic flux. Assuming that no field lines were trapped in the sphere initially, the induced currents act to cancel component of the magnetic field normal to the surface. The sphere thus behaves like a magnetically permeable sphere (see e.g. section J5.10 and J5.11, equations J5.106, J5.107, J5.115):

$$\boldsymbol{M} = \frac{\boldsymbol{m}}{4\pi a^3/3} = 3\left(\frac{\mu - \mu_0}{\mu + 2\mu_0}\right)\boldsymbol{H}_{\text{inc}} \tag{14.45}$$

with $\mu_r = \mu/\mu_0 = 0$ so that:

$$\boldsymbol{m} = -2\pi a^3 \boldsymbol{H}_{\text{inc}}. \tag{14.46}$$

The derivation is again very similar to the derivation we performed last time, with suitably chosen boundary conditions on $\boldsymbol{B}$ and $\boldsymbol{H}$.

If we then repeat the reasoning and algebra for this case of the conducting sphere (substituting *this* $\boldsymbol{p}$ and $\boldsymbol{m}$ into the expression we derived for the differential cross–section), we get:

$$\frac{d\sigma}{d\Omega} = k^4 a^6 \left|\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0 - \frac{1}{2}(\hat{\boldsymbol{n}} \times \hat{\boldsymbol{\epsilon}}^*) \cdot (\hat{\boldsymbol{n}}_0 \times \hat{\boldsymbol{\epsilon}}_0)\right|^2. \tag{14.47}$$

After much tedious but straightforward work, we can show (or rather you can show for homework) that:

$$\frac{d\sigma_\parallel}{d\Omega} = \frac{k^4 a^6}{2}\left|\cos\theta - \frac{1}{2}\right|^2 \tag{14.48}$$

$$\frac{d\sigma_\perp}{d\Omega} = \frac{k^4 a^6}{2}\left|1 - \frac{1}{2}\cos\theta\right|^2 \tag{14.49}$$

so that the total differential cross section is:

$$\frac{d\sigma}{d\Omega} = k^4 a^6 \left\{\frac{5}{8}(1 + \cos^2\theta) - \cos\theta\right\} \tag{14.50}$$

and the polarization is:

$$\Pi(\theta) = \frac{3\sin^2\theta}{5(1 + \cos^2\theta) - 8\cos\theta} \tag{14.51}$$

Finally, integrating the differential cross section yields the total cross-section:

$$\sigma = \frac{10\pi k^4 a^6}{3} = (4\pi a^2)(ka)^4 \frac{2.5}{3} \sim \sigma_{\text{dielectric}} \tag{14.52}$$

Figure 14.1: Differential cross–section and polarization of a small conducting sphere.

for $\epsilon_r >> 1$ curiously enough.

What do these equations tell us? The cross–section is strongly peaked backwards. Waves are reflected backwards more than forwards (the sphere actually casts a "shadow". The scattered radiation is polarized qualitatively alike the radiation scattered from the dielectric sphere, but with a somewhat different angular distribution. It is completely polarized perpendicular to the scattering plane when scattered through an angle of 60°, not 90°.

We see that dipole scattering will always have a characteristic k^4 dependence. By know you should readily understand me when I say that this is the result of performing a multipolar expansion of the reaction field (essentially an expansion in powers of kd where d is the characteristic maximum extent of the system) and keeping the first (dipole) term.

If one wishes to consider scattering from objects where $kd \sim 1$ or greater, one simply has to consider higher order multipoles (and one must consider the proper multipoles instead of simple expansions in powers of kd). If $kd >> 1$ (which is the case for light scattering from macroscopic objects, radar scattering from airplanes and incoming nuclear missiles, etc) then a whole different apparatus must be brought to bear. I could spend a semester (or a least a couple of weeks) just lecturing on the scattering of electromagnetic waves from spheres, let alone other shapes.

However, no useful purpose would be so served, so I won't. If you ever need to figure it out, you have the tools and can find and understand the necessary references.

14.4 Many Scatterers

It is, however, worthwhile to spend a moment considering a collections of identical scatterers at fixed spatial positions. Each scatterer then acts identically, but is scattering an electromagnetic field with its own (spatially dependent) phase at a given moment of time. The scattered fields then propagate freely, recombine, and form a total EM field that is measured by the detector. In order to evaluate the total differential cross–section we must sum the field amplitudes times the appropriate phases, project out the desired polarization moments, and *then* square.

A moment of quiet reflection[3] will convince you that in general:

$$\frac{d\sigma}{d\Omega} = \frac{k^4}{(4\pi\epsilon_0 E_0)^2} \left| \sum_j \left\{ \hat{\boldsymbol{\epsilon}}^* \cdot \boldsymbol{p}_j + (\hat{\boldsymbol{n}} \times \hat{\boldsymbol{\epsilon}}^*) \cdot \boldsymbol{m}_j/c \right\} e^{i\mathbf{q}\cdot\mathbf{x}_j} \right|^2 \tag{14.53}$$

where

$$\mathbf{q} = \mathbf{k}_0 - \mathbf{k}. \tag{14.54}$$

accomodates the *relative phase difference* between the field emitted by the scatterers at different locations. The geometry of this situation is pictured below.

In all directions but the forward direction, this depends on the distribution of scatterers and the nature of each scatterer. If we imagine all the scatterers to be alike (and assume that we are far from the collection) then this expression simplifies:

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma_0}{d\Omega}\mathcal{F}(\mathbf{q}) \tag{14.55}$$

[3] Sorry...

Figure 14.2: Geometry of multiple scatterers. The relative phase of two sources depends on the projection of the difference in wave vectors onto the vector connecting the scatterers.

where $\frac{d\sigma_0}{d\Omega}$ is the scattering cross-section of a single scatterer and the $\mathcal{F}(\mathbf{q})$ is called a "structure factor":

$$\mathcal{F}(\mathbf{q}) = \left| \sum_j e^{i\mathbf{q}\cdot\mathbf{x}_j} \right|^2 \tag{14.56}$$

$$= \sum_{i,j} e^{i\mathbf{q}\cdot(\mathbf{x}_j-\mathbf{x}_i)}. \tag{14.57}$$

This last expression is 1 on the diagonal $i = j$. If the (e.g.) atoms are uniformly but randomly distributed, the sum of the off-diagonal terms averages to zero and the total sum goes to N (the number of atoms). This is an incoherent superposition and the scattered intensitities add with negligible interference.

If the atoms are instead on a regular lattice, then "Bragg" scattering results. There will exist certain values of $\mathbf{q}$ that match the spacing between planes in such a way that whole rows of the matrix are 1. In those direction/wavelength combinations, the scattered intensity is of order N^2 and hence is much brighter. The scattered power distribution thus has bright spots in is corresponding to these directions, where constructive interference in the scattered waves occurs.

Structure factor sums occur in many branches of physics. If you think about it for a moment, you can easily see that it is possible to do a structure factor sum using the Green's function expansions you have studied. In electrodynamics and quantum multiple scattering theory these sums appear frequently in association with spatially fixed structures (like crystal lattices or molecules). In field theory, lattice sums are sometimes used as a discretized approximation for the continuum, and "lattice gauge" type field theories result. In these theories, our ability to do the structure factor sums is used to construct the Green's functions rather than the other way around. Either way, you should be familiar with the term and should think about the ways you might approach evaluating such a sum.

We are now done with our discussion of scattering from objects per se. It is well worth your while to *read J10.2 on your own.* I have given you the semi–quantitative argument for the blue sky; this section puts our simple treatment on firmer ground. It also derives the *perturbation theory of scattering* (using the Born approximation), and discusses a number of interesting current research topics (such as critical opalescence). I will probably assign one problem out of this section to help you out. However, perturbative scattering is easier to understand, and more useful, in the context of (scalar) quantum theory and so I will skip this section, expecting that you will see enough of it there.

You should also read J10.3. This presents *one* way to derive the Rayleigh expansion for a (scalar) plane wave in terms of free spherical waves (there are several). However, it goes further and addresses expansions of e.g. circularly polarized plane waves in terms of vector spherical harmonics! Lord knows why this is stuck off in this one section all by itself – I need to put the equivalent result for expansion in terms of Hansen solutions (which of course will be much more natural and will precompute most of the annoying parts of the algebra for us) in the sections on the Hansen functions and VSHs where it belongs, as it will actually be much simpler to understand there.

J10.4 redoes scattering from a sphere "right" in terms of VSHs, and again, if we wished to pursue this we would need to redo this in terms of Hansen functions to keep it simple. The primary advantage of reading this chapter is that it defines the *partial wave phase shifts* of scattering from a sphere, quantities that are in use in *precisely the same context* in quantum scattering theory in e.g. nuclear physics. SO, if you plan to go into nuclear physics you are well advised to read *this* chapter as well and work through it.

However, we cannot do this at this time because we had to go back and redo J7 and J8. Besides, we're doubtless a bit bored with multipoles and want to become excited again. We will therefore now move on to one of my favorite topics, *relativity theory*.

Part III

Relativistic Electrodynamics

Chapter 15

Special Relativity

15.1 Einstein's Postulates

By this time I certainly hope that you are familiar with the two postulates, due to Einstein, that lead to the theory of special relativity. They are:

1. The laws of nature are invariant with respect to the uniform translation of the coordinate system in which they are measured.
2. The speed of light is independent of the motion of the source.

Properly speaking, the second postulate is a consequence of the first, since if the speed of light depended on the motion of its source the laws of electrodynamics (which determine the speed of freely propagating electromagnetic waves) would depend on the inertial frame of the source, which contradicts the first postulate. For what it is worth, the first is not as obviously a consequence of the second: it seems entirely possible for *some* laws to depend on the velocity of the source and not contradict the second postulate, as long as they are not electrodynamical in nature. This has been the subject of considerable discussion, and I hesitate to state a religious view upon it.

I will, however, point out that in the opinion of Dirac, at least — the discovery of the uniform 3°K blackbody background explicitly contradicted the first postulate but not the second. You might amuse yourself, some quiet evening, by considering experiments that would measure your absolute velocity relative to the "rest" frame of this radiation. The second postulate (which is all we need) thus seems to be the safer of the two upon which to base our reasoning.

I strongly recommend that you read J11.1 — J11.2 on your own. They are "true facts" that will come in handy some day, and should astound and amaze you. Yes, Virginia, special relativity really really works.

For our purposes, we will begin with a *brief* review of the basic Lorentz transformation and an introduction to four vectors. Because we will do it again (correctly) in a week or so we won't take long now. We will also study four–velocity and four–momentum. This will suffice to give us the "flavor" of the theory and establish the geometricaly grounds for the matrix theory we will then derive.

As an application of this, we will study Thomas precession briefly and then go on to perform a detailed application of the theory to the dynamics of interacting charged particles and fields. We will spend the rest of the semester on this subject, in one form or another.

15.2 The Elementary Lorentz Transformation

To motivate the Lorentz transformation, recall the **Galilean transformation** between moving coordinate systems:

$$x_1' = x_1 - vt \tag{15.1}$$
$$x_2' = x_2 \tag{15.2}$$
$$x_3' = x_3 \tag{15.3}$$
$$t' = t \tag{15.4}$$

(where K is fixed and K' is moving in the 1–direction at speed v).

Then

$$F_j = m\ddot{x}_j = m\ddot{x}'_j = F'_j \tag{15.5}$$

or **Newton's Laws are covariant with respect to the Gallilean transformation.**

But

$$\frac{\partial}{\partial x_1} = \frac{\partial}{\partial x'_1} + \frac{1}{v}\frac{\partial}{\partial t'} \tag{15.6}$$

and so

$$\frac{\partial^2}{\partial x_1^2} = \frac{\partial^2}{\partial x_1'^2} + \frac{1}{v^2}\frac{\partial^2}{\partial t'^2} + \frac{2}{v}\frac{\partial^2}{\partial x'_1 \partial t'} \tag{15.7}$$

$$\frac{\partial^2}{\partial x_2^2} = \frac{\partial^2}{\partial x_2'^2} \tag{15.8}$$

$$\frac{\partial^2}{\partial x_3^2} = \frac{\partial^2}{\partial x_3'^2} \tag{15.9}$$

$$\frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial t'^2}. \tag{15.10}$$

Thus if

$$\left\{\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right\}\psi = 0 \tag{15.11}$$

then

$$\left\{\nabla'^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t'^2}\right\}\psi = \frac{-1}{v^2}\frac{\partial^2\psi}{\partial t'^2} - \frac{2}{v}\frac{\partial^2\psi}{\partial x'_1\partial t'} \neq 0 \tag{15.12}$$

(!) so that the wave equation, and hence Maxwell's equations which lead directly to the wave equation in free space, **are not covariant with respect to the Gallilean transformation!** They *already* determine the permitted velocity of a light wave, and do not allow that velocity to depend on anything but the properties of the medium through which the wave is transmitted.

The simplest linear transformation of coordinates is that preserves the form of the wave equation is easy to determine. It is one that keeps the speed of the (light) wave equal in both the K and the K' frames. Geometrically, if a flash of light is emitted from the (coincident) origins at time $t = t' = 0$, it will appear to expand like a sphere out from *both* coordinate origins, each in its own frame:

$$(ct)^2 - (x^2 + y^2 + z^2) = 0 \tag{15.13}$$

and

$$(ct')^2 - (x'^2 + y'^2 + z'^2) = 0 \tag{15.14}$$

are *simultaneous* constraints on the equations. Most generally,

$$(ct)^2 - (x^2 + y^2 + z^2) = \lambda^2 \left[(ct')^2 - (x'^2 + y'^2 + z'^2)\right] \tag{15.15}$$

where, $\lambda(v)$ describes a possible change of scale between the frames. If we insist that the coordinate transformation be homogeneous and symmetric between the frames[1], then

$$\lambda = 1. \tag{15.16}$$

Let us define

$$\begin{aligned} x_0 &= ct & (15.17)\\ x_1 &= x & (15.18)\\ x_2 &= y & (15.19)\\ x_3 &= z & (15.20)\\ (x_4 &= ict \text{ Minkowski metric}) & (15.21) \end{aligned}$$

Then we need a **linear transformation** of the coordinates that mixes $\mathbf{x}$ and (ct) in the direction of v in such a way that the length

$$\Delta s^2 = (x_0)^2 - (x_1^2 + x_2^2 + x_3^2) \tag{15.22}$$

[1]If we relax this requirement and allow for uniform expansions and/or contractions of the coordinate system, a more general group structure, the **conformal group**, results

is conserved *and* that goes into the Gallilean transformation as $v \to 0$. If we continue to assume that v is in the 1 direction, this leads to the **Lorentz transformation**:

$$\begin{aligned} x_0' &= \gamma(x_0 - \beta x_1) && (15.23)\\ x_1' &= \gamma(x_1 - \beta x_0) && (15.24)\\ x_2' &= x_2 && (15.25)\\ x_3' &= x_3 && (15.26)\end{aligned}$$

where at $x_1' = 0$,

$$x_1 = vt \to \beta = \frac{v}{c}. \tag{15.27}$$

Then

$$\Delta s^2 = \Delta s'^2 \tag{15.28}$$

leads to

$$x_0^2 - x_1^2 = \gamma^2(x_0^2 - x_1^2) + \gamma^2\beta^2(x_1^2 - x_0^2) \tag{15.29}$$

or

$$\gamma^2(1 - \beta^2) = 1 \tag{15.30}$$

so

$$\gamma = \frac{\pm 1}{\sqrt{1 - \beta^2}} \tag{15.31}$$

where we choose the + sign **by convention**. This makes $\gamma(0) = +1$. Finally,

$$\gamma(v) = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{15.32}$$

as we all know and love.

Now, let me remind you that when $v << c$,

$$\gamma(v) = 1 + \frac{1}{2}\frac{v^2}{c^2} + \ldots \tag{15.33}$$

to lowest surviving order in $\frac{v}{c}$. As we shall see, this is why "kinetic energy" in non–relativistic systems (being defined as the total energy minus the potential energy and the rest mass energy) is the usual $\frac{1}{2}mv^2$.

The inverse transformation (from K' to K) is also of some interest.

$$\begin{aligned} x_0 &= \gamma(x_0' + \beta x_1') && (15.34)\\ x_1 &= \gamma(x_1' + \beta x_0') && (15.35)\\ x_2 &= x_2' && (15.36)\\ x_3 &= x_3' && (15.37)\end{aligned}$$

which is perfectly symmetric, with $v \to -v$. It appears that which frame is at "rest" and which is moving is mathematically, at least, a matter of perspective.

Finally, if we let

$$\vec{\beta} = \frac{\mathbf{v}}{c} \tag{15.38}$$

(in an arbitrary direction) then we have but to use dot products to align the vector transformation equations with this direction:

$$x_0' = \gamma(x_0 - \vec{\beta} \cdot \mathbf{x}) \tag{15.39}$$

$$\mathbf{x}' = \mathbf{x} + \frac{\gamma - 1}{\beta^2}(\vec{\beta} \cdot \mathbf{x})\vec{\beta} - \gamma\vec{\beta}x_0 \tag{15.40}$$

I think that you should prove that this is correct as an exercise. Since the direction of $\vec{\beta}$ is arbitrary, it should suffice to show that this reduces the the form above for that direction and an arbitrary transverse direction.

Solution: Note that

$$\mathbf{x}_{\|} = \frac{(\vec{\beta} \cdot \mathbf{x})\vec{\beta}}{\beta^2} \tag{15.41}$$

Lorentz transform it according to this rule and one gets (by inspection)

$$\mathbf{x}_{\|}' = \gamma(\mathbf{x}_{\|} - \vec{\beta}x_0) \tag{15.42}$$

as one should. The x_0' transform is obvious. Finally, the other two ($\perp$) components do not get a contribution from γ. That is,

$$\mathbf{x}' = (\mathbf{x}_{\perp} + \mathbf{x}_{\|}) + \gamma\mathbf{x}_{\|} - \mathbf{x}_{\|} - \gamma\vec{\beta}x_0 \tag{15.43}$$

(reconstructing the result above directly) QED.

This is not the most general or convenient way to write the final transform. This is because γ and β are both related functions; it should not be necessary to use two parameters that are not independent. Also, the limiting structure of the transformation is not at all apparent without considering the functional forms in detail.

It is easy to see from the definition above that

$$(\gamma^2 - \gamma^2\beta^2) = 1. \tag{15.44}$$

The range of β is determined by the requirement that the transformation be non–singular and its symmetry:

$$0 \leq \beta < 1 \quad \text{so that} \quad 1 \leq \gamma < \infty. \tag{15.45}$$

If we think about functions that "naturally" parameterize these ranges, they are:

$$\cosh^2 \xi - \sinh^2 \xi = 1 \tag{15.46}$$

where

$$\beta = \tanh \xi = \frac{e^{-\xi} - e^{\xi}}{e^{-\xi} + e^{\xi}} \in [0, 1) \tag{15.47}$$

$$\gamma = \cosh \xi = \frac{1}{2}(e^{-\xi} + e^{\xi}) \in [1, \infty) \tag{15.48}$$

$$\gamma\beta = \sinh \xi = \frac{1}{2}(e^{-\xi} - e^{\xi}) \in [0, \infty). \tag{15.49}$$

The parameter ξ is called the **boost parameter** or **rapidity**. You will see this used frequently in the description of relativistic problems. You will also hear about "boosting" between frames, which essentially means performing a Lorentz transformation (a "boost") to the new frame. This will become clearer later when we generalize these results further. To give you something to meditate upon, consider in your minds the formal similarity between a "boost" and a "rotation" between the $\mathbf{x}_{\parallel}$ and x_0 coordinates where the rotation is through an imaginary angle $i\xi$. Hmmmm.

To elucidate this remark a wee tad more, note that in this parameterization,

$$x'_0 = x_0 \cosh \xi - x_{\parallel} \sinh \xi \tag{15.50}$$

$$x'_{\parallel} = -x_0 \sinh \xi + x_{\parallel} \cosh \xi \tag{15.51}$$

$$\mathbf{x}'_{\perp} = \mathbf{x}_{\perp}. \tag{15.52}$$

What is the 4×4 transformation matrix (in four dimensions) for this result? Does it look like a "hyperbolic rotation"[2] or what?

We have just determined that the (vector) coordinate system transforms a certain way. What, then, of vector fields, or any other vector quantity?

[2] "Hyperbolic" because of the relative minus sign between x^2 and ct^2. More on this later.

How do general vectors transform under a boost? This depends on the nature of the vectors. Many vectors, if not most, transform like the underlying coordinate description of the vectors. This includes the ones of greatest interest in physics. To make this obvious, we will have to generalize a vector quantity to **four dimensions**.

15.3 4-Vectors

Note well that we are not YET introducing proper notation for co- and contravariant tensors as we don't know what that means. Actually the notation for ordinary coordinates should be x^0, x^1, x^2, x^3 and we will have to determine whether any given 4-vector quantity that is a function of the coordinates transforms like the coordinate or like a *differential* of the coordinate in order to determine if it is co- or contravariant. Similarly, we have not yet discussed how to form the various dyadic products of co- and contravariant vectors – some will form scalars, some vectors, some second rank tensors. In other words, the results below are all correct, but the notation sucks and this suckiness will make certain parts of doing the algebra more difficult than it needs to be.

I may rewrite this whole combined multichapter stretch, as I'm not certain of the pegagogical value of presenting things incorrectly or in an elementary form and then correctly in an elegant form as Jackson does. In the meantime, please bear with the notation below allowing for the fact that much of it is just wrong.

Coordinate 4–vectors are (x_0, x_1, x_2, x_3).

Arbitrary 4–vectors are (A_0, A_1, A_2, A_3).

If the "arbitrary" vector transforms like the coordinates, then

$$A'_0 = \gamma(A_0 - \vec{\beta} \cdot \mathbf{A}) \tag{15.53}$$

$$A'_\parallel = \gamma(A_\parallel - \beta A_0) \tag{15.54}$$

$$\mathbf{A}'_\perp = \mathbf{A}_\perp \tag{15.55}$$

and

$$\begin{aligned} \Delta A^2 &= A_0^2 - (A_1^2 + A_2^2 + A_3^2) \\ &= A_0^2 - \mathbf{A} \cdot \mathbf{A} \end{aligned} \tag{15.56}$$

Figure 15.1: The Light Cone: Past, now, future, and elsewhere. Events. The world line.

is an **invariant** of the transformation. Note: whenever I boldface a vector quantity, I mean the 3D euclidean (cartesian) vector in ordinary space. In that case I will write the time (0) component explicitly. When I want to refer to a 4–vector generically, I will not boldface it (e. g. — $\mathbf{A}$ vs A).

Kids! Amaze your friends! Astound your neighbors! Show that

$$A'_0 B'_0 - \mathbf{A}' \cdot \mathbf{B}' = A_0 B_0 - \mathbf{A} \cdot \mathbf{B} \tag{15.57}$$

is an invariant of the Lorentz transformation for arbitrary 4–vectors A, B. This is (or will be) homework.

Now, we have a few definitions of "new words" to learn. Most of you probably already know them from one context or another, but we all need to agree at this point to talk a common language, so we will review the definitions carefully and avoid confusion.

Electromagnetic signals (and anything else travelling at speed c) travel on the **light cone**. An **event** is a coordinate $x = (x_0, \mathbf{x})$. We are usually interested in **causally connected** events on a **world line**. This might be, for example, the trajectory of a massive particle (like one on the tip of your nose) with $v < c$. Causally connected world line trajectories must live *inside* the light cone of each event that lies upon them.

Consider two events. If we define the **invariant interval**

$$S_{12}^2 = c^2(t_1 - t_2)^2 - |\mathbf{x}_1 - \mathbf{x}_2|^2 \tag{15.58}$$

then we have a

timelike separation $S_{12}^2 > 0 \quad \Rightarrow c^2(t_1 - t_2)^2 > |\mathbf{x}_1 - \mathbf{x}_2|^2$.

Both events are **inside** each other's light cone. These events can be "causally connected", because a light signal given off by one can reach the other from the "inside". In this case, a suitable Lorentz transformation can make $\mathbf{x}_1' = \mathbf{x}_2'$, but $t_1' \neq t_2'$ always.

spacelike separation $S_{12}^2 < 0 \quad \Rightarrow c^2(t_1 - t_2)^2 < |\mathbf{x}_1 - \mathbf{x}_2|^2$.

Both events are **outside** each other's light cone. These events are "causally disconnected", because a light signal given off by one can not reach the other. If nothing goes faster than light, then those particular events did not speak to one another. Note that this does *not* mean that *earlier* (and later) events on each world line to not connect. The events are disconnected, not the world lines themselves.

In this case, a suitable Lorentz transformation can make $t'_1 = t'_2$, but $\mathbf{x}'_1 \neq \mathbf{x}'_2$ always.

lightlike separation $S_{12}^2 = 0 \quad \Rightarrow c^2(t_1 - t_2)^2 = |\mathbf{x}_1 - \mathbf{x}_2|^2.$

Both events are *on* each other's light cone. These events are "causally connected" by electromagnetic radiation. The field produced by charges at one event are directly interacting with charges at the other event, and vice versa.

Note well that the event pairs considered above can be made spatially coincident, temporally coincident, or both, by suitably boosting the frame. Events with a timelike separation can be made spatially coincident. Events with a spacelike separation can be made to occur at the same time, or in either order. Events with a lightlike separation will always have a lightlike separation in all frames.

We are about to run into a profound philosophical difficulty. Physics is dedicated to dynamics – typically solving initial value problems and hence predicting the dynamical evolution of systems in time. Unfortunately, we just *eliminated time as an independent variable.* By making it a part of our geometry, it is no longer available as an *independent* parameter that we can use to write traditional equations of motion.

There are likely to other significant consequences of this decision, as many of the quantities studied in physics are tensor forms *defined with respect to spatial geometry.* That is, when I compute "charge" or "momentum" or "electric field" or a "rotation matrix", I'm computing 0th, 1st or 2nd rank tensors that inherit their directional character (or lack of it) from the underlying spatial coordinate system. Well, we've just made that underlying coordinate system *four* dimensional and so quantities like "momentum" and "electric field" will have to be reconsidered. We may need to find new "timelike" coordinates to associate with some of these, and perhaps reclassify others as different sorts of tensors.

Finally, we need to recover a "time" that can be used to write down *some* sort of equations of motion or we can't make a "physics". This will prove to be very difficult. For one thing, we can no longer expect to be able to solve initial value problems, as *time is now a symmetric coordinate*. The trajectories of particles are determined by their relativistic interaction connections and differential "equations of motion" with boundary conditions on a closed four dimensional hypersurface at four–infinity! That means that it is impossible *in principle* to predict future trajectories from only a knowledge of those trajectories in the past. It is amazing how few people in physics are willing to internally acknowledge that fact. Accept it. It is true. You will be happier for it.

Anyway, there are at least two ways around this (mathematical) difficulty. One is to introduce a "hypertime" – yet another dimension containing a parameter that can serve us as time has served in the past[3]. This, however, introduces a fifth dimension which we need (currently) like a fifth wheel. Maybe God lives in hypertime, but there are infinite difficulties associated with our trying to implement it in the complete absence of physical probes. Say hello to Plane Joe from Flatland. Leave it to masochistic theorists to play games with 10, 26, or even 4096 dimensional projective manifolds at least until you are ready to become one of them.

The second way is to introduce the **proper time**. This is the time **measured in the "rest frame" of a particle as it moves along its world line**. As such, it is still not an "absolute" time like we are used to but it is the closest that we can come to it.

Note well that proper time does not really solve our *philosophical* problems, because one must still ask how the "particle" measures time. If it carries with it a little "clock", that clock must have moving parts and some sort of associated period, and those parts have in turn their own proper time. If it is a *point* particle, its clock must either be in *internal* degrees of freedom – you begin to see why those theorists mentioned above work their

[3]Don't think too hard about this sentence or you'll start to go slightly nuts because it is self-referential and hence Gödelian.

way up to higher dimensional spaces – or else the particle infers the passage of time from what it "sees" of the rest of the Universe via its interaction connections and doesn't really have a proper time at all because it cannot have its own proper clock.

It does, however, solve our immediate mathematical problem (that of finding a suitable parameter in terms of which to describe the evolution of a system) so we'll go with it anyway.

15.4 Proper Time and Time Dilation

Suppose we have a particle moving with a velocity $\mathbf{v}$ in a given coordinate system K. In a time dt (in that system) it moves $d\mathbf{x} = \mathbf{v}dt$. Then its **invariant infinitesimal interval** is

$$(ds)^2 = (cdt)^2 - |d\mathbf{x}|^2 = c^2 dt^2(1 - \beta^2). \tag{15.59}$$

In the *particular* frame where the particle is at *rest* ($dx' = 0$) we define the **proper time** to be

$$d\tau = dt' \tag{15.60}$$

so that

$$(ds)^2 = c^2(d\tau)^2. \tag{15.61}$$

Thus the proper time is just the time experienced by the particle in its own rest frame.

From the relations above, it is easy to see that

$$d\tau = dt\sqrt{1 - \beta^2(t)} = \frac{dt}{\gamma(t)} \tag{15.62}$$

and to find the interval between two events on some world line it is necessary to integrate:

$$\begin{aligned} t_2 - t_1 &= \int_{\tau_1}^{\tau_2} \frac{d\tau}{\sqrt{1 - \beta^2(\tau)}} \\ &= \int_{\tau_1}^{\tau_2} \gamma(\tau) d\tau. \end{aligned} \tag{15.63}$$

If β is constant (so the frames are inertial) then we get the usual **time dilation**

$$\Delta t = \gamma \Delta\tau \tag{15.64}$$

or

$$\Delta\tau = \frac{\Delta t}{\gamma} \tag{15.65}$$

but this is not true if the particle is accelerating. Applying it without thought leads to the "twin paradox". However, the full integral relations will be valid even if the two particles are accelerating (so that $\beta(\tau)$). You will need to evaluate these relations to solve the twin paradox for one of your homework problems.

Finally, I want to note (without discussing it further at this time) that proper time dilation leads to a relativistic correction to the usual doppler shift. Or should I say that the non–relativistic doppler shift is just a low velocity limit of the correct, time dilated result.

Now that we have some notion of what an infinitesimal time interval is, we could go ahead and try to defince 4–dimensional generalizations of momentum and energy. First, however, we will learn how *velocities* Lorentz transform.

15.5 Addition of Velocities

If we form the infinitesimal version of the Lorentz transformation of coordinates:

$$\begin{aligned} dx_0 &= \gamma(dx_0' + \beta dx_1') && (15.66)\\ dx_1 &= \gamma(dx_1' + \beta dx_0') && (15.67)\\ dx_2 &= dx_2' && (15.68)\\ dx_3 &= dx_3' && (15.69) \end{aligned}$$

Point P is moving at velocity $\mathbf{u}'$ in frame K', which is in turn moving at velocity $\mathbf{v} = v\hat{1}$ with respect to the "rest" frame K. We need to determine u (the velocity of P in K). We will express the problem, as usual, in coordinates $\parallel$ and $\perp$ to the direction of motion, exploiting the obvious azimuthal symmetry of the transformation about the $\hat{1}$ direction.

Note that

$$u_i = c\frac{dx_i}{dx_0} \tag{15.70}$$

Figure 15.2: P has $\mathbf{u}' = (u', \theta', \phi')$ in K' frame. K' is moving in the 1 direction at $v = c\beta$. $\gamma(v)$ changes frames. We want $\mathbf{u}(u, \theta, \phi)$.

for $i = 0 \ldots 3$. Then

$$
\begin{aligned}
u_{\|} &= c\frac{\gamma(dx_1' + \beta dx_0')}{\gamma(dx_0' + \beta dx_1')} \\
&= c\frac{\left\{\frac{dx_1'}{dx_0'} + \beta\right\}}{\left\{1 + \beta\frac{dx_1'}{dx_0'}\right\}} \\
&= \frac{u_{\|} + v}{1 + \frac{\mathbf{u}'\cdot\mathbf{v}}{c^2}}.
\end{aligned} \tag{15.71}
$$

Similarly, $u_\perp$ (e.g. — u_2) is given by

$$
\begin{aligned}
u_2 &= \frac{cdx_2'}{\gamma(dx_0' + \beta dx_1')} \\
&= \frac{u_2'}{\gamma(1 + \frac{\mathbf{u}'\cdot\mathbf{v}}{c^2}}
\end{aligned} \tag{15.72}
$$

or

$$
\mathbf{u}_\perp = \frac{\mathbf{u}_\perp}{\gamma\left\{1 + \frac{\mathbf{u}'\cdot\mathbf{v}}{c^2}\right\}}. \tag{15.73}
$$

We see, then, that the velocity changes in *both* the $\|$ *and* the $\perp$ directions.

Note also that if $|\mathbf{u}'|$ and $|\mathbf{v}| << c$, then

$$
\frac{\mathbf{u}'\cdot\mathbf{v}}{c^2} << 1 \tag{15.74}
$$

and

$$\gamma \approx 1 \tag{15.75}$$

so that we recover the Gallilean result,

$$\begin{aligned} \mathbf{u}_{\parallel} &= \mathbf{u}'_{\parallel} + \mathbf{v} && (15.76) \\ \mathbf{u}_{\perp} &= \mathbf{u}'_{\perp}. && (15.77) \end{aligned}$$

What about the other limit? If $|\mathbf{u}'| = c$, then

$$|\mathbf{u}| = c \tag{15.78}$$

as you should verify on your own. *This is Einstein's second postulate!* We have thus proven explicitly that the speed of light (and the speed of anything else travelling at the speed of light) is *invariant* under Lorentz coordinate transformations. This is their entire motivation.

We observe that the three spatial components of "velocity" do *not* seem to transform like a four vector. Both the $\parallel$ and the $\perp$ components are mixed by a boost. We can, however, make the velocity into a four vector that does. We *define*

$$\begin{aligned} U_0 &= \frac{dx_0}{d\tau} = \frac{dx_0}{dt}\frac{dt}{d\tau} \\ &= c\gamma(u) && (15.79) \\ \mathbf{U} &= \frac{d\mathbf{x}}{d\tau} = \frac{d\mathbf{x}}{dt}\frac{dt}{d\tau} \\ &= \mathbf{u}\gamma(u) && (15.80) \end{aligned}$$

where $\gamma(u)$ is evaluated using the magnitude of $\mathbf{u}$. It is an exercise to show that this transforms like the coordinate 4–vector x.

Now we can "guess" that the 4–momentum of a particle will be $\sim mU$. To prepare us for this, observe that

$$U = (U_0, \mathbf{U}) = (\gamma_u c, \gamma_u \mathbf{u}) \tag{15.81}$$

are just the γ_u–scaled "velocities" of the particle:

Figure 15.3: Note that $\gamma_u \geq 1$ so that each component of the 4–velocity is always "larger" than associated Cartesian components, even though (as usual) the length of the four velocity is invariant. What is its invariant length?

15.6 Relativistic Energy and Momentum

We seek a relativistic generalization of momentum (a vector quantity) and energy. We know that in the low speed limit, $v << c$,

$$\mathbf{p} = m\mathbf{u} \tag{15.82}$$

$$E = E(0) + \frac{1}{2}mu^2 \tag{15.83}$$

where $E(0)$ is a constant allowed by Newton's laws (since forces depend only on energy *differences*).

The only possible form for this generalization of these equations consistent with our requirement that the laws of nature remain invariant are:

$$\mathbf{p} = \mathcal{M}(u)\mathbf{u} \tag{15.84}$$

$$E = \mathcal{E}(u), \tag{15.85}$$

that is, the mass and the energy must become functions of the speed only, and leave the vector character of the velocity alone. A boost cannot change the direction of the momentum of a particle, and any (scalar) functional variation in its magnitude can be thrown into the "mass" term.

This immediately yields the limiting forms:

$$\mathcal{M}(0) = m \tag{15.86}$$

$$\frac{\partial \mathcal{E}}{\partial u^2}(0) = \frac{m}{2} \tag{15.87}$$

where we have assumed that there is no pathology in the functions at the origin.

There are several possible ways to evaluate the full forms of these functions. Jackson's (based on scattering theory) is tedious and conceals the structure of the result. Furthermore, after telling us that selecting clever initial directions with an eye to simplifying the algebra "lacks motivation" he derives a result by selecting particular initial directions. The guy loves algebra, what can I say. Feel free to study his approach. It works.

I, on the other hand, am too lazy to spend most of a period deriving a result that is "obvious" in the correct notation. I am therefore going to "give" you the result and motivate it, and then verify it trivially be expressing it as a four–vector. This works nearly as well and is not anywhere near as painful.

We begin by considering elastic scattering theory. An elastic collision of two identical particles must conserve momentum and energy in all inertial frames. In the center of mass frame (which we will consider to be K')

$$\mathbf{p}'_{ia} + \mathbf{p}'_{ib} = \mathbf{p}'_{fa} + \mathbf{p}'_{fb} \tag{15.88}$$

$$E'_{ia} + E'_{ib} = E'_{fa} + E'_{fb} \tag{15.89}$$

relate the intial and final momenta and energy of the two identical particles.

Now,

$$\mathbf{u}'_{ia} = \mathbf{v} = -\mathbf{u}'_{ib} \tag{15.90}$$

and

$$\mathbf{u}_{fa} = \mathbf{v}' = \mathbf{u}'_{fb} \tag{15.91}$$

by definition in the center of mass system.

A moments quiet reflection (egad, another pun!) should convince you that in terms of the general transformation:

$$\mathcal{M}(v)\mathbf{v} - \mathcal{M}(v)\mathbf{v} = \mathcal{M}(v')\mathbf{v}' - \mathcal{M}(v')\mathbf{v}' \tag{15.92}$$

Figure 15.4: $\theta' = 30^\circ$ and $\beta^2 = \frac{1}{3}$. The dashed lines are the results of a Gallilean transformation from K' to K. Note that the scattering is more forward than expected because of the Lorentz contraction of the longitudinal distances seen by the particles.

$$\mathcal{E}(v) + \mathcal{E}(v) = \mathcal{E}(v') + \mathcal{E}(v'). \tag{15.93}$$

For what it is worth, if the collision is elastic and the particles are identical before and after the collision, $v = v'$ and all the mass terms are the same. We will denote the scattering angle in K' as θ'.

We thus begin with

$$\mathcal{M}(v)\mathbf{v} - \mathcal{M}(v)\mathbf{v} = \mathcal{M}(v)\mathbf{v} - \mathcal{M}(v)\mathbf{v} \tag{15.94}$$

$$\mathcal{E}(v) + \mathcal{E}(v) = \mathcal{E}(v) + \mathcal{E}(v) \tag{15.95}$$

where v is the speed of the incoming and outgoing particles. Now, $\mathcal{M}(v)$ must be a scalar function of v, and in the limit $v \to 0$ must turn into

$$\lim_{v \to 0} \mathcal{M}(v) = m. \tag{15.96}$$

The only scalar function of v we have encountered so far with this behavior is $\gamma(v)$, so we should sensibly guess

$$\mathcal{M}(v) = \gamma(v)m \tag{15.97}$$

which has the exactly correct limiting behavior.

Thus

$$\mathbf{p} = \gamma m \mathbf{u} \tag{15.98}$$

is a reasonable guess to be the generalization of momentum we seek. It is easy to verify that this is a consistent choice, and that it indeed results in conservation of momentum in all inertial frames.

To get the energy equation, we use the same approach. Recall that a binomial expansion of γ is given by

$$\lim_{v\to 0} \gamma(v) = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} = 1 + \frac{1}{2}\frac{v^2}{c^2} + \ldots \tag{15.99}$$

We need to keep the first non–constant term because we recall that physics is always "independent" of absolute energy scale. Then it should be clear that

$$\lim_{v\to 0} \mathcal{E}(v) = \gamma(v)\mathcal{E}(0) \approx \mathcal{E}(0) + \frac{1}{2}\mathcal{E}(0)\frac{v^2}{c^2} \approx \mathcal{E}(0) + \frac{1}{2}mv^2 \tag{15.100}$$

as it must in order to yield the low velocity limit of kinetic energy if and only if

$$\mathcal{E}(0) = mc^2. \tag{15.101}$$

There are several questions to be answered at this point, some experimentally and some theoretically. We need to measure the rest masses and theoretically verify that only this transformation correctly preserves the energy momentum conservation laws in elastic collisions as required. Beyond that, there are still some uncertainties. For example, there could in principal be an additional constant energy added to the energy term that was not scaled by γ and the laws of physics would still be expressible, since they are not sensitive to absolute energy scale. We will take advantage of that freedom in several instances to add or subtract an infinite *theoretical* constant in order to make the rest mass come out to the observed *experimental* mass m. This is called renormalization.

To obtain the same result a different way, we turn to the notation of 4–vectors. We observe that the common factor of γ above in both E and $\mathbf{p}$ also occurs when one makes velocity into a four vector. This suggests that energy and momentum can similarly be made into four vectors that transform like the coordinates under a boost. If we try the combination

$$
\begin{aligned}
p_0 &= mcU_0 = \frac{E}{c} && (15.102)\\
\mathbf{p} &= m\mathbf{U} && (15.103)
\end{aligned}
$$

we see that it works exactly. It results in an invariant

$$p_0^2 - \mathbf{p} \cdot \mathbf{p} = (m_c)^2. \tag{15.104}$$

It is easy to see the value of the invariant when $v = 0$; you should verify explicitly that it holds when $v \neq 0$ as well. Practically speaking, it suffices to show that this length is invariant when one wishes to show that its components transform like the coordinates under the action of a boost (why is that?).

The total energy can thus be expressed in terms of the three momentum as

$$E = \sqrt{c^2p^2 + m^2c^4}. \tag{15.105}$$

Finally, it is sometimes convenient to be able to get the velocity of the particle in terms of its energy and momentum

$$\mathbf{u} = \frac{c^2\mathbf{p}}{E} \tag{15.106}$$

which follows directly from the definitions.

This completes our review of "elementary relativity theory". We shall now proceed to develop the theory in a new, *geometric* language which is suitable to our much more sophisticated needs. To do this, we will need to begin by generalizing the notion of a four dimensional vector space with a set of transformations that leave an appropriately defined "length" invariant.

Chapter 16

The Lorentz Group

16.1 The Geometry of Space–Time

Recall that a great deal of simplification of the kinematics of classical non–relativistic mechanics occurs when one considers the *group structure* of transformations with respect to the underlying coordinates. Specifically, the group of **inversions, translations** and **rotations** of a given coordinate system leave the **norm** (length) of a given vector invariant. These transformations form the Euclidean group in three dimensions, E_3.

For those of you who led deprived childhoods, a **group** $\mathcal{G}$ is a set of mathematical objects $(a, b, c \ldots)$ with a rule of composition, or group product, $(a \circ b)$ such that:

1. Every product of a pair of elements in the group is also in the group. That is, if $a, b \in \mathcal{G}$ then $c = a \circ b \in \mathcal{G}$). This property is called **closure**.

2. The group must contain a special element called the **identity** $I \in \mathcal{G}$ such that $a \circ I = a$ for all $a \in \mathcal{G}$.

3. Every element of the group $\mathcal{G}$ must have an **inverse**, also in $\mathcal{G}$. If $a \in \mathcal{G}$ then $\exists a^{-1} \in \mathcal{G}$ such that $a \circ a^{-1} = I$.

4. The group product must be **associative**. That is, $a \circ (b \circ c) = (a \circ b) \circ c, \forall a, b, c \in \mathcal{G}$.

If the group product **commutes** ($a \circ b = b \circ a$) the group is said to be **Abelian**[1] otherwise the group is said to be **non–Abelian**, which is sensible enough. A **Lie** group is a **continuous** group[2] such as the group of infinitesimal transformations. It necessarily has an uncountable infinity of elements. There are also discrete (but countably infinite) groups, finite groups, and everything in between. There are also "semi–groups" (which do not, for example, contain an inverse). Finally, one can construct "non–associative" structures like groups from non–associative algebras like the octonions. Multiplication over the reals forms a continuous Abelian group. Rotations form a non–Abelian Lie group. Multiplication over rational numbers forms a countably infinite group. The set of rotations and inversions that leave a square invariant form a finite (point) group. The "renormalization group" you will hear much about over the years is not a group but a semi–group — it lacks an inverse.

However, our purpose here is not, however, to study group theory *per se*. One could study group theory for four years straight and still only scratch the surface. It is somewhat surprising that, given the importance of group theory in physics, we don't offer a single course in it, but then again, it's not *that* surprising...

With that in mind, we can decide what we are looking for. We seek initially the set of transformations in four dimensions that will leave

$$s^2 = x_0^2 - (\mathbf{x} \cdot \mathbf{x}) \tag{16.1}$$

invariant for a single event x with respect to a particular coordinate origin. These transformations form a group called the **homogeneous Lorentz group**. It consists of ordinary rotations in the spatial part, the Lorentz transformations we have just learned that mix space and time, and several discrete transformations such as space inversion(s) and time inversion.

The set of transformations that leave the quantity

$$s^2(x, y) = (x_0 - y_0)^2 - \left\{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2)\right\} \tag{16.2}$$

invariant form the **inhomogeneous Lorentz**[3] or **Poincaré group**. It consists of the homogeneous group (including the "improper" transformations

[1]Wikipedia: http://www.wikipedia.org/wiki/Abelian group. ;
[2]Wikipedia: http://www.wikipedia.org/wiki/Lie group. ,
[3]Wikipedia: http://www.wikipedia.org/wiki/Lorentz group. ,

that include spatial reflection and time reversal) and *uniform translations of the origin*. If anyone cares, the Lorentz group is the generalized orthogonal group O(1,3). The proper subgroup of the Lorentz group (the one that is simply connected spatially (no odd inversions) and contains the identity) is SO(1,3) the special orthogonal group. If time's direction is also preserved we add a +, $\mathrm{SO}^+(1,3)$. This nomenclature is defined here for your convenience but of course the wikinote reference contains active links to a lot of this in detail.

We will define $s(x, y)$ to be the norm of relativistic space–time. This quantity may be considered to be the invariant "distance" (squared) between two events, x and y, and of course is one of the fundamental objects associated with the construction of differentials. Since quantities that are unchanged by a geometric transformation are called **scalars** it is evident that $s(x, y)$ is a 4–scalar. Since the first postulate states that the laws of physics must be invariant under homogeneous (at least) Lorentz transformations, they must ultimately be based on Lorentz scalars. Indeed, the Lagrangian densities upon which field theories are based are generally constructed to be Lorentz scalars. This is a strong constraint on allowed theories.

These scalars are, however, formed out of 4–vectors (as we see above) or, more generally, the contraction of 4–tensors. We must, therefore, determine the general transformation properties of a tensor of arbitrary rank to completely determine a theory. In the part of this book devoted to mathematical physics is an entire chapter that discusses tensors, in particular the definitions of covariant and contravariant tensors, how to contract (Einstein sum) pairs of tensors to form tensors of lower rank, and the role of the metric tensor in defining tensor coordinate frames and transformations thereupon. We will *not repeat* this review or introduction (depending on the student) and urge students to at this time spend an hour or so working through this chapter before continuing (even if you've seen it before).

16.2 Tensors in 4 Dimensions

Let us now consider the *specific* nature of tensors on four-dimensional spacetime. Tensors of rank k^4 are categorized (for each coordinate index) by their transformation properties relative to a transformation of the underlying coordinate system $x \to x'$ as defined above. This transformation is implicit in all the discussion below.

A **scalar** (tensor of rank zero) is unchanged by such a transformation. *This is not a trivial statement!* It is trivial for scalar numbers like π, no doubt, but in physics the *interesting* part of this requirement occurs when discussing the scalars that result *algebraically* from fully contracting products of tensors over all of their indices using the *metric tensor*. This will be made quite clear below.

For a **vector** (tensor of rank one) we have two possibilities. Either it transforms like the coordinate itself and we have a

contravariant vector (A^0, A^1, A^2, A^3) such that

$$A^\alpha = \frac{\partial \bar{x}^\alpha}{\partial x^\beta} A^\beta \tag{16.3}$$

(noting that all the indices are on *top*, along with the new *primed* coordinate). This makes the differential transformation relationship to the underlying ordinary (contravariant) coordinates explicit and is obviously an identity for those coordinates.

Alternatively, we have a

covariant vector (B_0, B_1, B_2, B_3) such that

$$B_\alpha = \frac{\partial x^\beta}{\partial \bar{x}^\alpha} B_\beta \tag{16.4}$$

(with the coordinate indices on top and the new primed coordinate on the bottom). Again, note that this is precisely what we expect – the transformation is in the *opposite* sense of that of the underlying coordinates. We need in both cases, of course, to figure out the matrix of e.g. $\frac{\partial x^\beta}{\partial \bar{x}^\alpha}$ explicitly.

[4]The **rank** of a tensor is determined by the number of indices it has. Scalars are 0th rank, vectors are 1st rank, 2D matrices are 2nd rank, and our old friend ϵ_{ijk} is a third rank fully antisymmetric tensor.

In a moment we will see explicitly what exactly the difference is between these two types of first rank tensors. First, however, we should note that

contravariant tensors of rank 2 transform like

$$F^{\alpha\beta} = \frac{\partial \bar{x}^\alpha}{\partial x^\gamma}\frac{\partial \bar{x}^\beta}{\partial x^\delta}F^{\gamma\delta}. \tag{16.5}$$

Similarly, we have

covariant tensors of rank 2

$$G_{\alpha\beta} = \frac{\partial x^\gamma}{\partial \bar{x}^\alpha}\frac{\partial x^\delta}{\partial \bar{x}^\beta}G_{\gamma\delta} \tag{16.6}$$

and

mixed tensors of rank 2

$$H^\alpha_\beta = \frac{\partial \bar{x}^\alpha}{\partial x^\gamma}\frac{\partial x^\delta}{\partial \bar{x}^\beta}H^\gamma_\delta. \tag{16.7}$$

It is clearly a trivial exercise to determine the co/contra variant transformation properties of higher rank tensors. We can form higher rank tensors by means of an *outer* (dyadic) *product*, where we simply take two tensors of some rank and multiply them out componentwise, preserving products of any underlying basis vectors as they occur. For example we can construct a second rank tensor by:

$$F^{\alpha\beta} = A^\alpha B^\beta \tag{16.8}$$

where α and β run over the full range of index values. Note well that this defines a *square matrix* in this case of basis vector dyads as objects such as $\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}$, ... occur.

One important question is whether *all* e.g. second rank tensors can be written as products of first rank tensors. It is not the general case that this is possible, but in many of our uses of these ideas in physics it will be. In this case the generalized product forms a *division algebra* where we can *factor* e.g. second rank tensors into first rank tensors in various ways. Division algebras are discussed in the Mathematical Physics section as well, and interested students should return there to read about *geometric* *algebras*, the result of fully generalizing the notion of complex numbers to complex spaces of arbitrary dimension while preserving the factorizability of the algebraic objects.

In addition to *extending* the rank of tensor objects by forming dyadic, triadic, or n-adic products of tensors, we can *reduce* the rank of tensors by means of a process called *contraction*. A contraction of two tensors is the result of setting two of the indices (typically they must be a covariant/contravariant pair) to be equal and performing the Einstein summation over the shared range. This reduces the rank of the expression by one relative to that of its constituents, hence the term "contraction". An expression can be contracted over several components at a time when doing algebra so second rank tensors can be contracted to form a 4-scalar, for example, or third rank tensors can be contracted to first.

Our familiar notion of multiplying a vector by a matrix to produce a vector in proper tensor language is to form the outer product of the matrix (second rank tensor) and the vector (first rank tensor), set the rightmost indices to be equal and sum over that index to produce the resulting first rank tensor.

Hence we define our **scalar product** to be the **contraction** of a covariant and contravariant vector.

$$B \cdot A = B_\alpha A^\alpha \tag{16.9}$$

Note that I've introduced a sort of "sloppy" convention that a single quantity like B or A can be a four-vector in context. Clearly the expression on the right side is less ambiguous!

Then:

$$\begin{aligned} B' \cdot A' &= \frac{\partial x^\gamma}{\partial \bar{x}^\alpha} B_\gamma \frac{\partial \bar{x}^\alpha}{\partial x^\delta} A^\delta \\ &= \frac{\partial x^\gamma}{\partial x^\delta} B_\gamma A^\delta \\ &= \delta_{\gamma\delta} B_\gamma A^\delta \\ &= B_\delta A^\delta = B \cdot A \end{aligned} \tag{16.10}$$

and the desired invariance property is proved. Hmmm, that was pretty easy! Maybe there is something to this notation thing after all!

16.3 The Metric Tensor

The section above is still very generic and little of it depends on whether the tensors are three or four or ten dimensional. We now need to make them work for the *specific* geometry we are interested in, which is one where we will ultimately be seeking transformations that preserve the *invariant interval*:

$$(ds)^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \tag{16.11}$$

as this is the one that directly encodes an invariant speed of light.

From this point on, we must be careful not to confuse $x \cdot x = x^2$ and $x^2 = y$, etc. Contravariant indices should be clear from context, as should be powers. To simplify life, algebraically indices are always greek (4–vector) or roman italic (3–vector) while powers are still powers and hence are generally integers.

Let us write this in terms of only contravariant pieces dx^μ. This requires that we introduce a *relative minus sign* when contracting out the components of the *spatial part of the differential only*. We can most easily encode this requirement into a special matrix (tensor) called the *metric tensor* as:

$$(ds)^2 = g_{\alpha\beta}\; dx^\alpha dx^\beta \tag{16.12}$$

The tensor g obviously satisfies the following property:

$$g_{\alpha\beta} = g_{\beta\alpha} \tag{16.13}$$

(that is, it is symmetric) because the multiplication in the Einstein summation is ordinary multiplication and hence commutative. It is called the **metric tensor** because it defines the way **length** is **measured**.

At this point if we were going to discuss general relativity we would have to learn what a *manifold*[5] s. Technically, a manifold is a coordinate system that may be curved but which is *locally* flat. By locally flat I mean very specifically that one can cover the entire space with "patches" in the neighborhood of points where the coordinate system is locally Euclidean (e.g. Cartesian). An example of a curved space manifold is the surface of a sphere (think the surface of the earth). When we look down at the

[5]Wikipedia: http://www.wikipedia.org/wiki/Manifold. i

ground beneath our feet, it looks quite flat and we can draw triangles on it that appear to have interior angles that sum to π and we can draw a map of (say) our county that more or less accurately encodes distances on the ground in terms of distances measured on the map. However, if we take too *big* a patch all of this breaks down. The angles in a triangle sum to strictly *more* than π radians. Maps have to be distorted and chopped into pieces to correctly represent distances on the ground as distances on the flat 2-dimensional map. This is how a manifold works – we can work with it in the *local neighborhood* of any point as if it is flat, but if we go too far we have to work harder and correct for its curvature, where "too far" is obviously defined in terms of the scale of its curvature and some common sense.

General relativity introduces the hypothesis that gravitational fields bend space-time. However, this bending is very, very slight unless one is in a very *strong* gravitational field, and this bending preserves a local smoothness of space-time so that space-time, although it is no longer strictly Euclidean, is still a manifold and we can do all sorts of transformations in a very general way as long as we restrict the results to a locally flat patch.

In our discussion of *special* relativity we will assume from the beginning that our space–time *is* flat and not bent by strong gravitational fields. In this case the metric tensor can be expressed in a very simple form. We will use the Lorentz metric (as opposed to the Minkowski metric that uses $x^4 = ict$ instead of x^0). Using our definitions of the $\mu = 0, 1, 2, 3$ coordinates, g in the differentials above is just:

$$g_{00} = 1, g_{11} = g_{22} = g_{33} = -1 \tag{16.14}$$

and we see that it is not just symmetric, it is diagonal.

The contravariant and mixed metric tensors for flat space–time are the same (this follows by considering the $\frac{\partial x^\alpha}{\partial x^\beta}$ coordinate transformation matrices that define co- and contra-variance):

$$g_{\alpha\beta} = g_\alpha^\beta = g^{\alpha\beta}. \tag{16.15}$$

Finally, the contraction of any two metric tensors is the "identity" tensor,

$$g_{\alpha\gamma} g^{\gamma\beta} = \delta_\alpha^\beta = \delta_{\alpha\beta} = \delta^{\alpha\beta}. \tag{16.16}$$

Since we want $(ds)^2$ to be (to contract to) a scalar, it is clear that:

$$x_\alpha = g_{\alpha\beta} x^\beta \tag{16.17}$$

$$x^\alpha = g^{\alpha\beta} x_\beta \tag{16.18}$$

or the metric tensor can be used to raise or lower arbitrary indices, converting covariant indices to contravariant and vice–versa:

$$F^{\mu\alpha\nu} = g^{\alpha\beta} F^{\mu\nu}_{\beta} \tag{16.19}$$

This is an important trick! Note well that in order to perform a contraction that reduces the rank of the expression by one, the indices being summed *must* occur as a co/contra pair (in either order). If both are covariant, or both are contravariant, one or the other must be raised or lowered by contracting it with the metric tensor before contracting the overall pair! We use this repeatedly in the algebra in sections below.

Finally we are in a position to see how covariant and contravariant vectors differ (in this metric). We have already seen that "ordinary" vectors must linearly transform like contravariant vectors. Given a contravariant vector (A^0, A^1, A^2, A^3) we thus see that

$$A_0 = A^0, A_1 = -A^1, A_2 = -A^2, A_3 = -A^3 \tag{16.20}$$

or

$$A^\alpha = (A^0, \mathbf{A}), A_\alpha = (A^0, -\mathbf{A}). \tag{16.21}$$

Covariant vectors are just spatially inverted contravariant vectors. Note that this definition, together with our definition of the general scalar product, reconstructs the desired invariant:

$$B \cdot A = B_\alpha A^\alpha = (B^0 A^0 - \mathbf{B} \cdot \mathbf{A}) \tag{16.22}$$

This tells us how ordinary quantities transform. However, we are also interested in how tensor differentials transform, since these are involved in the construction of a dynamical system. By considering the chain rule we see that

$$\frac{\partial}{\partial \bar{x}^\alpha} = \frac{\partial x^\beta}{\partial \bar{x}^\alpha} \frac{\partial}{\partial x^\beta} \tag{16.23}$$

or, differentiation by a **contra**variant coordinate transforms like a **covariant vector operator**. This is more or less the definition of covariant, in fact. Similarly, differentiation with respect to a **co**variant vector coordinate transforms like a **contravariant vector operator**. This also follows from the above by using the metric tensor,

$$\frac{\partial}{\partial x_\alpha} = g_{\alpha\beta} \frac{\partial}{\partial x^\beta}. \tag{16.24}$$

It is tedious to write out all of the pieces of partial derivatives w.r.t. various components, so we (as usual, being the lazy sorts that we are) introduce a "simplifying" notation. It does, too, after you get used to it.

$$\begin{aligned}
\partial^\alpha &= \frac{\partial}{\partial x_\alpha} = (\frac{\partial}{\partial x^0}, -\boldsymbol{\nabla}) & (16.25)\\
\partial_\alpha &= \frac{\partial}{\partial x^\alpha} = (\frac{\partial}{\partial x^0}, +\boldsymbol{\nabla}). & (16.26)
\end{aligned}$$

Note that we have cleverly indicated the co/contra nature of the vector operators by the placement of the index on the bare partial.

We cannot resist writing down the 4–divergence of a 4–vector:

$$\partial^\alpha A_\alpha = \partial_\alpha A^\alpha = \frac{\partial A^0}{\partial x^0} + \boldsymbol{\nabla} \cdot \mathbf{A} = \frac{1}{c}\frac{\partial A^0}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{A} \tag{16.27}$$

which looks a *lot* like a continuity equation or a certain well–known gauge condition. (Medidate on just what A^μ would need to be for either of these equations to be realized as a four-scalar). Hmmmmmm, I say.

Even more entertaining is the 4–Laplacian, called the D'Lambertian operator:

$$\begin{aligned}
\Box &= \partial_\alpha \partial^\alpha = \frac{\partial^2}{\partial x^{02}} - \nabla^2 & (16.28)\\
&= \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 & (16.29)
\end{aligned}$$

which just happens to be the (negative of the) **wave operator**! *Hmmmmmmmm!* By strange coincidence, certain objects of great importance in electrodynamics "just happen" to be Lorentz scalars! Remember that I *did* say above that part of the point of introducing this lovely tensor notation was to make the various transformational symmetries of physical quantities manifest, and this appears to be true with a vengeance!

That was the "easy" part. It was all geometry. Now we have to do the messy part and derive the infinitesimal transformations that leave scalars in this metric invariant.

16.4 Generators of the Lorentz Group

Let

$$x = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{16.30}$$

be a column vector. Note that we no longer indicate a vector by using a vector arrow and/or boldface – those are reserved for the spatial part of the four-vector only. Then a "matrix" scalar product is formed in the usual way by

$$(a, b) = \tilde{a}b \tag{16.31}$$

where $\tilde{a}$ is the (row vector) transpose of a. The metrix tensor is just a matrix:

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{16.32}$$

and $g^2 = \overleftrightarrow{\mathbf{I}}$. Finally,

$$gx = \begin{pmatrix} x^0 \\ -x^1 \\ -x^2 \\ -x^3 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{16.33}$$

In this compact notation we define the scalar product *in this metric* to be

$$a \cdot b = (a, gb) = (ga, b) = \tilde{a}gb = a^\alpha g_{\alpha\beta} b^\beta = a^\alpha b_\alpha. \tag{16.34}$$

We seek the set (group, we hope) of linear transformations that leaves $(x, gx) = x \cdot x$ invariant. Since this is the "norm" (squared) of a four vector, these are "length preserving" transformations in this four dimensional metric. That is, we want all matrices A such that

$$x' = Ax \tag{16.35}$$

leaves the norm of x invariant,

$$x' \cdot x' = \tilde{x}' g x' = \tilde{x} g x = x \cdot x \tag{16.36}$$

or

$$\tilde{x}\tilde{A}gAx = \tilde{x}gx \tag{16.37}$$

or

$$\tilde{A}gA = g. \tag{16.38}$$

Clearly this last condition is sufficient to ensure this property in A.

Now,

$$\det\left|\tilde{A}gA\right| = \det|g|\,(\det|A|)^2 = \det|g| \tag{16.39}$$

where the last equality is required. But $\det|g| = -1 \neq 0$, so

$$\det|A| = \pm 1 \tag{16.40}$$

is a **constraint** on the allowed matrices (transformations) A. There are thus two **classes** of transformations we can consider. The

proper Lorentz transformations with $\det|A| = +1$; and

improper Lorentz transformations with $\det|A| = \pm 1$.

Proper L. T.'s contain the identity (and thus can form a group by themselves), but **improper** L. T.'s can have either sign of the determinant. This is a signal that the metric we are using is "indefinite". Two examples of improper transformations that illustrate this point are spatial inversions (with $\det|A| = -1$) and $A = -I$ (space *and* time inversion, with $\det|A| = +1$).

In very general terms, the proper transformations are the continuously connected ones that form a Lie group, the improper ones include one or more inversions and are not equal to the product of any two proper transformations. The proper transformations are a subgroup of the full group — this is not true of the improper ones, which, among other things, lack the identity. With this in mind, let us review the properties of infinitesimal linear transformations, preparatory to deducing the particular ones that form the homogeneous Lorentz group.

16.4.1 Infinitesimal Transformations

We seek (Lie) groups of continous linear transformations,

$$x' = T_a x \tag{16.41}$$

or

$$x'^{\mu} = f^{\mu}(x; a) \tag{16.42}$$

for $\mu = 1, 2, \ldots n$. We require that the $a = a_1, \ldots, a_r$ are r real numbers (parameters) that characterize the transformation. r must be **minimal** ("essential").

Examples of transformations of importance in physics (that you should already be familiar with) include

$$\begin{aligned} x' &= T_d x \\ &= x + d \end{aligned} \tag{16.43}$$

where $d = (d_1, \ldots, d_n)$. This is the (n parameter) **translation** group in n dimensions. Also,

$$x'^{i} = R_{ij} x^{j} \tag{16.44}$$

where

$$R\tilde{R} = I, \qquad \det |R| > 0, \qquad i = 1, 2, 3 \tag{16.45}$$

is the (three parameter) **rotation group**.

An **infinitesimal transformation** in one of the parameters is defined by

$$T_{a(0)+\epsilon} = I_{\epsilon} + \mathcal{O}(\epsilon^2). \tag{16.46}$$

In this definition, $a(0)$ are the (r–parameter) values associated with the identity transformation I. These can be chosen to be zero by suitably choosing the parameter coordinates. The infinitesimal parameters ϵ_u are taken to zero, so that $\epsilon^2 = \epsilon_u \epsilon_u$ (summed) is neglible. Thus

$$I_{\epsilon} = I + \epsilon_u Q_u \tag{16.47}$$

where

$$Q_u = f^{\mu_u}(x) \frac{\partial}{\partial x^{\mu}} \tag{16.48}$$

and

$$f^{\mu_u}(x) = \left. \frac{\partial f^{\mu}(x, a)}{\partial a_u} \right|_{a=a(0)}. \tag{16.49}$$

Putting this all together,

$$\begin{aligned} x' &= (T_{a(0)+\epsilon}) x = (I + \epsilon_u Q_u)\, x \\ &= I x + \epsilon_u Q_u x \\ &= x + \epsilon_u \left. \frac{\partial f^{\mu}(x, a)}{\partial a_u} \right|_{a=a(0)} \frac{\partial x}{\partial x^{\mu}} \end{aligned} \tag{16.50}$$

(summed over $\mu = 0, \ldots, 3$ in four dimensional space–time and $u = 0, \ldots, r$). Thus (unsurprisingly)

$$x'^\mu = x^\nu \delta^\mu_\nu + \epsilon_u \left.\frac{\partial f^\nu}{\partial a_u}\right|_{a=a(0)} g^\mu_\nu \tag{16.51}$$

which has the form of the first two terms of a Taylor series. This is characteristic of infinitesimal linear transformations.

One can easily verify that

$$I_\epsilon I_{\epsilon'} = I_{\epsilon'} I_\epsilon \tag{16.52}$$

(infinitesimal transformations commute) and that

$$I_\epsilon^{-1} = I_{-\epsilon} \tag{16.53}$$

(to order ϵ^2). They thus have an identity, an inverse, and can be shown to be associative.

The continuous transformation group (mentioned above) follows immediately from making d_u (the displacement of coordinates) infinitesimal and finding finite displacements by integration. The rotation group (matrices) are a little trickier. They are

$$I_\epsilon = I + gS \tag{16.54}$$

where

$$\tilde{S} = -S, \qquad S_{ij} = \epsilon_k \left.\frac{\partial R_{ij}}{\partial a_k}\right|_{a(0} . \tag{16.55}$$

The infinitesimal S are antisymmetric and traceless (in 3D), so they have only three independent parameters (that are thus "essential"). We can write them generally as

$$S_{ij} = \epsilon_{ijk}\; d\omega_k \tag{16.56}$$

where the $d\omega_k$ is the infinitesimal parameter and where ϵ_{ijk} is the antisymmetric unit tensor. Thus, if

$$dx_i = x'_i - x_i = S_{ij} x_j = \epsilon_{ijk}\; x_j\; d\omega_k \tag{16.57}$$

we see that

$$d\boldsymbol{x} = \boldsymbol{x} \times d\boldsymbol{\omega} \tag{16.58}$$

A moment of thought should convince you that $d\boldsymbol{\omega}$ is the infinitesimal (vector) rotation angle, with direction that points along the axis of rotation.

To obtain the rotation **group** we must show that *every* rotation can be obtained by integrating $I_{d\omega}$. This follows by writing an arbitrary rotation or product of rotations as a single rotation about a fixed axis. For $d\boldsymbol{\omega}$ parallel to this axis $\boldsymbol{\Omega}$, this is obviously true, as I show next. Since any rotation can be written this way, the rotations indeed form a group.

The integration proceeds like:

$$R_{\boldsymbol{\Omega}} = \lim_{\Delta\omega \to 0} (R_{\Delta\omega})^{\Omega/\Delta\omega} \tag{16.59}$$

where $\Delta\omega = |\Delta\boldsymbol{\omega}|$ and $\Omega = |\boldsymbol{\Omega}|$. We can parameterize this as

$$R_{\boldsymbol{\Omega}} = \lim_{m \to \infty} (I + \frac{1}{m}\Omega S_0)^m = e^{\Omega S_0} \tag{16.60}$$

where

$$(S_0)_{ij} = \epsilon_{ijk} \frac{\Delta\omega_k}{\Delta\omega}. \tag{16.61}$$

Believe it or not, this was one of the primary things we wanted to show in this aside. What it shows is that rotations about an arbitrary axis can be written as an *exponential* that can be thought of as the *infinite product* of a series of *infinitesimal transformations* where each transformation has various nice properties.

With these known results from simpler days recalled to mind, we return to the homogeneous, proper Lorentz group. Here we seek the infinitesimal linear transformations, etc. in *four* dimensions. Algebraically one proceeds almost identically to the case of rotation, but now in four dimensions and with the goal of preserving length in a different metric. A general infinitesimal transformation can be written compactly as:

$$I_\epsilon = I + gL \tag{16.62}$$

where (as before) $\tilde{gL} = -gL$ (and hence gL is traceless), L is infinitesimal, and where g is the usual metric tensor (that follows from all the annoying derivatives with respect to the parameters and coordinates).

Thus

$$A = \lim_{m \to \infty} \left(I + \frac{1}{m}L\right)^m = e^L \tag{16.63}$$

defines the form of a general transformation matrix associated with a given "direction" in the parameter space constructed from an infinite product of infinitesimal transformations, each of which is basically the leading term of a Taylor series of the underlying coordinate function transformation in terms of the parameters. This justifies the "ansatz" made by Jackson. The matrices L are called the **generators** of the linear transformation.

Thus, whenever we write

$$A = e^L \tag{16.64}$$

where the L's are (to be) the generators of the Lorentz group transformations we should remember what it stands for. Let's find the distinct L. Each one is a 4×4 real, traceless matrix that is (as we shall see) antisymmetric in the spatial part (since gL is antisymmetric from the above).

To construct A (and find the distinct components of L) we make use of its properties. Its determinant is

$$\det|A| = \det\left|(e^L)\right| = e^{\mathrm{Tr}\ L} = \pm 1 \tag{16.65}$$

(This follows from doing a similarity transformation to put A in diagonal form. L is necessarily then diagonal. Similarity transformations do not alter the determinant, because

$$\det\left|S^{-1}MS\right| = \det\left|S^{-1}\right| \det|M| \det|S| = \det|M|\,. \tag{16.66}$$

If L is diagonal, then the last equation follows from the usual properties of the exponential and the definition of the exponential of a matrix.)

If L is real then $\det|A| = -1$ is excluded by this result. If L is traceless (and only if, given that it is real), then

$$\det|A| = +1 \tag{16.67}$$

which is *required* to be true for proper Lorentz transformations (recall from last time). Making L a traceless 4x4 matrix therefore suffices to ensure that we will find only proper Lorentz transformations.

Think back to the requirement that:

$$\tilde{A}gA = g \tag{16.68}$$

in order to preserve the invariant interval where

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{16.69}$$

and L is a real, traceless, 4×4 matrix.

If we multiply from the right by A^{-1} and the left by g, this equation is equivalent also to

$$g\tilde{A}g = A^{-1}. \tag{16.70}$$

Since $\tilde{A} = e^{\tilde{L}}$, $A^{-1} = e^{-L}$, and $I = g^2$:

$$g\tilde{A}g = e^{g^2\tilde{L}} = e^{g\tilde{L}g} = e^{-L} \tag{16.71}$$

or

$$g\tilde{L}g = -L. \tag{16.72}$$

(This can also easily be proven by considering the "power series" or product expansions of the exponentials of the associated matrices above, changing the sign/direction of the infinitesimal series.)

Finally, if we multiply both sides from the *left* by g and express the left hand side as a transpose, we get

$$\widetilde{gL} = -gL. \tag{16.73}$$

From this we see that the matrix gL is traceless and antisymmetric as noted/expected from above. If we mentally factor out the g, we can without loss of generality write L as:

$$L = \begin{pmatrix} 0 & L_{01} & L_{02} & L_{03} \\ L_{01} & 0 & L_{12} & L_{13} \\ L_{02} & -L_{12} & 0 & L_{23} \\ L_{04} & -L_{13} & -L_{23} & 0 \end{pmatrix}. \tag{16.74}$$

This matrix form satisfies all the constraints we deduced above for the generators. Any L of this form will make an A that preserves the invariant interval (length) of a four vector. There are exactly six **essential** parameters as expected. Finally, if we use our intuition, we would *expect* that the L_{ij} for $i, j = 1, 2, 3$ form the *rotation subgroup* and describe physical rotations.

So this is just great. Let us now separate out the individual couplings for our appreciation and easy manipulation. To do that we define six fundamental matrices (called the **generators** of the group from which we can construct an arbitrary L and hence A. They are basically the individual matrices with unit or zero components that can be scaled by the six parameters $L_{\mu\nu}$. The particular choices for the signs make certain relations work out nicely:

$$S_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{16.75}$$

$$S_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{16.76}$$

$$S_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{16.77}$$

$$K_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{16.78}$$

$$K_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{16.79}$$

$$K_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{16.80}$$

The matrices S_i generate **rotations** in the spatial part and the matrices K_i generate **boosts**. Note that the squares of these matrices are diagonal and either $+1$ or -1 in the submatrix involved:

$$S_1^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{16.81}$$

and

$$K_1^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{16.82}$$

etc. From this we can deduce that

$$S_i^3 = -S_i \tag{16.83}$$

$$K_i^3 = K_i. \tag{16.84}$$

Note that these relations are *very similar* to the multiplication rules for unit pure complex or pure real numbers.

The reason this is important is that if we form the dot product of a vector of these generators with a spatial vector (effectively decomposing a vector parameter in terms of these matrices) in the exponential expansion, the following relations can be used to reduce powers of the generators.

$$(\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{S})^3 = -\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{S} \tag{16.85}$$

and

$$(\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{K})^3 = \hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{K} \tag{16.86}$$

In these expressions, $\hat{\boldsymbol{\epsilon}}$ an arbitrary unit vector, and these expressions effectively match up the generator axes (which were arbitrary) with the direction of the parameter vector for rotation or boost respectively. After the reduction (as we shall see below) the exponential is, in fact, a well-behaved and easily understood matrix!

It is easy (and important!) to determine the commutation relations of these generators. They are:

$$[S_i, S_j] = \epsilon_{ijk} S_k \tag{16.87}$$

$$[S_i, K_j] = \epsilon_{ijk} K_k \tag{16.88}$$

$$[K_i, K_j] = -\epsilon_{ijk} S_k. \tag{16.89}$$

The first set are immediately recognizable. They tells us that "two rotations performed in both orders differ by a rotation". The second and third show that "a boost and a rotation differ by a boost" and "two boosts differ by a rotation", respectively. In quotes because that is somewhat oversimplified, but it gets some of the idea across.

These are the generators for the groups $SL(2, C)$ or $O(1, 3)$. The latter is the group of relativity as we are currently studying it.

A question that has been brought up in class is "where is the factor i in the generators of rotation" so that $\boldsymbol{S} \times \boldsymbol{S} = i\boldsymbol{S}$ as we might expect from considering spin and angular momentum in other contexts. It is there, but subtly hidden, in the fact that $S_i^2 = -I_i$ *in the projective block* of the rotation matrices only. Matrices appear to be *a* way to represent geometric algebras, as most readers of this text should already know from their study of the (quaternionic) Pauli spin matrices. We won't dwell on this here, but note well that the Pauli matrices $I, \sigma_1, \sigma_2, \sigma_3$ are isomorphic to the unit quaternions $1, i, j, k$ via the mapping $I \to 1$, $\sigma_1\sigma_2 \to i$, $\sigma_3\sigma_1 \to j$, $\sigma_2\sigma_3 \to k$ as the reader can easily verify[6] Note well that:

$$\sigma_3\sigma_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{16.90}$$

is both real and, not at all coincidentally, the structure of an S sub-block.

With these definitions in hand, we can easily decompose L in terms of the $\boldsymbol{S}$ and the $\boldsymbol{K}$ matrices. We get:

$$L = -\boldsymbol{\omega} \cdot \boldsymbol{S} - \boldsymbol{\xi} \cdot \boldsymbol{K} \tag{16.91}$$

where $\boldsymbol{\omega}$ is a (finite) rotation around an axis in direction $\hat{\omega}$ and where $\boldsymbol{\xi}$ is a (finite) boost in direction $\hat{\xi}$. Thus the **completely general** form of A is

$$A = e^{-\boldsymbol{\omega} \cdot \boldsymbol{S} - \boldsymbol{\xi} \cdot \boldsymbol{K}}. \tag{16.92}$$

The (cartesian) components of $\boldsymbol{\omega}$ and $\boldsymbol{\xi}$ are now the six free parameters of the transformation.

[6] And should! That's right, you students, you know who I'm talking to. So here's a question for you: Are $I, \sigma_3\sigma_1$ a real isomorphism to complex numbers? What would the various results of the introduction to complex numbers look like expressed in terms of these two matrices? What *in particular* does multiplying by a unimodular "complex number" such as $\cos(\theta)I + \sin(\theta)\sigma_3\sigma_1$ look like? Hmmm... veeeery interesting.

Let us see that these are indeed the familiar boosts and rotations we are used to. After all, this exponential notation is not transparent. Suppose that $\boldsymbol{\omega} = 0$ and $\boldsymbol{\xi} = \xi\hat{x}$. Then $L = -\xi K_1$ and

$$\begin{aligned} A &= e^L = I - \xi K_1 + \frac{1}{2!}(\xi K_1)^2 - \frac{1}{3!}(\xi K_1)^3 + \ldots \\ &= (I - K_1^2) - K_1(\xi + \frac{1}{3!}\xi^3 + \ldots) + K_1^2(I + \frac{1}{2!}\xi^2 + \ldots) \\ &= (I - K_1^2) - K_1 \sinh(\xi) + K_1^2 \cosh(\xi) \end{aligned} \tag{16.93}$$

or (in matrix form)

$$A = \begin{pmatrix} \cosh(\xi) & -\sinh(\xi) & 0 & 0 \\ -\sinh(\xi) & \cosh(\xi) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{16.94}$$

which (ohmygosh!) is our old friend the Lorentz transformation, just like we derived it *a la* kiddy–physics–wise. As an exercise, show that the $\boldsymbol{\omega} = \omega\hat{x}, \boldsymbol{\xi} = 0$ result is a rotation around the x axis. Note that the step of "adding and subtracting" S_1^2 is essential to reconstructing the series of the sine and cosine, just like the K_1 was above for cosh and sinh.

Now, a boost in an *arbitrary* direction is just

$$A = e^{-\boldsymbol{\xi}\cdot\boldsymbol{K}}. \tag{16.95}$$

We can certainly parameterize it by

$$\boldsymbol{\xi} = \hat{\beta}\tanh^{-1}\beta \tag{16.96}$$

(since we know that $\beta = \hat{\xi}\tanh\xi$, inverting our former reasoning for $|\boldsymbol{\beta}| \in [0,1]$. Then

$$A(\boldsymbol{\beta}) = e^{-\hat{\beta}\cdot\boldsymbol{K}\tanh^{-1}\beta}. \tag{16.97}$$

I can do no better than quote Jackson on the remainder:

"It is left as an exercise to verify that ..."

$$A(\boldsymbol{\beta}) = \begin{pmatrix} \gamma & -\gamma\beta_1 & -\gamma\beta_2 & -\gamma\beta_3 \\ -\gamma\beta_1 & 1 + \frac{(\gamma-1)\beta_1^2}{\beta^2} & \frac{(\gamma-1)\beta_1\beta_2}{\beta^2} & . \\ . & -\frac{(\gamma-1)\beta_1\beta_2}{\beta^2} & . & . \\ . & . & . & . \end{pmatrix} \tag{16.98}$$

(etc.) which is just the explicit full matrix form of

$$x^{0\prime} = \gamma(x^0 - \boldsymbol{\beta}\cdot\boldsymbol{x}) \tag{16.99}$$

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{(\gamma-1)}{\beta^2}(\boldsymbol{\beta}\cdot\boldsymbol{x})\boldsymbol{\beta} - \gamma\boldsymbol{\beta}x^0 \tag{16.100}$$

from before.

Now, we have enough information to construct the exact form of a simultaneous boost and rotation, but this presents a dual problem. When we go to factorize the results (like before) the components of independent boosts and rotations do not commute! If you like,

$$A(\boldsymbol{\beta},0)A(0,\boldsymbol{\omega}) \neq A(0,\boldsymbol{\omega})A(\boldsymbol{\beta},0) \tag{16.101}$$

and we cannot say anything trivial like

$$A(\boldsymbol{\beta},\boldsymbol{\omega}) = A(\boldsymbol{\beta},0)A(0,\boldsymbol{\omega}) \tag{16.102}$$

since it depends on the order they were performed in! Even worse, the product of two boosts is equal to a single boost and a rotation (if the boosts are not in the same direction)!

The worst part, of course, is the algebra itself. A useful exercise for the algebraically inclined might be for someone to construct the general solution using, e.g. – mathematica.

This suggests that for *rotating* relativistic systems (such as atoms or orbits around neutron stars) we may need a kinematic correction to account for the successive frame changes as the system rotates.

The atom perceives itself as being "elliptically deformed". The consequences of this are observable. This is known as "Thomas precession".

16.5 Thomas Precession

We must begin our discussion by noting that the magnetic moment of an electron is (according to the "Uhlenbeck-Goudsmit hypothesis")

$$\boldsymbol{\mu} = \frac{ge}{2mc}\mathbf{s} \tag{16.103}$$

where $\boldsymbol{s}$ is the (half integer) spin of the electron in units of $\hbar$ and where g is the "g–factor" introduced to accomodate two distinct results. The splitting of the observed spectra in an applied magnetic field $\mathbf{B}$ via the anomalous Zeeman interaction:

$$U_{\rm AZ} = -\frac{ge}{2mc}\mathbf{s}\cdot\mathbf{B} \tag{16.104}$$

was correctly predicted only if $g = 2$. On the other hand (as we shall see), the simple classical argument that led to this result also led to a spin orbit interaction

$$U_{\rm SO} = \frac{g}{2m^2c^2}(\mathbf{s}\cdot\mathbf{L})\frac{1}{r}\frac{dV}{dr} \tag{16.105}$$

(where $\mathbf{L} = m(\mathbf{r}\times\mathbf{v})$ is the orbital angular momentum of the electron) that was a factor of, curiously enough, 2 too large. That is, the fine–structure intervals observed in nature were only half the theoretically predicted values. If $g = 1$ was chosen instead, the splittings were correct but the Zeeman effect was then normal (instead of anomalous, as observed).

I don't have the time to go into more detail on, for example, what the Zeeman effect (splitting of energy levels in an applied magnetic field) is. In any event, it is strictly a quantum effect, and you should study it soon in elementary quantum theory, if you haven't already.

Thomas (who taught for years over at NC State) showed in 1927 that the discrepancy is due to a relativistic kinematic correction like that we previously considered. In a nutshell, the rest frame of the electron rotates as well as translates (boosts) and we must therefore take into account both kinematical effects. This results in an additional (Thomas) "precession" of the frames. When Thomas precession is taken into account, not only are both the fine structure and anomalous Zeeman effect in atoms accomodated, but a deeper understanding of the spin–orbit interaction in nuclear physics (and rotating frames in general) also results.

Let us begin by (naívely) deriving the spin–interaction energy. Suppose the electron moves with velocity $\mathbf{v}$ in external fields $\mathbf{E}$ and $\mathbf{B}$. Then the torque on the electron in its rest frame is just

$$\left(\frac{d\mathbf{s}}{dt}\right)_{\text{rest frame}} = \boldsymbol{\mu}\times\mathbf{B}' \tag{16.106}$$

where $\mathbf{B}'$ is the magnetic field in that frame.

As we will show very soon, the magnetic field transforms like

$$\mathbf{B}' = \left(\mathbf{B} - \frac{\mathbf{v}}{c} \times \mathbf{E}\right) \tag{16.107}$$

to order v^2/c^2. Then

$$\left(\frac{d\mathbf{s}}{dt}\right)_{\text{rest frame}} = \boldsymbol{\mu} \times \left(\mathbf{B} - \frac{\mathbf{v}}{c} \times \mathbf{E}\right). \tag{16.108}$$

Associated with this torque there is an interaction energy

$$U' = -\boldsymbol{\mu} \cdot \left(\mathbf{B} - \frac{\mathbf{v}}{c} \times \mathbf{E}\right). \tag{16.109}$$

The electric force $e\mathbf{E}$ is very nearly the negative gradient of a spherically averaged potential energy $V(r)$. For one electron atoms this is exact; it is a good approximation for all the others. Thus we will try using

$$e\mathbf{E} = -\frac{\mathbf{r}}{r}\frac{dV}{dr} \tag{16.110}$$

in the equation for the spin interaction energy:

$$U' = -\frac{ge}{2mc}\mathbf{s} \cdot \mathbf{B} + \frac{g}{2m^2c^2}(\mathbf{s} \cdot \mathbf{L})\frac{1}{r}\frac{dV}{dr} \tag{16.111}$$

(where $\mathbf{L} = m(\mathbf{r} \times \mathbf{v})$ for the orbiting electron). This gives the anomalous Zeeman effect correctly (from the first term) but the spin orbit (fine structure) splitting is a factor of two too large. Too bad!

The error is, in a nutshell, that we have assumed the electron to be in a "rest" frame (that is, a frame travelling in a straight line) when that frame is, in fact, rotating. There is an additional correction to vector quantities that arises from the rotation of the frame. This correction, in macroscopic systems, gives rise to things like coriolis force.

Let us recall (from classical mechanics) that if a coordinate system rotates at some angular velocity $\boldsymbol{\omega}$, the *total* rate of change of any vector quantity is given by

$$\left(\frac{d\mathbf{G}}{dt}\right)_{\text{non-rot}} = \left(\frac{d\mathbf{G}}{dt}\right)_{\text{rest frame}} + \boldsymbol{\omega} \times \mathbf{G}. \tag{16.112}$$

This is a geometric relation that says that a vector in a non–rotating frame is related to the same vector expressed in a (rotating) "rest" frame by adding its time rate of change in *direction* resulting from the rotation of the frame. A moment of quiet reflection should convince you that this should have the magnitude

$$G\frac{d\theta}{dt}$$

and should be perpendicular to $\boldsymbol{\omega}$ and $\mathbf{G}$. This just adds the rotation *of* the frame to the vector *in* the frame to get the vector in a non–rotated frame.

Well, as I noted above, the expression we have given above for the time rate of change of the spin was correct for the field and moment *expressed in the rest frame of the electron.* In the *lab* (non–rotating) frame, which is where we measure its energy, we therefore should have:

$$\left(\frac{d\mathbf{s}}{dt}\right)_{\mathrm{non-rot}} = \mathbf{s} \times \left(\frac{ge\mathbf{B}'}{2mc} - \boldsymbol{\omega}_T\right) \tag{16.113}$$

where ω_T is the angular velocity of the precession of the frames. This adds a $(\mathbf{s} \cdot \boldsymbol{\omega}_T)$ correction to the interaction energy:

$$U = -\frac{ge}{2mc}\mathbf{s} \cdot \mathbf{B} + \frac{g}{2m^2c^2}(\mathbf{s} \cdot \mathbf{L})\frac{1}{r}\frac{dV}{dr} + (\mathbf{s} \cdot \boldsymbol{\omega}_T). \tag{16.114}$$

U is thus the *laboratory* potential energy of interaction. What, then, is the correct value of $\boldsymbol{\omega}_T$?

To answer that we must consider carefully what defines the "rest" frame of the *accelerating* electron. We will do so by chopping the motion of the electron into infinitesimal segments. If the electron is moving at velocity $\mathbf{v}(t) = c\boldsymbol{\beta}$ at any instant of time t, then at $t + \delta t$ the electron is moving at $\mathbf{v}(t) = c(\boldsymbol{\beta} + \delta\boldsymbol{\beta})$. To get from the lab frame (x) to the instantaneous rest frame of the electron (x') we must therefore boost:

$$x' = A(\boldsymbol{\beta})x \tag{16.115}$$

(at t) or

$$x'' = A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})x \tag{16.116}$$

(at $t + \delta t$). Note that for *each* of these transformations, there is no rotation, just the boost.

The coordinate frame precession is going to be determined by the Lorentz transformation between these two (infinitesimally separated) results:

$$x'' = A_T x' \tag{16.117}$$

where (as I hope is obvious)

$$A_T = A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})A^{-1}(\boldsymbol{\beta}) = A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})A(-\boldsymbol{\beta}). \tag{16.118}$$

To evaluate this (in the limit of vanishing δt) we will pick an initial $\boldsymbol{\beta}$ along the 1 direction and add to it $\delta\boldsymbol{\beta}$ in the 1–2 plane. Clearly this is general, for a suitable initial orientation of the coordinate system.

Then

$$A(-\boldsymbol{\beta}) = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{16.119}$$

and (keeping only first order terms in $\delta\boldsymbol{\beta}$)

$$A(\boldsymbol{\beta} + \delta\boldsymbol{\beta}) = \begin{pmatrix} \gamma + \gamma^3\beta\delta\beta_1 & -(\gamma\beta + \gamma^3\delta\beta_1) & -\gamma\delta\beta_2 & 0 \\ -(\gamma\beta + \gamma^3\delta\beta_1) & \gamma + \gamma^3\beta\delta\beta_1 & \frac{(\gamma-1)}{\beta}\delta\beta_2 & 0 \\ -\gamma\delta\beta_2 & \frac{(\gamma-1)}{\beta}\delta\beta_2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{16.120}$$

We multiply these matrices together to obtain:

$$A(\boldsymbol{\beta} + \delta\boldsymbol{\beta}) = \begin{pmatrix} 1 & -\gamma^2\delta\beta_1 & -\gamma\delta\beta_2 & 0 \\ -\gamma^2\delta\beta_1 & 1 & \frac{(\gamma-1)}{\beta}\delta\beta_2 & 0 \\ -\gamma\delta\beta_2 & \frac{(\gamma-1)}{\beta}\delta\beta_2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{16.121}$$

(Note that the action of $A(-\boldsymbol{\beta})$ is only in the upper left corner). Finally, if we decompose this in terms of the S and K matrices, we get:

$$A_T = I - \left(\frac{\gamma - 1}{\beta^2}\right)(\boldsymbol{\beta} \times \delta\boldsymbol{\beta}) \cdot \mathbf{S} - (\gamma^2\delta\boldsymbol{\beta}_\parallel + \gamma\delta\boldsymbol{\beta}_\perp) \cdot \mathbf{K} \tag{16.122}$$

where $\delta\boldsymbol{\beta}_\parallel$ and $\delta\boldsymbol{\beta}_\perp$ are the components of $\delta\boldsymbol{\beta}$ parallel to and perpendicular to $\boldsymbol{\beta}$, respectively.

To first order in $\delta\boldsymbol{\beta}$, we see that the total transformation A_T is equivalent to a boost and a rotation:

$$A_T = A(\Delta\boldsymbol{\beta})R(\Delta\boldsymbol{\Omega}) \tag{16.123}$$

which can be performed in either order (because they are "infinitesimal" and hence commute to first order. In this expression,

$$A(\Delta\boldsymbol{\beta}) = I - \Delta\boldsymbol{\beta} \cdot \mathbf{K} \tag{16.124}$$

and

$$R(\Delta\boldsymbol{\Omega}) = I - \Delta\boldsymbol{\Omega} \cdot \mathbf{S}. \tag{16.125}$$

Obviously,

$$\Delta\boldsymbol{\beta} = \gamma^2 \delta\boldsymbol{\beta}_{\parallel} + \gamma\delta\boldsymbol{\beta}_{\perp} \tag{16.126}$$

and

$$\Delta\boldsymbol{\Omega} = \left(\frac{\gamma - 1}{\beta^2}\right)(\boldsymbol{\beta} \times \delta\boldsymbol{\beta}) = \frac{\gamma^2}{\gamma + 1}\boldsymbol{\beta} \times \delta\boldsymbol{\beta}. \tag{16.127}$$

Finally we see explicitly that at least for infinitesimal transformations, a pure Lorentz boost $A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})$ is equivalent to a boost to an infinitesimally differing frame $A(\boldsymbol{\beta})$ followed by a simultaneous infinitesimal boost *and* rotation.

Now comes the tricky part. The equation of motion for the spin that we began with (in the "rest frame") can be expected to hold provided that the evolution of the rest frame is described by a series of infinitesimal boosts alone (*without* rotations). In other words, we have to add the relativistic equivalent of counterrotating the frames (like we did above with the $\boldsymbol{\omega}_T \times \mathbf{G}$ term). These "relativistically nonrotating coordinates" are related to the instantaneous rest frame coordinates of the electron by the infinitesimal boost

$$x''' = A(\Delta\boldsymbol{\beta})\left\{A(\boldsymbol{\beta})x = x'\right\} \tag{16.128}$$

alone. In terms of the lab coordinates,

$$x''' = R(-\Delta\boldsymbol{\Omega})A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})x. \tag{16.129}$$

Thus the "rest" system of coordinates of the electron are defined by x'''. They are rotated by $-\Delta\boldsymbol{\Omega}$ relative to the boosted laboratory axes x''. If a physical vector $\mathbf{G}$ has a (proper) time rate of change of $d\mathbf{G}/d\tau$ in the rest frame, the precession of the rest frame axes with respect to the laboratory makes the total time rate of change

$$\left(\frac{d\mathbf{G}}{dt}\right)_{\rm non-rot} = \left(\frac{d\mathbf{G}}{dt}\right)_{\rm rest\ frame} + \boldsymbol{\omega} \times \mathbf{G} \tag{16.130}$$

as before with

$$\boldsymbol{\omega}_T = \lim_{\delta t \to 0} \frac{\Delta \boldsymbol{\Omega}}{\delta t} = \frac{\gamma^2}{\gamma+1} \frac{\mathbf{a} \times \mathbf{v}}{c^2} \tag{16.131}$$

(Recall that the connection to laboratory time is $d\mathbf{G}/dt = \gamma^{-1} d\mathbf{G}/d\tau$ in the rest frame itself).

The acceleration perpendicular to the instantaneous velocity appears in this expression because it is this quantity that produces the "rotation" in the infinitesimal transformation between frames that occured in the infinitesimal time interval. Note that this is a purely kinematical effect, and has nothing to do with the laws of nature, just like the non-relativistic "coriolis force" and "centrifugal force". If one wishes to relate the laws of nature as measured in some accelerating frame to those measured in a non-accelerating frame, then it is necessary to insert a fictitious "force" (or more properly interaction "energy") that is kinematic in origin.

In this case, curiously enough, the laws of nature are known in the accelerating frame, and the fictitious force appears in the lab frame, where it is not properly speaking fictitious. However, it *is* still kinematic. That is, there is no actual energy associated with the fictitious interaction (whatever *that* means); however, this interaction is necessary nonetheless if we wish to obtain the equation of motion from the energy equation alone without explicit consideration of the transformations of frames.

To conclude, for electrons the acceleration is caused by the (screened) Coulomb force on the electron that keeps it bound. Thus

$$\boldsymbol{\omega}_T = \frac{-1}{2c^3} \frac{\mathbf{r} \times \mathbf{v}}{m} \frac{1}{r} \frac{dV}{dr} = \frac{-1}{2m^2c^2} \mathbf{L} \frac{1}{r} \frac{dV}{dr}. \tag{16.132}$$

This has exactly the same form as the "rest frame" spin orbit interaction with half the magnitude and the opposite sign. It beautifully cancels the extra factor of 2. The final result is:

$$U' = -\frac{ge}{2mc} \mathbf{s} \cdot \mathbf{B} + \frac{(g-1)}{2m^2c^2} (\mathbf{s} \cdot \mathbf{L}) \frac{1}{r} \frac{dV}{dr}. \tag{16.133}$$

With $g = 2$, *both* the spin–orbit interaction *and* the anomalous Zeeman effect are correctly predicted in accord with what is experimentally observed. Relativistic effects, which are generally thought of as being "subtle", are not subtle at all when it comes to kinematics. The relativistic kinematic correction is as large as the other quantities naturally present independent of the particular orbit or speed of the electron.

This effect is even more pronounced in atomic nuclei. There the electromagnetic forces are much weaker than the binding nuclear forces, and can be neglected to lowest order. However, even uncharged neutrons experience a spin–orbit interaction

$$U_{rmN} = -\frac{1}{2M^2c^2}\mathbf{s}\cdot\mathbf{L}\frac{1}{r}\frac{dV_{\rm N}}{dr} \tag{16.134}$$

that is now *purely* kinematic and has nothing whatsoever to do with the electromagnetic force! There will be a small electromagnetic correction to this for protons. This simple prediction is in reasonable agreement with what is observed in many nucleii for simple models for $V_{\rm N}$. Unfortunately, the system is actually so complicated that this simple minded, single particle description itself is not really valid.

This is just a drop in the proverbial bucket of accelerated systems. Clearly, accelerated, relativistic systems have a much more involved structure than that described by the Lorentz transformations alone. This becomes even more so when Einstein's revered equivalence principal is invoked, so that gravitational force and "real" acceleration are not (locally) distinguishable. But that is general relativity and far beyond the scope of this course.

16.6 Covariant Formulation of Electrodynamics

We are now ready to get serious about electrodynamics. We have developed a beautiful, geometric system for describing the *coordinates* in terms of which electrodynamics must be formulated for the speed of light to be an invariant. We have developed a group of coordinate transformations that preserves that invariance. Now we must face up to the fact that our original equations of electrodynamics are not in a "covariant" formulation that makes these constraints and transformation properties manifest. For example, we do not yet know how the electric and magnetic fields themselves transform under a LT!

Let us then reformulate our basic equations in 4–tensor form. We will make the equations themselves 4–scalars, 4–vectors, or 4–tensors of higher rank so that we can simply look at them and deduce their transformation properties. In addition, we will simplify the notation when possible.

We begin at the beginning. All we really know about electromagnetic fields is their (defined) action on a charged particle:

$$\frac{d\mathbf{p}}{dt} = q\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \tag{16.135}$$

(in 3–vector notation). Well, we know that the 3–vector momentum is just part of a 4–vector momentum:

$$p^\alpha = (p^0, \mathbf{p}) = m(U^0, \mathbf{U}) \tag{16.136}$$

(where $p^0 = E/c$). Also, we don't know what "t" is (since that depends on the choice of frame) so we need to use "τ" instead in our definition.

Thus we can write

$$\frac{d\mathbf{p}}{d\tau} = \frac{q}{c}\left(U^0\mathbf{E} + \mathbf{U} \times \mathbf{B}\right). \tag{16.137}$$

The left hand side tells us the rate of change of the (spatial) momentum, which is just part of a four vector. The time component should tell us how the energy changes with proper time:

$$\frac{dp^0}{d\tau} = \frac{q}{c}\mathbf{U} \cdot \mathbf{E}. \tag{16.138}$$

Now, if this energy–force 4–vector equation is to be covariant (so its transformed form is still a 4–vector) then the right hand sides must form a 4–vector too. Thus we must be able to express it (as a contraction of co and contra variant tensors) so that this property is "manifest". We know (experimentally) that charge is a Lorentz scalar; that is, charge is invariant under LT's. $(U^0, \mathbf{U})$ forms a contravariant 4–vector.

From this we can deduce the 4–tensor form for the electromagnetic field! Since the space parts $\mathbf{U} \cdot \mathbf{E}$ form the time component of a four vector, $\mathbf{E}$ must be the time–space part of a tensor of rank two. That is,

$$\mathbf{E} \cdot \mathbf{U} = F^{0\beta}U_\beta. \tag{16.139}$$

We could easily find **B** in a similar fashion and could eventually work out the electromagnetic field strength tensor. However, it is more constructive to keep on making four vectors, etc. out of the rest of the relations at hand.

For example, we already have observed that the continuity equation is a covariant 4–scalar:

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{J} = 0. \tag{16.140}$$

To make it's covariance manifest, we define a 4–current

$$J^\alpha = (c\rho, \mathbf{J}) \tag{16.141}$$

so that

$$\partial_\alpha J^\alpha = 0 \tag{16.142}$$

is the continuity equation. Note that (as Jackson remarks) this only works because electric charge is a Lorentz invariant and so is a *four*–dimensional volume element (since $\det A = +1$).

Next, consider the wave equations for the potentials in the Lorentz gauge (note well that Jackson for no obvious reason I can see *still* uses Gaussian units in this part of chapter 11, which is goiing to make this a pain to convert below – bear with me):

$$\begin{aligned} \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} - \nabla^2 \phi &= \frac{\rho}{\epsilon_0} = \frac{J^0}{\epsilon_0 c} \\ &= \mu_0 \frac{J^0}{\mu_0 \epsilon_0 c} = \mu_0 (c J^0) \qquad (16.143) \\ & \qquad (16.144) \end{aligned}$$

so that:

$$\frac{1}{c^2}\frac{\partial^2 (\phi/c)}{\partial t^2} - \nabla^2 (\phi/c) = \mu_0 J^0 \tag{16.145}$$

$$\frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}}{\partial t^2} - \nabla^2 \boldsymbol{A} = \mu_0 \boldsymbol{J} \tag{16.146}$$

Therefore, if we form the 4–vector potential

$$A^\alpha = (\frac{\phi}{c}, \mathbf{A}) \tag{16.147}$$

then we can write the various 4–relations:

$$\partial_\alpha A^\alpha = \frac{1}{c}\frac{\partial A^0}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{A} \tag{16.148}$$

(which is the 4–scalar Lorentz gauge condition)

$$\Box A^\alpha = \partial_\beta \partial^\beta A^\alpha = \mu_0 J^\alpha \tag{16.149}$$

(the 4–vector inhomogeneous electromagnetic wave equation constructed from the 4–scalar D'Lambertian wave operator – the set of four wave equations for ϕ and the components of $\boldsymbol{A}$ above).

Now we can construct the components of **E** and **B** from the covariant 4–vector potential. For example, we know that:

$$\boldsymbol{E} = -\boldsymbol{\nabla}\phi - \frac{\partial \boldsymbol{A}}{\partial t} \tag{16.150}$$

where $\phi = cA^0$, so

$$E_x = -c\frac{\partial A_x}{\partial(ct)} - \frac{\partial c A^0}{\partial x} = -c(\partial^0 A^1 - \partial^1 A^0) \tag{16.151}$$

and similarly, since $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$:

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = -(\partial^2 A^3 - \partial^3 A^2) \tag{16.152}$$

etc.

The components of the electric and magnetic fields (all six of them) thus transform like the components of a **second rank, antisymmetric, traceless field strength tensor**[7] :

$$F^{\alpha\beta} = \partial^\alpha A^\beta - \partial^\beta A^\alpha. \tag{16.153}$$

In explicit component form,

$$F^{\alpha\beta} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}. \tag{16.154}$$

[7]Wikipedia: http://www.wikipedia.org/wiki/Electromagnetic tensor. Note that I'm not completely comfortable with the signs for the covariant form of the potential in the Wikipedia article, although its main conclusions are sound enough.

The tensor with two covariant indices (formed by two contractions with g) is obtained by replacing $\boldsymbol{E}$ with $-\boldsymbol{E}$.

$$F_{\alpha\beta} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix}. \tag{16.155}$$

Another important version of this tensor is the **dual field strength tensor** $\mathcal{F}^{\alpha\beta}$. In terms of the totally antisymmetric tensor of the fourth rank and the normal field strength tensor it is given by:

$$\mathcal{F}^{\alpha\beta} = \frac{1}{2}\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z/c & -E_y/c \\ B_y & -E_z/c & 0 & E_x/c \\ B_z & E_y/c & -E_x/c & 0 \end{pmatrix}. \tag{16.156}$$

This is obtained from the basic contravariant field strength tensor by the substitutions $\boldsymbol{E} \to \boldsymbol{B}, \boldsymbol{B} \to -\boldsymbol{E}$. Consideration of the section on magnetic monopoles shows that this is indeed a particular duality transformation obtained in free space with the "rotation" parameter equal to $\pi/2$ (in J6.151).

Finally, we must write Maxwell's equations in covariant form. The inhomogeneous equations are (recall)

$$\boldsymbol{\nabla}\cdot\boldsymbol{E} = \frac{\rho}{\epsilon_0} \tag{16.157}$$

$$\boldsymbol{\nabla}\times\boldsymbol{B} - \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} = \mu_0 \boldsymbol{J} \tag{16.158}$$

The quantity on the right is proportional to the four current. The quantity on the left must therefore contract a 4–derivative with the field strength tensor. You should verify that

$$\partial_\alpha F^{\alpha\beta} = \mu_0 J^\beta \tag{16.159}$$

exactly reconstructs the inhomogeneous equation for each component of J^β.

The homogeneous equations

$$\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0 \tag{16.160}$$

$$\boldsymbol{\nabla}\times\boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \tag{16.161}$$

also form a four vector (of zero's) and must hence be the contraction of a field strength tensor. But which one? Well, the second homogeneous equation requires that $\mathbf{B} \to -\mathbf{E}$ and both require that $\mathbf{E} \to \mathbf{B}$, so it must be the dual:

$$\partial_\alpha \mathcal{F}^{\alpha\beta} = 0. \tag{16.162}$$

If we feel compelled to write everything in terms of the field strength tensor itself, this can be done. The result is the four equations

$$\partial^\alpha F^{\beta\gamma} + \partial^\beta F^{\gamma\alpha} + \partial^\gamma F^{\alpha\beta} = 0 \tag{16.163}$$

where α, β, γ are any three of the four indices 0,1,2,3. However, this equation is a third rank tensor on the left, and its reduction by symmetry to a tensor of first rank is not manifest. It is ugly, if you will.

Now that we have written Maxwell's equations (and the consequences of ME) in four dimensional form (remarking all the while that they are unusually beautiful and concise in this notation) we are done. Before we go on to deduce (from these results) how electric and magnetic fields LT, however, we should complete the question with which we began the discussion, namely, how does Newton's law become covariant? The answer is (now that we know what the field strength tensor is)

$$\frac{dp^\alpha}{d\tau} = m\frac{dU^\alpha}{d\tau} = \frac{q}{c}F^{\alpha\beta}U_\beta. \tag{16.164}$$

The time–component equation is just the work–energy theorem, and the space equations are Newton's law.

As a postscript to our discussion (recalling that sometimes the fields propagate in some medium and not free space) we note that in this case the homogeneous equation(s) remain unchanged, but the inhomgeneous equations are modified (by using $\boldsymbol{H}$ and $\boldsymbol{D}$ instead of $\boldsymbol{B}$ and $\boldsymbol{E}$). The inhomogeneous equation is then

$$\partial_\alpha G^{\alpha\beta} = \mu J^\beta \tag{16.165}$$

where hopefully the definition of $G^{\alpha\beta}$ is obvious (that is, substitute $v = \sqrt{1/\epsilon\mu}$ for c throughout in appropriate places, or if you prefer recapitulate the entire derivation using $\boldsymbol{H}$ and $\boldsymbol{D}$ from the beginning).

Let us pause for a moment of religious silence and contemplate a great wonder of nature. This is the scientist's version of "prayer in school".

16.7 The Transformation of Electromagnetic Fields

Now that we have this in hand, we can easily see how to transform the electric and magnetic fields when we boost a frame. Of course, that does not guarantee that the result will be simple.

To convert $F^{\alpha\beta}$ from K to K', we must contract its indices with the transformation tensors,

$$F'^{\alpha\beta} = \frac{\partial x'^{\alpha}}{\partial x^{\gamma}} \frac{\partial x'^{\beta}}{\partial x^{\delta}} F^{\gamma\delta}. \tag{16.166}$$

Note that since A is a *linear* transformation:

$$A^{\alpha}{}_{\gamma} = \frac{\partial x'^{\alpha}}{\partial x^{\gamma}} \tag{16.167}$$

(where I have deliberately inserted a space to differentiate the first index from the second) we can write this in terms of the components of A as:

$$\begin{aligned} F'^{\alpha\beta} &= A^{\alpha}{}_{\gamma} F^{\gamma\delta} A_{\delta}{}^{\beta} \\ &= A^{\alpha}{}_{\gamma} F^{\gamma\delta} \tilde{A}^{\beta}{}_{\delta} \end{aligned} \tag{16.168}$$

or (in a compressed notation):

$$F' = AF\tilde{A} \tag{16.169}$$

This is just a specific case of the general rule that A can be used in general to transform any nth rank tensor by contracting it appropriately with each index.

As we saw in our discussion of Thomas precession, we will have occasion to use this result for the particular case of a pure boost in an arbitrary direction that we can without loss of generality pick to be the 1 direction. Let's see how this goes. Recall that A for a pure boost in the one direction is the matrix formed with a lower right quadrant identity and an upper left quadrant 2×2 with γ on the diagonal and $-\gamma\beta$ on the corners). Thus: so:

$$\begin{aligned} F'^{01} &= A^{0}{}_{0} F^{01} A_{1}{}^{1} + A^{0}{}_{1} F^{10} A_{1}{}^{1} \\ -\frac{E_1'}{c} &= -\gamma^2 \frac{E_1}{c} - \gamma^2\beta^2 \frac{E_1}{c} \\ E_1' &= (\gamma^2 + \gamma^2\beta^2) E_1 \\ E_1' &= E_1 \end{aligned} \tag{16.170}$$

Note that we have extracted the ordinary cartesian components of $\boldsymbol{E}$ and $\boldsymbol{B}$ from F after transforming it. I leave the rest of them to work out yourself. You should be able to show that:

$$E'_1 = E_1 \tag{16.171}$$
$$E'_2 = \gamma(E_2 - \beta B_3) \tag{16.172}$$
$$E'_3 = \gamma(E_3 + \beta B_2) \tag{16.173}$$
$$B'_1 = B_1 \tag{16.174}$$
$$B'_2 = \gamma(B_2 + \beta E_3) \tag{16.175}$$
$$B'_3 = \gamma(B_3 - \beta E_2) \tag{16.176}$$

The component of the fields in the direction of the boost is unchanged, the perpendicular components of the field are mixed (almost as if they were space–time pieces) by the boost. If you use instead the general form of A for a boost and express the components in terms of dot products, you should *also* show that the general transformation is given by:

$$\mathbf{E}' = \gamma(\mathbf{E} + \boldsymbol{\beta} \times \mathbf{B}) - \frac{\gamma^2}{\gamma+1}\boldsymbol{\beta}(\boldsymbol{\beta} \cdot \mathbf{E}) \tag{16.177}$$
$$\mathbf{B}' = \gamma(\mathbf{B} - \boldsymbol{\beta} \times \mathbf{E}) - \frac{\gamma^2}{\gamma+1}\boldsymbol{\beta}(\boldsymbol{\beta} \cdot \mathbf{B}). \tag{16.178}$$

A purely electric or magnetic field in one frame will thus be a mixture of electric and magnetic fields in another. We see that truly, there is little reason to distinguish them. We have to be a little careful, of course. If there is a monopolar (static) electric field in any frame, we cannot transform it completely into a magnetostatic field in another, for example. Why? Because the equations above will lead to some mixture for all $\beta < 1$, and $\beta < 1$ in nature as a constraint.

I encourage you to review the example given in Jackson and meditate upon the remarks therein. We will not spend valuable class time on this, however. Instead we will end this, after all, purely mathematical/geometrical *kinematical* interlude (no Hamiltonians or Lagrangians = no physics) and do some *physics*. Let us deduce the covariant dynamics of relativistic particles in (assumed fixed) electromagnetic fields.

Chapter 17

Relativistic Dynamics

17.1 Covariant Field Theory

We are interested in deducing the dynamics of point charged particles in "given" (i. e. — fixed) electromagnetic fields. We already "know" the answer, it is given by the covariant form of Newton's law, that is:

$$\frac{dp^\alpha}{d\tau} = m\frac{dU^\alpha}{d\tau} = \frac{q}{c}F^{\alpha\beta}U_\beta. \tag{17.1}$$

From this we can find the 4–acceleration,

$$\frac{dU^\alpha}{d\tau} = \frac{q}{mc}F^{\alpha\beta}U_\beta \tag{17.2}$$

which we can integrate (in principle) to find the 4–trajectory of the particle in question.

However, this is not useful to us. *Real* physicists don't use Newton's law anymore. This is nothing against Newton, it is just that we need Hamilton's or Lagrange's formulation of dynamics in order to construct a quantum theory (or even an elegant classical theory). Our first chore, therefore, will be to generalize the arguments that lead to the Euler–Lagrange or Hamilton equations of motion to four dimensions.

17.1.1 The Brute Force Way

Recall that the Lagrangian path to the dynamics of a particle (which is most easily made covariant, since it uses $(q(t), \dot{q}(t), t)$ as its variables) is based on the **Action**

$$A = \int_{t_0}^{t_1} L(q(t), \dot{q}(t), t)dt. \tag{17.3}$$

By requiring that A be an extremum as a *functional* of the system trajectory, we obtain the **Euler–Lagrange equations**

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i}\right) = 0. \tag{17.4}$$

These are equivalent to Newton's law for suitable definitions of L and the force.

The simplest way to make this relativistic is to express it in terms of the proper time and then require that the action A be extremal in all frames. Then,

$$A = \int_{\tau_0}^{\tau_1} \gamma L d\tau \tag{17.5}$$

is the action, since $dt = \gamma d\tau$. We now must remark

1. If the extremal condition is to be invariant with respect to LT's, then A must be invariant (and hence a 4–scalar);
2. Therefore (since $d\tau$ is invariant) γL must be invariant;
3. Finally, since γ is just a number, L must be a 4–scalar.

This final conclusion greatly constrains the possible forms of L.

Note well that it is not clear that this argument (from Jackson) is valid. γ, while a number, is not invariant under boosts – indeed, it is parametrically related to the boost parameter! It is also perfectly clear that the first statement is false – while it is true that if A is a four-scalar with respect to boosts that its extremums must be preserved, it is equally true that this condition is not necessarily unique – all that is *necessary* is that a boost monotonically scale the action in such a way that the extremal property is preserved!

A weaker (but still sufficient) argument might then be:

> *If* L is a 4-scalar, and γ is a monotonic function independent of the 4-coordinates, 4-velocities, and τ, then the property of a given trajectory resulting in an extremum of the action is preserved.

In my opinion this is clearer and still adequate for our purposes. L being a 4-scalar (0th rank tensor in 4-space w.r.t. the Lorentz transformation) is *sufficient* to produce an invariant extremum of the action A, even if the *numerical values* of A vary under a boost. To prove that it is also *necessary* very likely involves an exercise in the calculus of variations that distributes the derivatives over γL – similar exercises are already in the homework for this chapter in other contexts.

Either way, we will now assert that the Lagrangian of a free particle must be a 4-scalar (and hence must be formed out of the full contraction of tensors of higher rank), and will remain alert in the work below for any sort of inconsistency that might be related to γ.

Obviously, we want it to reproduce the classical non–relativistic theory in the appropriate limit, that is, a free particle should have constant energy and momentum, or, equivalently, 4–velocity. The simplest (not "only" as Jackson states) Lorentz invariant function of the 4–velocity is it's quadratic form:

$$U^\alpha U_\alpha = c^2 \tag{17.6}$$

Of course, any polynomial functional of this quadratic is also a possible scalar, but they are not the thing to try first. Thus a reasonable guess for the Lagrangian is

$$L = (\text{constant})c^2\gamma^{-1} = -mc^2\sqrt{1 - \frac{u^2}{c^2}}. \tag{17.7}$$

If we now crunch through the Euler–Lagrange equation we find that this choice for the constant leads to

$$\frac{d}{dt}(\gamma m\mathbf{u}) = 0 \tag{17.8}$$

which is indeed Newton's law for a free particle, but with the *relativistic* form of the three–momentum.

If one chooses a frame where the particle is initially at rest, a trajectory where it remains at rest will yield the least action (you should check this).

This is because γ^{-1} is maximal when $\beta = 0$ (and the Lagrangian has a minus sign).

Now, suppose that the charged particle is in a electromagnet potential. If it were moving slowly in a scalar potential Φ only, its potential energy would be $V = q\Phi$. The non–relativistic Lagrangian in this case should be just $T - V$ (where T is the free particle Lagrangian). The interaction part of the relativistic Lagrangian must therefore reduce to $-q\Phi$ in this non–relativistic limit.

We must find a Lorentz invariant (scalar) form for γL_{int} that reduces to $-q\Phi$ for non–relativistic velocities. Since Φ is the time component of a four vector potential, we can guess that the correct generalization must involve the four vector potential A^α. Since it must be a scalar, it must involve the scalar product of A^α with some four vector(s). The only ones avaliable are x^α and U^α.

The correct γL_{int} depends only on the U^α. If it depended on the coordinates as well, then the physics would not be translationally invariant and the results of our calculation might well depend on where we chose the origin. This does not seem reasonable. Once again, this does not uniquely determine it, it only determines the simplest (linear) form to within a sign and a constant:

$$\gamma L_{\text{int}} = -\frac{q}{c} U_\alpha A^\alpha \tag{17.9}$$

or

$$L_{\text{int}} = -q\Phi + \frac{q}{c}\vec{u}\cdot\vec{A}. \tag{17.10}$$

There could be additional terms involving polynomials of this quantity, the product $A^\alpha A_\alpha$ (which is indeed present in some theories) and other scalar reductions of field and charge/field tensor quantities. Linearity, in either the vector potential or the velocity, is an axiom and not logically necessary.

The complete relativistic Lagrangian for a charged particle is thus

$$L = -mc^2\sqrt{1 - \frac{u^2}{c^2}} + \frac{q}{c}\vec{u}\cdot\vec{A} - q\Phi. \tag{17.11}$$

It should take you about one hour to show that this yields the correct relativistic Lorentz force law. The free particle part is obvious, the electric field is obvious. You will have to work a bit, using

$$\frac{d}{dt} = \left(\frac{\partial}{\partial t} + \vec{u}\cdot\boldsymbol{\nabla}\right) \tag{17.12}$$

to squeeze $-\vec{u} \times (\boldsymbol{\nabla} \times \vec{A})$ out of the remainder. I suggest that you simply work out the terms by expanding them as far as they go and reassembling the pieces, but some of you may know enough vector algebra to do it better ways. This will be on the next assignment, so feel free to start.

The canonical momentum $\vec{P}$ conjugate to the position coordinates x is obtained (as usual) from

$$P_i = \frac{\partial L}{\partial u_i} = \gamma m u_i + \frac{q}{c} A_i. \tag{17.13}$$

This result,

$$\vec{P} = \vec{p} + \frac{q}{c}\vec{A} \tag{17.14}$$

(where $\vec{p}$ is the relativistic kinetic momentum of the particle) is extremely important to remember, as it is a necessary ingredient in the construction of either a quantum theory or an elegant classical theory. Placing the particle in a field alters its canonical "momentum".

We make the Hamiltonian out of the Lagrangian and the canonical momentum via

$$H = \vec{P} \cdot \vec{u} - L \tag{17.15}$$

The basic result here has too many variables. We must eliminate $\vec{u}$ in favor of $\vec{A}$ and $\vec{P}$. Note that

$$\vec{u} = \frac{c\vec{P} - q\vec{A}}{\sqrt{\left(\vec{P} - \frac{q}{c}\vec{A}\right)^2 + m^2c^2}} \tag{17.16}$$

(something that is a wee chore to prove, of course, but it is straightforward algebra). With even more tedious algebra, you can show that the Hamiltonian is:

$$H = \sqrt{(c\vec{P} - q\vec{A})^2 + m^2c^4} + q\Phi = W. \tag{17.17}$$

From this result, Hamilton's equations of motion should reproduce the Lorentz force law. See that it does (although the relationship between the EL equations and Hamilton's equations makes the result obvious). Note that if we interpret the Hamiltonian (as usual) as the total energy W of the particle, this result is related to the free particle energy by $\vec{p} \to (\vec{P} - \frac{q}{c}\vec{A})$ and the addition of the scalar potential energy $q\Phi$. This is actually just a single change in the four–vector momentum:

$$(W - q\Phi)^2 - (c\vec{P} - q\vec{A})^2 = m^2c^4 = p^\alpha p_\alpha \tag{17.18}$$

(which has the usual form if

$$p^\alpha = \left(\frac{E}{c}, \vec{p}\right) = \left(\frac{1}{c}(W - e\Phi), \vec{P} - \frac{q}{c}\vec{A}\right) \tag{17.19}$$

). This also makes the invariance properties of the Hamiltonian manifest.

It is really annoying to obtain the invariance properties of things after the fact. It is also annoying (although perhaps useful) to have the three vector coordinates $\vec{x}, \vec{u}$ hanging around at this point. So let us rederive these results using only four–vectors and suitable scalar reductions.

17.1.2 The Elegant Way

We can write the free particle Lagrangian using only scalar reductions of suitable 4–vectors:

$$L_{\text{free}} = -\frac{mc}{\gamma}\sqrt{U_\alpha U^\alpha} \tag{17.20}$$

(which is still $-mc^2\gamma^{-1}$). The action is thus

$$A = -mc\int_{\tau_0}^{\tau_1} \sqrt{U_\alpha U^\alpha}\, d\tau. \tag{17.21}$$

The variations on this action must be carried out subject to the constraint

$$U_\alpha U^\alpha = c^2 \tag{17.22}$$

which severely limits the allowed solutions. We write this as

$$\begin{aligned}
\frac{d(U_\alpha U^\alpha)}{d\tau} &= 0 \\
\frac{dU_\alpha}{d\tau}U^\alpha + U_\alpha\frac{dU^\alpha}{d\tau} &= 0 \\
\frac{dU_\alpha}{d\tau}g^{\alpha\beta}g_{\beta\alpha}U^\alpha + U_\alpha\frac{dU^\alpha}{d\tau} &= 0 \\
U_\beta\frac{dU^\beta}{d\tau} + U_\alpha\frac{dU^\alpha}{d\tau} &= 0 \\
2U_\alpha\frac{dU^\alpha}{d\tau} &= 0 \\
U_\alpha\frac{dU^\alpha}{d\tau} &= 0
\end{aligned} \tag{17.23}$$

Now,

$$\sqrt{U_\alpha U^\alpha} d\tau = \sqrt{\frac{dx_\alpha}{d\tau}\frac{dx^\alpha}{d\tau}} d\tau = \sqrt{g^{\alpha\beta} dx_\alpha dx_\beta} \tag{17.24}$$

which is an infinitesimal length in four–space. The latter expression does not explicitly contain $d\tau$. We can thus parameterize the action in terms of a path–parameter s that increases monotonically with τ but is otherwise arbitrary. Then

$$A = -mc \int_{s_0}^{s_1} \sqrt{g^{\alpha\beta} \frac{dx_\alpha}{ds}\frac{dx_\beta}{ds}} ds. \tag{17.25}$$

We are clearly making progress. We have taken a perfectly good expression and made in unrecognizable. To make you a little happier, note that this has now got the form of

$$A = \int \tilde{L} ds \tag{17.26}$$

where $\tilde{L}$ is a scalar "Lagrangian" written in terms of an *independent* free parameter. This might be progress after all, since we have quashed the annoying γ^{-1}.

If we now do the calculus of variations thing and get the Euler-Lagrange equations in *four* dimensions:

$$\frac{d}{ds}\left(\frac{d\tilde{L}}{\partial\left(\frac{dx_\alpha}{ds}\right)}\right) - \partial^\alpha \tilde{L} = 0 \tag{17.27}$$

(for $\alpha = 0, 4$). Applying them to the Langrangian in this action, they turn out to be:

$$\begin{aligned}
mc\frac{d}{ds}\frac{\partial\left\{g^{\delta\beta}\frac{dx_\beta}{ds}\frac{dx_\delta}{ds}\right\}^{\frac{1}{2}}}{\partial\left(\frac{dx_\alpha}{ds}\right)} &= 0 && (17.28)\\
\frac{mc}{2}\frac{d}{ds}\left\{\frac{\frac{dx^\alpha}{ds} + \frac{dx^\alpha}{ds}}{\sqrt{\frac{dx_\beta}{ds}\frac{dx^\beta}{ds}}}\right\} &= 0 && (17.29)\\
mc\frac{d}{ds}\left\{\frac{\frac{dx^\alpha}{ds}}{\sqrt{\frac{dx_\beta}{ds}\frac{dx^\beta}{ds}}}\right\} &= 0. && (17.30)
\end{aligned}$$

This still does not have the constraint above imposed on it. We impose the constraint by identifying ds with $d\tau$ in such a way that the constraint is simultaneously satisfied:

$$\begin{aligned} \sqrt{\frac{dx_\alpha}{ds}\frac{dx^\alpha}{ds}}ds &= cd\tau \\ \frac{d}{d\tau} &= \frac{c}{\sqrt{\frac{dx_\alpha}{ds}\frac{dx^\alpha}{ds}}}\frac{d}{ds} \end{aligned} \tag{17.31}$$

(which requires both $ds = d\tau$ and $U_\alpha U^\alpha = c^2$). If you like, this constraint picks out of all possible path parameterizations the one that follows the proper time while keeping the four vector velocity scalar product Lorentz invariant. For free particles this is a lot of work, but it is paid back when we include an interaction.

If we multiply the Euler-Lagrange equation (in terms of s) from the left by:

$$\frac{c}{\sqrt{\frac{dx_\alpha}{ds}\frac{dx^\alpha}{ds}}} \tag{17.32}$$

and use the constraint to convert to τ, the result (for the equation of motion) is:

$$m\frac{c}{\sqrt{\frac{dx_\beta}{ds}\frac{dx^\beta}{ds}}}\frac{d}{ds}\left\{\frac{c}{\sqrt{\frac{dx_\beta}{ds}\frac{dx^\beta}{ds}}}\frac{d}{ds}x^\alpha\right\} = 0 \tag{17.33}$$

or

$$m\frac{d^2x^\alpha}{d\tau^2} = 0 \tag{17.34}$$

which certainly looks it has the right form.

We can include an interaction. Just as before, γL_{int} must be a Lorentz scalar. When we make a parameterized version of the Lagrangian, the part under the integral must be a 4–scalar. The covariant form of the result is (hopefully obviously)

$$A = -\int_{s_0}^{s_1}\left\{mc\sqrt{g^{\delta\beta}\frac{d_\delta}{ds}\frac{dx_\beta}{ds}} + \frac{q}{c}\frac{dx_\beta}{ds}A^\beta\right\}ds. \tag{17.35}$$

The "four Lagrangian" in this equation is

$$\tilde{L} = -\left\{mc\sqrt{g^{\delta\beta}\frac{d_\delta}{ds}\frac{dx_\beta}{ds}} + \frac{q}{c}\frac{dx_\beta}{ds}A^\beta\right\}. \tag{17.36}$$

As before we construct the Euler-Lagrange equation.

$$mc\frac{d}{ds}\left\{\frac{\frac{dx^\alpha}{ds}}{\sqrt{\frac{dx_\beta}{ds}\frac{dx^\beta}{ds}}}+\frac{q}{c}A^\alpha\right\}-\frac{q}{c}\frac{x_\beta}{ds}\partial^\alpha A^\beta \quad = \quad 0 \tag{17.37}$$

$$\tag{17.38}$$

Again we multiply through from the left by

$$\frac{c}{\sqrt{\frac{dx_\alpha}{ds}\frac{dx^\alpha}{ds}}} \tag{17.39}$$

and convert to τ to get:

$$m\frac{d^2x^\alpha}{d\tau^2}+\frac{q}{c}\frac{dA^\alpha}{d\tau}-\frac{q}{c}\frac{dx_\beta}{d\tau}\partial^\alpha A^\beta = 0 \tag{17.40}$$

The derivative $dA^\alpha/d\tau$ is a bit jarring. However, if we express this total derivative in terms of partials we observe that:

$$\frac{dA^\alpha}{d\tau} = \frac{dx_\beta}{d\tau}\frac{\partial}{\partial x_\beta}A^\alpha = \frac{dx_\beta}{d\tau}\partial^\beta A^\alpha \tag{17.41}$$

Substituting, the equation of motion becomes:

$$\frac{d(mU^\alpha)}{d\tau} = m\frac{d^2x^\alpha}{d\tau^2} = \frac{q}{c}(\partial^\alpha A^\beta - \partial^\beta A^\alpha)\frac{dx_\beta}{d\tau} = \frac{q}{c}F^{\alpha\beta}U_\beta. \tag{17.42}$$

which is, as expected, the Lorentz force law in covariant form! How lovely!

To make a Hamiltonian in this notation, we must first make the canonical momentum:

$$P^\alpha = -\frac{\partial\tilde{L}}{\partial\left(\frac{x_\alpha}{ds}\right)} = mU^\alpha + \frac{q}{c}A^\alpha \tag{17.43}$$

which is a covariant version of the *complete* set of interaction equations from the previous section (it does both energy and 3–momentum).

There are several ways to make a Hamiltonian (recall that in general there is what amounts to gauge freedom, minimally the ability to add an arbitrary constant which naturally does not affect the resulting differential equations). One is[1]:

$$\tilde{H} = U_\alpha P^\alpha + \tilde{L} \tag{17.44}$$

[1]Note that I've rearranged this slightly to avoid having to do lots of stuff with g sandwiches below.

Again, we must eliminate:

$$U^\alpha = \frac{1}{m}\left(P^\alpha - \frac{q}{c}A^\alpha\right) \tag{17.45}$$

in favor of P^α, A^α. Thus:

$$\tilde{L} = -mc\sqrt{\frac{1}{m^2}\left(P_\alpha - \frac{q}{c}A_\alpha\right)\left(P^\alpha - \frac{q}{c}A^\alpha\right)} - \frac{q}{mc}\left(P_\alpha - \frac{q}{c}A_\alpha\right)A^\alpha \tag{17.46}$$

and

$$\begin{aligned}\tilde{H} &= \frac{1}{m}\left(P_\alpha - \frac{q}{c}A_\alpha\right)P^\alpha - mc\sqrt{\frac{1}{m^2}\left(P_\alpha - \frac{q}{c}A_\alpha\right)\left(P^\alpha - \frac{q}{c}A^\alpha\right)} \\ &\quad -\frac{q}{mc}\left(P_\alpha - \frac{q}{c}A_\alpha\right)A^\alpha \qquad (17.47) \\ &= \frac{1}{m}\left(P_\alpha - \frac{q}{c}A_\alpha\right)\left(P^\alpha - \frac{q}{c}A^\alpha\right) - c\sqrt{\left(P_\alpha - \frac{q}{c}A_\alpha\right)\left(P^\alpha - \frac{q}{c}A^\alpha\right)} \qquad (17.48)\end{aligned}$$

This Hamiltonian in four dimensions is no longer an energy since it is obviously a 4–scalar and energy transforms like the time–component of a four vector. However, it works. Hamilton's equations (in four dimensions) lead again directly to the relativistic Lorentz force law:

$$\begin{aligned}\frac{dx^\alpha}{d\tau} = \frac{\partial\tilde{H}}{\partial P_\alpha} &= \frac{1}{m}\left(P^\alpha - \frac{q}{c}A^\alpha\right) \qquad (17.49) \\ \frac{dP^\alpha}{d\tau} = -\frac{\partial\tilde{H}}{\partial x_\alpha} = -\partial^\alpha\tilde{H} &= \frac{q}{mc}\left(P_\alpha - \frac{q}{c}A_\alpha\right)\partial^\alpha A^\beta \qquad (17.50)\end{aligned}$$

There is a bit of algebra involved in deriving this result. For example, one has to recognize that:

$$p^\alpha = mU^\alpha = P^\alpha - \frac{q}{c}A^\alpha \tag{17.51}$$

and $p_\alpha p^\alpha = m^2c^2$ and apply this to eliminate unwanted terms *judiciously*, that is *after* differentiation. If you apply it too early (for example at the beginning) you observe the puzzling result that:

$$\begin{aligned}\tilde{H} &= \frac{1}{m}p_\alpha p^\alpha - c\sqrt{p_\alpha p^\alpha} \\ &= \frac{1}{m}m^2c^2 - c\sqrt{m^2c^2} \\ &= mc^2 - mc^2 \\ &= 0 \qquad (17.52)\end{aligned}$$

which leads one to the very Zen conclusion that the cause of all things is Nothing (in four dimensions, yet)!

We are left with a rather mystified feeling that the algebraic hand is quicker than the eye. Somehow an equation whose *four-scalar value* is zero has a functional form, a *structure*, that leads to non–zero, covariant equations of motion. Also (as already remarked) this Hamiltonian is not unique. Obviously one can add an arbitrary four-scalar constant to the equation and get no contribution from the derivatives (just as one can in nonrelativistic classical physics). There are other gauge freedoms – ultimately there several other ways of writing the Hamiltonian associated with the given Lagrangian; all of them yield a constant value that is not the energy when evaluated and yield the correct equations of motion when processed.

Finally there exist what are called *singular Lagrangians* – Lagrangians for which the generalized coordinates do not always map into generalized conjugate variables! Dirac was (unsurprisingly) the first to formally identify this in the context of constrained systems (systems described by a Lagrangian and constraints with Lagrange multipliers for which the Hesse determinant vanishes); Bergmann (at Syracuse) also made major contributions to the formal development of the concept. However the roots of the problem date much further back to e.g. Noether's theorem. I have a couple of papers on this that I've collected from the web, although the idea is also discussed in various monographs and textbooks on mechanics.

It is worth pointing out that there was at one point considerable work being done here at Duke on the idea – N. Mukunda, Max Lohe, (both friends of mine) worked on the idea with Larry Biedenharn (my advisor); Biedenharn also published work with Louck on the subject, and of course Mukunda and Sudarshan's book on classical mechanics remains a "classic". Since Dirac's time the notion that the "right" unified field theory will have certain interesting properties related to this has been batted around.

This points out an ongoing problem in relativistic quantum theories. These theories are generally based on a Hamiltonian, but manifestly covariant Hamiltonians for a given system cannot in general be uniquely derived from first principles as the mapping between velocities and momenta is not always one-to-one. Thus even when a covariant Lagrangian density can be constructed, the associated Hamiltonian is not obvious or necessarily unique.

This is just one (although it is one of the most fundamental) obstacles to overcome when developing a relativistic quantum field theory.

17.2 Motion of a Point Charge in a Static Magnetic Field

Now that we have obtained the various covariant forms of the Lorentz force law, we can easily determine the trajectories of charged particles in various fixed fields. In fact, we could have done this weeks ago (if not years) even without knowing the covariant forms.

In a static magnetic field, the equations of motion are:

$$\frac{dE}{dt} = 0 \tag{17.53}$$
$$\frac{d\vec{p}}{dt} = \frac{q}{c}\vec{v} \times \vec{B} \tag{17.54}$$

for the energy and momentum, respectively (arranged like pieces of a four vector for clarity). Clearly the speed of the particle is constant since the force is perpendicular to the motion and does no work. γ is therefore also constant. Thus

$$\frac{d\vec{v}}{dt} = \vec{v} \times \vec{\omega_B} \tag{17.55}$$

where

$$\vec{\omega_B} = \frac{q\vec{B}}{\gamma mc} = \frac{qc\vec{B}}{E} \tag{17.56}$$

is the gyration or precession (cyclotron) frequency. The motion described by this equation is a circular motion perpendicular to $\vec{B}$ coupled to a uniform motion parallel to $\vec{B}$.

This is too droll for words (and in fact you have probably already taught it to your kids in kiddy physics) but it does yield one important result. The magnitude of the momentum perpendicular to $\vec{B}$ is

$$cp_\perp = qBa \tag{17.57}$$

where a is the radius of the circular helix. From this (in, for example, a bubble chamber, where the track can be photographed) and a knowledge (or guess) as the the charge, the transverse momentum can be measured. Measuring other things (like the rate of change of the curvature of the track) can yield the mass of the particle from a knowledge of its momentum. From these humble traces the entire picture we currently have of the sub–atomic zoo has been built up.

Sections 12.2-12.4 are too simple to waste time on. 12.5-12.6 are interesting but important only to plasma people. 12.7 is redundant of things we will do correctly later. Thus we skip to 12.8, leaving you to read any or all of the intermediate material on your own. We will skip 12.9. Finally, we will do 12.10–12.11 to complete chapter 12.

17.3 Building a Relativistic Field Theory

We have not quite finished the job of building a proper relativistic field theory of electromagnetism. That is because we would like to be able to obtain all of the equations of motion (that is, physics) describing the system from a covariant action principle. We have done that for the particles *in* the fields, but what about the fields themselves? In fact, since the particles produce (and hence modify) the fields, we do not even have the correct solutions for the particles alone, yet. Let us see if we can develop a suitable Lagrangian for the fields that leads, ideally, to Maxwell's equations.

The **Rules** for building a field theory Lagrangian are of interest in and of themselves, since they are quite general. The rules are:

1. Take the position and velocity coordinates for continuous space time and replace them with field variables.
2. Label the field variables with discrete (coordinate direction) labels and with continuous (position) variables.
3. Replace the "velocity" with the 4–gradient.
4. Require the action to be stationary w.r.t. variations in the field variables themselves and their gradients.

That is,

$$i \;\rightarrow\; x^\alpha, k \tag{17.58}$$

$$
\begin{aligned}
q_i &\rightarrow \phi_k(x) && (17.59)\\
\dot{q}_i &\rightarrow \partial^\alpha \phi_k(x) && (17.60)\\
L = \sum_i L_i(q_i, \dot{q}_i) &\rightarrow \int \mathcal{L}(\phi_k, \partial^\alpha \phi_k) d^3x && (17.61)\\
\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i} &\rightarrow \partial^\beta \frac{\partial \mathcal{L}}{\partial \partial^\beta \phi_k} = \frac{\partial \mathcal{L}}{\partial \phi_k}. && (17.62)
\end{aligned}
$$

When we make an action integral, we integrate $\mathcal{L}$ over time, making the total integral four dimensional. We therefore call $\mathcal{L}$ the *Lagrangian density* in four dimensions. Note that the action will be covariant provided the Lagrangian *density* is a 4–scalar. This is what I have meant whenever I have inadvertantly called the "Lagrangian" a scalar. Good, clean, relativistic theories with or without particles are made out of scalar Lagrangian densities, not Lagrangians per se:

$$
A = \int \int \mathcal{L} d^3x dt = \int \mathcal{L} d^4x. \tag{17.63}
$$

We now do the usual dance. We know that $\mathcal{L}$ for the fields must be a scalar. We also know that Maxwell's equations relate the fields to the currents that produce them, and also link the electric and magnetic fields. We thus need to build a theory out of $F^{\alpha\beta}$, A^α, J^α. Various ways we can do this include

$$
F_{\alpha\beta} F^{\alpha\beta}
$$

$$
J_\alpha A^\alpha
$$

$$
F_{\alpha\beta} \mathcal{F}^{\alpha\beta}
$$

and still messier pieces like

$$
F_{\alpha\beta} J^\alpha A^\beta.
$$

The first two terms are invariant under the transformations of the full Lorentz group. The third is not a scalar under inversion, but a pseudoscalar (odd under inversion). We reject it. The last is a mess. We reject it. We want a term quadratic in the 4-gradients of the fields. This is the first term. We want a source term to couple the fields and the particles. The second term does that.

So, we try a Lagrangian density with just these two terms, with unknown constants Q and R that have to be determined so that they correctly reconstruct Maxwell's equations in whatever system of units we like:

$$\mathcal{L} = -QF_{\alpha\beta}F^{\alpha\beta} - RJ_\alpha A^\alpha. \tag{17.64}$$

We need to take derivatives of $\mathcal{L}$ with respect to $\partial^\beta A^\alpha$, so it is useful to write this Lagrangian in the form:

$$\mathcal{L} = -Qg_{\lambda\mu}g_{\nu\sigma}(\partial^\mu A^\sigma - \partial^\sigma A^\mu)(\partial^\lambda A^\nu - \partial^\nu A^\lambda) - RJ_\alpha A^\alpha. \tag{17.65}$$

When we form $\partial\mathcal{L}/\partial(\partial^\beta A^\alpha)$ we get delta functions whenever α and β are equal to a pair of the indices above. We therefore get four terms:

$$\frac{\partial\mathcal{L}}{\partial(\partial^\beta A^\alpha)} = -Qg_{\lambda\mu}g_{\nu\sigma}\left\{\delta^\mu_\beta\delta^\sigma_\alpha F^{\lambda\nu} - \delta^\sigma_\beta\delta^\mu_\alpha F^{\lambda\nu} + \delta^\lambda_\beta\delta^\nu_\alpha F^{\mu\sigma} - \delta^\nu_\beta\delta^\lambda_\alpha F^{\mu\sigma}\right\} \tag{17.66}$$

where the first two terms come from delta functions formed from the first term and the second two terms come from delta functions formed from the second term.

$g_{\alpha\beta}$ is symmetric (in fact, diagonal). The $F^{\alpha\beta}$ is antisymmetric. When we do the sums against the δ–functions, the four terms make identical contributions:

$$\frac{\partial\mathcal{L}}{\partial(\partial^\beta A^\alpha)} = -4QF_{\beta\alpha} = 4QF_{\alpha\beta}. \tag{17.67}$$

The other part of the E–L equation (which corresponds in position space to the "potential", or "force" term) is

$$\frac{\partial\mathcal{L}}{\partial A^\alpha} = -RJ_\alpha. \tag{17.68}$$

Therefore the equations of motion for the electromagnetic field can be written

$$4Q\partial^\beta F_{\beta\alpha} = RJ_\alpha. \tag{17.69}$$

If one checks back in one's notes, one sees that this is indeed the covariant form of the inhomogeneous Maxwell's equations if $Q = 1/4$ and $R = \mu_0$:

$$\partial^\beta F_{\beta\alpha} = \mu_0 J_\alpha \tag{17.70}$$

follows from the Lagrangian density:

$$\mathcal{L} = -\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta} - \mu_0 J_\alpha A^\alpha. \tag{17.71}$$

Therefore the Lagrangian we have constructed yields the inhomogeneous Maxwell equations, but not the homogeneous ones. That is okay, though, because we have constructed the $F^{\alpha\beta}$ in terms of the A^α in such a way that the homogeneous ones are satisfied *automatically*! To observe that this miracle is true, we recall the covariant form of the homogeneous equations:

$$\partial_\alpha \mathcal{F}^{\alpha\beta} = 0. \tag{17.72}$$

Also,

$$\mathcal{F}^{\alpha\beta} = \frac{1}{2}\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta}. \tag{17.73}$$

Thus

$$\begin{aligned}
\partial_\alpha \mathcal{F}^{\alpha\beta} &= \frac{1}{2}\partial_\alpha \epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta} \\
&= \partial_\alpha \epsilon^{\alpha\beta\gamma\delta}\partial_\gamma A_\delta \\
&= \epsilon^{\alpha\beta\gamma\delta}\partial_\alpha\partial_\gamma A_\delta
\end{aligned} \tag{17.74}$$

is the first term. But $\partial_\alpha\partial_\gamma$ is symmetric, while $\epsilon^{\alpha\beta\gamma\delta}$ is antisymmetric in the same two indices, so the contraction on the two indices vanishes (work it out term by term if you doubt it).

Thus the homogeneous equations are satisfied by our *definition* of $\mathcal{F}^{\alpha\beta}$ quite independent of any dynamics. In four dimensions, all of the inhomogeneous source terms must appear in equations with the form of the inhomogeneous equation above, and only one of these equations can result from the action principle. The similarity transformation to the fields we observe is thus the "natural" form of the ME's, and in four dimensions the homogeneous equations are really not independent as far as the action principle is concerned. Note that this is fundamentally because our field strength tensor derives from the definition of the magnetic field as the curl of the vector field $\vec{A}$ (which is divergenceless) which is built into the definition.

As a final quixotic note, observe that if we take the 4–divergence of both sides of the inhomogeneous Maxwell equations:

$$\partial^\alpha\partial^\beta F_{\beta\alpha} = \mu_0\partial^\alpha J_\alpha \tag{17.75}$$

the left hand side vanishes because again, a *symmetric* differential operator is contracted with a *completely antisymmetric* field strength tensor. Thus

$$\partial^\alpha J_\alpha = 0, \tag{17.76}$$

which, by some strange coincidence, is the charge–current conservation equation in four dimensions. Do you get the feeling that something very deep is going on? This is what I love about physics. Beautiful things are really *beautiful!*

We will now blow off the "proca" Lagrangian, which would be appropriate if the photon had a mass. It doesn't, but if it did you would need to read this chapter. It might, of course, so you should probably read the chapter anyway, but it currently (bad pun) doesn't so I'm going to make light of it (worse pun) and continue.

If we had one more month, we would now study the covariant forms of the stress tensor. It is important, but it is also quite difficult, and necessitates a broader discussion than we can now afford. To treat the subject properly, we would need to treat parts of chapter 17 simultaneously, and we would need to do a *lot* of algebra. This would mean that we would miss (in all probability) being able to learn the Liénard–Wiechart potential, which is far more important. We will therefore content ourselves with merely defining the stress tensor, remarking on some of its properties without proof, and moving on. You are responsible for working your way through this chapter, according to your needs, inclinations, and abilities, on your own.

17.4 The Symmetric Stress Tensor

Imagine a big blob of jelly. Imagine poking it on a side. The whole thing wiggles and distorts, as the force of your poke acts on the entire blob of jelly. The mathematical mechanism that describes how your poke is distributed is calle the **stress tensor** of the material. It tells how energy and momentum are connected *by the medium itself.*

The same concept can be generalized to a four dimensional medium, where the "jelly" is space time itself. Let us now study what an electromagnetic stress tensor is, and how it relates to electromagnetic "pokes". Recall that

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \tag{17.77}$$

is the canonical momentum corresponding to the variable q_i in an arbitrary

Lagrangian. The Hamiltonian is given, in this case, by

$$H = \sum_i p_i \dot{q}_i - L \tag{17.78}$$

as usual. If $\partial L/\partial t = 0$ then one can show that $\partial H/\partial t = 0$. For four dimensional fields we should probably have a Lagrangian and Hamiltonian density whose 3–integral are the usual Lagrangian and Hamiltonians. The Hamiltonian is the energy of a particle or system, so it should transform like the zeroth component of a four vector. Thus, since

$$H = \int \mathcal{H} d^3x \tag{17.79}$$

and $d^4x = d_0 x d^3 x$, then $\mathcal{H}$ must transform like the time component of a second rank tensor. If we define the Hamiltonian density $\mathcal{H}$ in terms of the Lagrangian density $\mathcal{L}$ of a field, then

$$\mathcal{H} = \sum_k \frac{\partial \mathcal{L}}{\partial\left(\frac{\partial \phi_k}{\partial t}\right)} \frac{\partial \phi_k}{\partial t} - \mathcal{L}. \tag{17.80}$$

Well, great! The first factor in the sum is the conjugate momentum by definition, and the second is the generalized "velocity". Since $\mathcal{H}$ must transform like the time component of a second rank tensor (and the time derivative appears in this equation) it appears that the covariant generalization of the Hamiltonian density is something that puts a covariant derivative there, instead. We try

$$T^{\alpha\beta} = \sum_k \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \phi_k)} \partial^\beta \phi_k - g^{\alpha\beta}\mathcal{L}. \tag{17.81}$$

This is called the **canonical stress tensor**, and is related to the stress tensor defined and studied in Chapter 6. This tensor has the covariant function of telling us how the energy and momentum carried by the electromagnetic field transform.

What is this tensor? It is, in fact, highly non–trivial. The best we can do is note that *if* we assume that only free fields are present *and* that the free fields are localized in some finite region of space (neither assumption is particularly physical), then we can show that

$$\int T^{00} d^3x = \frac{1}{2}\int (\epsilon_0 \mathbf{E}^2 + \frac{1}{\mu_0}\mathbf{B}^2) d^3x = E_{\text{field}} \tag{17.82}$$

and

$$\int T^{0i} d^3x = \epsilon_0 c \int (\mathbf{E} \times \mathbf{B})_i d^3x = \frac{1}{c} \int (\mathbf{E} \times \mathbf{H})_i d^3x = c P^i_{\text{field}} \tag{17.83}$$

which are the "usual" expressions for the energy and momentum of the free field. At least if I got the change to SI units right...

What, you might ask, is this good for? Well, aside from this correspondance (which is full of holes, by the way), we can write the energy–momentum conservation law

$$\partial_\alpha T^{\alpha\beta} = 0. \tag{17.84}$$

This is proven in Jackson, with a discussion of some of its shortcomings.

One of these is that it is not symmetric. This creates difficulties when we consider the angular momentum carried by the field. Since the angular momentum density is important when we go to create photons (which must have quantized angular momenta), it is worthwhile to construct the **symmetric stress tensor**

$$\Theta^{\alpha\beta} = \left(g^{\alpha\mu} F_{\mu\lambda} F^{\lambda\beta} + \frac{1}{4} g^{\alpha\beta} F_{\mu\lambda} F^{\mu\lambda} \right) \tag{17.85}$$

in terms of which we can *correctly* construct a covariant generalization of the energy momentum conservation law

$$\partial_\alpha \Theta^{\alpha\beta} = 0 \tag{17.86}$$

and the angular momentum tensor

$$M^{\alpha\beta\gamma} = \Theta^{\alpha\beta} x^\gamma - \Theta^{\alpha\gamma} x^\beta \tag{17.87}$$

which is therefore conserved. This form of the stress tensor can also be directly coupled to source terms, resulting in the covariant form of the work energy theorem for the combined system of particles and fields.

This is about all we will say about this at this time. I realize that it is unsatisfactory and apologize. If we had one more semester together, we could do it properly, but we don't. Therefore, it is on to

17.5 Covariant Green's Functions

Just when you thought it was safe to go back into the classroom, along comes Jaws himself. Green's functions are your *friends!*

The inhomogeneous Maxwell equations are now compactly written as

$$\partial_\alpha F^{\alpha\beta} = \mu_0 J^\beta. \tag{17.88}$$

From the definition of the field strength tensor, this is

$$\Box A^\beta - \partial^\beta(\partial_\alpha A^\alpha) = \mu_0 J^\beta \tag{17.89}$$

If the potentials satisfy the Lorentz condition, $\partial_\alpha A^\alpha = 0$ and therefore

$$\Box A^\beta = \mu_0 J^\beta \tag{17.90}$$

Do you get the feeling that there is something mystical about space–time notations? Do you remember what a pain in the butt this was to derive the hard way?

To solve this inhomogeneous differential equation, we construct simultaneously a Green's function

$$\Box D(x, x') = \delta^{(4)}(x - x') \tag{17.91}$$

and the associated integral equation over the source term:

$$A^\alpha(x) = A_I^\alpha + \mu_0 \int d^4x' D(x - x') J^\alpha(x') \tag{17.92}$$

(where the inhomogeneous term A_I^α depends on the Green's function and is the "boundary" term or the free potential from inhomogeneous sources outside the region of integration).

Next week we will concentrate on the integral equation solutions themselves. Now let us see how to construct the appropriate (covariant) Green's function. As usual, the principle part of the Green's function can involve only the absolute distance between the points. Thus if $y^\alpha = x^\alpha - x'^\alpha$ we seek solutions to

$$\Box D(y) = \delta^{(4)}(y). \tag{17.93}$$

There are several ways we could go about solving this equation. They are all equivalent at some level or another. For example, we have already solved this equation for a single fourier component in Chapter 9. We could transform this result and obtain a four dimensional result. However, a more general procedure is to construct the solution from scratch.

The four dimensional fourier transform of the desired Green's function is defined by

$$D(y) = \frac{1}{(2\pi)^4} \int d^4k \tilde{D}(k) e^{-ik\cdot y} \tag{17.94}$$

where $k \cdot y = k_0 y_0 - \boldsymbol{k} \cdot \boldsymbol{y}$. The four dimensional delta function is

$$\delta^4(y) = \frac{1}{(2\pi)^4} \int d^4k d^{-ik\cdot y} \tag{17.95}$$

so (taking the $\Box$ of $D(y)$ under the integral and equating factors)

$$\tilde{D}(k) = -\frac{1}{k \cdot k}. \tag{17.96}$$

We therefore know that the Green's function has the form

$$D(y) = \frac{-1}{(2\pi)^4} \int d^4k \frac{e^{-ik\cdot y}}{k \cdot k}. \tag{17.97}$$

The integrand in this expression is singular when $k \cdot k = 0$. Recall that the presence of singularities means that we have to decide how to treat them to get a well-defined result. There are several ways to do this, and each has a physical interpretation. If we integrate over the "time" component k_0 first, we get

$$D(y) = -\frac{1}{(2\pi)^4} \int d^3k e^{i\boldsymbol{k}\cdot\boldsymbol{y}} \int dk_0 \frac{e^{-ik_0 y_0}}{k_0^2 - \kappa^2} \tag{17.98}$$

where $|\mathbf{k}| = \kappa$. Now the singularities live in a single 1–D integral that we can easily evaluate via contour integration and the method of residues provided that we select a suitable contour.

Let's do the integral carefully (in case your contour integration is bit rusty). Note that the poles of this integral are both real. This means that the integral is ambiguous – it can be assigned any of several possible values depending on how we choose to evaluation it. It is beyond the scope of these notes to evaluate the consequences of making and physically interpreting each of these choices. Instead we will choose to include *both* poles completely within a standard contour closed in the upper or lower half plane respectively, and then take limits such that the poles return to the real axis

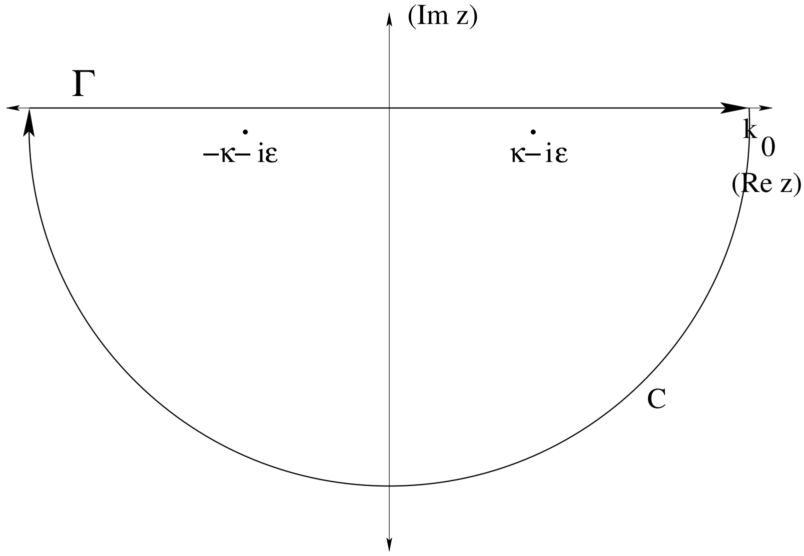

Figure 17.1: Contours for evaluating the Green's function in 4–dimensions.

after the integral because this particular choice leads us very simply to the advanced and retarded forms of the Green's function that we already obtained when discussing the fourier transform of the incoming or outgoing spherical Green's functions for the Helmholtz equation.

First we have to decide which way to close the contour. Examining the integrand, we note that if $y^0 = x^0 - x'^0 > 0$ the integrand vanishes on a lower-half contour like C in the figure above. We displace the poles down slightly so that they lie inside the contour Γ: $\pm\kappa \to \pm\kappa - i\epsilon$. Finally, let $z = k_0 + ik_i$ be a complex variable such that the real axis is k_0.

$$\oint_\Gamma dz \frac{e^{-iy_0 z}}{z^2 - \kappa^2} = \int_{-\infty}^{\infty} dk_0 \frac{e^{-ik_0 y_0}}{k_0^2 - \kappa^2} + \int_C dz \frac{e^{-iy_0 z}}{z^2 - \kappa^2} \tag{17.99}$$

As noted, the integral over C clearly vanishes for $y_0 > 0$. Thus:

$$\begin{aligned}
\int_{-\infty}^{\infty} dk_0 \frac{e^{-ik_0 y_0}}{k_0^2 - \kappa^2} &= \oint_\Gamma dz \frac{e^{-iy_0 z}}{z^2 - \kappa^2} \\
&= \lim_{\epsilon \to 0} (-2\pi i) \mathrm{Res} \frac{e^{-izy_0}}{(z - (\kappa - i\epsilon))(z + (\kappa + i\epsilon)} \\
&= -2\pi i \left\{ \frac{e^{-i\kappa y_0}}{2\kappa} + \frac{e^{i\kappa y_0}}{-2\kappa} \right\} \\
&= -2\pi \frac{sin(\kappa y_0)}{\kappa} \qquad (17.100)
\end{aligned}$$

We can then write the Green's function as

$$\begin{aligned} D(z) &= \frac{\theta(y_0)}{(2\pi)^3}\int d^3k e^{i\mathbf{k}\cdot\mathbf{z}}\frac{\sin(\kappa z_0)}{\kappa} \\ &= \frac{\theta(y_0)}{(2\pi)^3}\int_0^\infty \kappa^2 d\kappa \int_{-1}^{1} d(\cos(\theta))\int_0^{2\pi} d\phi e^{i\kappa R\cos(\theta)}\frac{\sin(\kappa y_0)}{\kappa} \\ &= \frac{\theta(y_0)}{(2\pi)^2}\int_0^\infty \kappa^2 d\kappa \int_{-i\kappa R}^{i\kappa R} d(i\kappa R\cos(\theta))\frac{e^{i\kappa R\cos(\theta)}}{i\kappa R}\frac{\sin(\kappa y_0)}{\kappa} \\ &= \frac{\theta(y_0)}{2\pi^2 R}\int_0^\infty d\kappa \sin(\kappa R)\sin(\kappa y_0) \end{aligned} \tag{17.101}$$

where $R = |\mathbf{x}-\mathbf{x}'|$ is the spatial separation of the points x and x'.

Using a trig identity (or if you prefer expanding the sin's in terms of exponentials and multiplying out, then changing variables and exploiting the fact that only even terms survive) to extend the integral to $-\infty$ we can write this as:

$$D(z) = \frac{\theta(y^0)}{4\pi R}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty} d\kappa \left(e^{i(y_0-R)\kappa} - e^{i(y_0+R)\kappa}\right)\right\}. \tag{17.102}$$

These remaining integrals are just one dimensional Dirac delta functions. Evaluating, we get:

$$D_r(x-x') = \frac{\theta(x^0 - x'^0)}{4\pi R}\left\{\delta(x^0 - x'^0 - R) + \delta(x^0 - x'^0 + R)\right\} \tag{17.103}$$

where we have now labelled it with "r" for "retarded". The source event x' is always at an *earlier time* than the observation event x. This means that the domain of the support of the Heaviside function just happens to be disjoint from the support of the second delta function. We can therefore simplify this to:

$$D_r(x-x') = \frac{\theta(x^0 - x'^0)}{4\pi R}\delta(x^0 - x'^0 - R) \tag{17.104}$$

which is just what we got before from Fourier transforming the outgoing stationary wave Green's function, as it should be.

If we had chosen the other contour, identical arguments would have led us to the *advanced* Green's function:

$$D_a(x-x') = \frac{\theta[-(x_0 - x_0')]}{4\pi R}\delta(x_0 - x_0' + R) \tag{17.105}$$

The other *possible* contours (enclosing only one or the other of the two singularities, using a contour that avoids the singularities on the real axis instead of displacing the singularities) would yield still other possible Green's functions. Just as an arbitrary normalized sum of outgoing and incoming Green's functions resulted in an acceptable Green's function before, an arbitrary sum of advanced and retarded Green's functions are acceptable here. However, the inhomogeneous term of the integral equation is a functional of the Green's function selected!

For what it is worth, the Green's functions can be put in covariant form. One almost never uses them in that form, and it isn't pretty, so I won't bother writing it down. We can now easily write down formal solutions to the wave equation for *arbitrary* currents (*not* just harmonic ones):

$$A^{\alpha}(x) = A^{\alpha}_{\text{in}}(x) + \mu_0 \int d^4x' D_r(x - x') J^{\alpha}(x') \tag{17.106}$$

and

$$A^{\alpha}(x) = A^{\alpha}_{\text{out}}(x) + \mu_0 \int d^4x' D_a(x - x') J^{\alpha}(x'). \tag{17.107}$$

In these equations, the inhomogeneous terms are the radiation field incident upon (radiated from) the *four*-volume of space-time containing the four-current that are *not* connected to the four-current in that four-volume by the retarded Green's function.

It is a worthwhile exercise to meditate upon what might be a suitable form for the inhomogeneous terms if one considerst the integration four-volume to be *infinite* (with no inhomogeneous term at all) and then split the infinite volume up into the interior and exterior of a finite four-volume, as we did with incoming and outgoing waves before, especially when there are many charges and they are permitted to interact.

Dirac noted that choosing a "retarded" Green's function, just as choosing an "outgoing wave" Green's function before, results in a somewhat misleading picture given that the actual physics is completely time-reversal symmetric (indeed, independent of using a *mixed* version of the Green's functions in either case). He therefore introduced the "radiation field" as the *difference* between the "outgoing" and the "incoming" inhomogenous terms given the contraint that the actual vector potential is the same regardless of the choice of Green's function used::

$$A^{\alpha}_{\text{radiation}} = A^{\alpha}_{\text{out}} - A^{\alpha}_{\text{in}} = \frac{4\pi}{c} \int d^4x' D(x - x') J^{\alpha}(x') \tag{17.108}$$

where

$$D(z) = D_r(z) - D_a(z). \tag{17.109}$$

In some fundamental sense, only the radiation fields are "physical" – they are the *change* in the vector potential at an event produced symmetrically by any given four-current due to its past *and its future* motion. This is a critical aspect of the interpretation of radiation reaction as being produced by transfer of momentum both *to* a charge (event) from other charges in its past and *from a charge* to those *same* charges in its future.

Chapter 18

Radiation from Point Charges

To summarize from the last chapter, two useful Green's functions for the inhomogeneous wave equation:

$$\Box A^\alpha = \mu_0 J^\alpha \tag{18.1}$$

are

$$D_r(x - x') = \frac{\theta(x^0 - x'^0)}{4\pi R}\delta(x^0 - x'^0 - R) \tag{18.2}$$

(the retarded Green's function) and

$$D_a(x - x') = \frac{\theta[-(x^0 - x'^0)]}{4\pi R}\delta(x^0 - x'^0 + R) \tag{18.3}$$

(the advanced Green's function). The integral equations associated with these Green's functions were:

$$A^\alpha(x) = A^\alpha_{\text{in}}(x) + \mu_0 \int d^4x' D_r(x - x') J^\alpha(x') \tag{18.4}$$

and

$$A^\alpha(x) = A^\alpha_{\text{out}}(x) + \mu_0 \int d^4x' D_a(x - x') J^\alpha(x'). \tag{18.5}$$

For the moment, let us ignore Dirac's observations and the radiation field and focus instead on only the "normal" causally connected retarded potential produced by a single charged particle as it moves in the absence of external potentials. This potential is "causal" in that the effect (the potential field) follows the cause (the motion of the charge) in time, where the advanced potential has the effect preceding the cause, so to speak. Let me emphasize that this is not a particularly consistent assumption (again, we the theory is manifestly time symmetric so "past" and "future" are pretty much arbitrary namings of two opposed directions), but it yields some very nice results, as well as some problems. In that case:

$$A^\alpha(x) = \mu_0 \int d^4x' D_r(x-x')J^\alpha(x') \tag{18.6}$$

where the four–current of a point charge e is found from

$$J(\boldsymbol{x}', t) = \left\{ \begin{array}{c} c\rho(\boldsymbol{x}', t) \\ \boldsymbol{J}(\boldsymbol{x}', t) \end{array} \right\} = \left\{ \begin{array}{c} ec\ \delta[\boldsymbol{x}' - \boldsymbol{r}(t)] \\ e\boldsymbol{v}\delta[\boldsymbol{x}' - \boldsymbol{r}(t)] \end{array} \right\} \tag{18.7}$$

in the lab/rest frame K or (in covariant form):

$$J^\alpha(x') = ec \int d\tau U^\alpha(\tau)\delta^{(4)}\left([x' - r(\tau)]\right) \tag{18.8}$$

where

$$U = \gamma \left(\begin{array}{c} c \\ \boldsymbol{v} \end{array} \right) = \frac{dr}{d\tau} \tag{18.9}$$

Note that the δ function in these expressions simply forces the particle to be found at the correct location at each (proper) time. The $r(\tau)$ function is the trajectory of the particle. Its τ derivative is the four–velocity. This yields (when the γ's have all been accounted for) the rest frame expression.

To do the integral, we need the "manifestly covariant" form of the retarded Green's function. Note that:

$$\begin{aligned} \delta[(x-x')^2] &= \delta[(x^0 - x'^0)^2 - |\boldsymbol{x} - vx'|^2] \\ &= \delta[(x^0 - x'^0 - R)(x^0 - x'^0 + R)] \\ &= \frac{1}{2R}[\delta(x_0 - x_0' - R) + \delta(x_0 - x_0' + R)] \end{aligned} \tag{18.10}$$

(where $R = |\boldsymbol{x} - \boldsymbol{x}'|$). In terms of this, D_r is given by

$$D_r(x-x') = \frac{1}{2\pi}\theta(x_0 - x_0')\delta[(x-x')^2]. \tag{18.11}$$

Again, the second delta–function makes no contribution because of the opposing θ–function. Thus

$$\begin{aligned} A^\alpha(x) &= \frac{\mu_0 c}{2\pi} \int d^4x' \theta(x_0 - x_0')\delta([x - x']^2) \\ &\qquad \times e \int d\tau U^\alpha(\tau)\delta^{(4)}\left([x' - r(\tau)]\right) \qquad (18.12) \\ &= \frac{e\mu_0 c}{2\pi} \int d\tau U^\alpha(\tau)\theta[x_0 - r_x(\tau)]\delta\left\{[x - r(\tau)]^2\right\}. \qquad (18.13) \end{aligned}$$

The vector potential at a point gets a contribution only where–when that point lies on the light cone in the future (picked out by the θ function) of the world line of the charge (picked out be the δ function). The contribution is proportional to $eU^\alpha(\tau)$ at that (retarded) time. It dies off like $1/R$, although that is obscured by the form of the δ function.

To evaluate this (and discover the embedded R), we use the rule (from way back at the beginning of the book, p. 30 in J1.2)

$$\delta[f(x)] = \sum_i \frac{\delta(x - x_i)}{\left| \left(\frac{df}{dx}\right)_{x=x_i} \right|} \tag{18.14}$$

where the $x = x_i$ are the non–degenerate zeros of $f(x)$. $f(x)$ is assumed to be "smooth". Then if we let

$$f(\tau) = [x - r(\tau)]^2 \tag{18.15}$$

(which is zero when $\tau = \tau_p$ in the past) then

$$\frac{d}{d\tau}[x - r(\tau)]^2 = -2[x - r(\tau)]_\beta U^\beta(\tau) \tag{18.16}$$

and therefore

$$\delta([x - r(\tau)]^2) = \frac{\delta(\tau - \tau_p)}{|-2[x - r(\tau)]_\beta U^\beta(\tau)|} = \frac{\delta(\tau - \tau_p)}{2[x - r(\tau)]_\beta U^\beta(\tau)} \tag{18.17}$$

From this we see that

$$A^\alpha(x) = \frac{e\mu_0 c}{4\pi} \left. \frac{U^\alpha(\tau)}{U \cdot [x - r(\tau)]} \right|_{\tau=\tau_p} \tag{18.18}$$

where τ_p is the proper time in the past of x when the light cone of the charge contains the event x. This potential (and its other forms above) are called the *Liénard–Wiechert potentials.* In non–covariant form, they are obtained from the identity

$$\begin{aligned} U \cdot (x - r) &= U_0[x_0 - r_0(\tau_p)] - \mathbf{U} \cdot [\mathbf{x} - \mathbf{r}(\tau_p)] \\ &= \gamma c R(1 - \beta \cdot \mathbf{n}) \end{aligned} \tag{18.19}$$

where $\mathbf{n}$ is a unit vector in the direction of $\mathbf{x} - \mathbf{r}(\tau)$ and where $\beta = \mathbf{v}(\tau)/c$ as usual.

Recall that $A = (\phi/c, \boldsymbol{A})$. Thus:

$$A^0(x) = \frac{e\mu_0 c}{4\pi} \left. \frac{\gamma c}{\gamma c R(1 - \boldsymbol{\beta} \cdot \hat{\boldsymbol{n}})} \right|_{\mathrm{ret}} \tag{18.20}$$

and

$$\phi(\boldsymbol{x}, t) = cA^0 = \frac{e}{4\pi\epsilon_0} \left. \frac{1}{R(1 - \boldsymbol{\beta} \cdot \hat{\boldsymbol{n}})} \right|_{\mathrm{ret}} \tag{18.21}$$

where all quantities (e.g. $\boldsymbol{\beta}, R$) must be evaluated at the retarded time where the event x is on the light cone of a point on the particle trajectory.

Similarly

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{x}, t) &= \frac{e\mu_0 c}{4\pi} \left. \frac{\gamma c \boldsymbol{\beta}}{\gamma c R(1 - \boldsymbol{\beta} \cdot \hat{\boldsymbol{n}})} \right|_{\mathrm{ret}} \\ &= \frac{e}{4\pi} \sqrt{\frac{\mu_0}{\epsilon_0}} \left. \frac{\boldsymbol{\beta}}{R(1 - \boldsymbol{\beta} \cdot \hat{\boldsymbol{n}})} \right|_{\mathrm{ret}} \end{aligned} \tag{18.22}$$

where again things must be evaluated at retarded times on the particle trajectory. Note well that both of these manifestly have the correct non-relativistic form in the limit $|\boldsymbol{\beta}| << 1$.

We can get the fields from the 4–potential in any of these forms. However, the last few forms we have written are compact, beautiful, intuitive, and have virtually no handles with which to take vector derivatives. It is simpler to return to the integral form, where we can let ∂^α act on the δ and θ functions.

$$\partial^\alpha A^\beta = \frac{e\mu_0 c}{2\pi} \int d\tau U^\beta(\tau) \theta[x_0 - r_0(\tau)] \partial^\alpha \delta\left([x - r(\tau)]^2\right) \tag{18.23}$$

where

$$\partial^\alpha \delta[f] = \partial^\alpha f \cdot \left(\frac{d}{df}\delta[f]\right) = \partial^\alpha f \cdot \frac{d\tau}{df} \cdot \frac{d}{d\tau}\delta[f]. \tag{18.24}$$

Again, we let $f = [x - r(\tau)]^2$. Then

$$\partial^\alpha \delta[f] = -\frac{(x-r)^\alpha}{U \cdot (x-r)} \frac{d}{d\tau}\delta[f] \tag{18.25}$$

This is inserted into the expression above and integrated by parts:

$$\begin{aligned} \partial^\alpha A^\beta &= -\frac{e\mu_0 c}{2\pi} \int d\tau U^\beta(\tau) \theta[x_0 - r_0(\tau)] \frac{(x-r)^\alpha}{U \cdot (x-r)} \frac{d}{d\tau}\delta[f] \\ &= \frac{e\mu_0 c}{2\pi} \int d\tau \frac{d}{d\tau}\left\{U^\beta(\tau) \frac{(x-r)^\alpha}{U \cdot (x-r)}\right\} \theta[x_0 - r_0(\tau)] \delta([x - r(\tau)]^2) \end{aligned} \tag{18.26}$$

There is no contribution from the θ function because the derivative of a theta function is a delta function with the same arguments

$$\frac{d}{d\tau}\theta(x_0 - r_0(\tau)) = \delta[x_0 - r_0(\tau)] \tag{18.27}$$

which constrains the other delta function to be $\delta(-R^2)$. This only gets a contribution at $R = 0$ (on the world line of the charge), but we already feel uncomfortable about the field there, which we suspect is infinite and meaningless, so we exclude this point from consideration. Anywhere *else* the result above is exact.

We can now do the integrals (which have the same form as the potential integrals above) and construct the field strength tensor:

$$F^{\alpha\beta} = \frac{e\mu_0 c}{4\pi}\,\frac{e}{U\cdot(x-r)}\frac{d}{d\tau}\left\{\frac{(x-r)^\alpha U^\beta - (x-r)^\beta U^\alpha}{U\cdot(x-r)}\right\}\Bigg|_{\rm ret} \tag{18.28}$$

This whole expression must be evaluated after the differentiation at the retarded proper time τ_p.

This result is beautifully covariant, but not particularly transparent for all of that. Yet we will need to find explicit and useful forms for the fields for later use, even if they are not as pretty. Jackson gives a "little" list of ingredients (J14.12) to plug into this expression when taking the derivative to get the result, which is obviously quite a piece of algebra (which we will skip):

$$\boldsymbol{E}(\boldsymbol{x},t) = \frac{e\mu_0}{4\pi c^2}\left[\frac{(\hat{\boldsymbol{n}}-\boldsymbol{\beta})}{\gamma^2(1-\boldsymbol{\beta}\cdot\hat{\boldsymbol{n}})^3R^2}\right]_{\rm ret} + \frac{e\mu_0}{4\pi c^3}\left[\frac{\hat{\boldsymbol{n}}\times\left((\hat{\boldsymbol{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}\right)}{(1-\boldsymbol{\beta}\cdot\hat{\boldsymbol{n}})^3R}\right]_{\rm ret} \tag{18.29}$$

and

$$\boldsymbol{B}(\boldsymbol{x},t) = \frac{1}{c}(\hat{\boldsymbol{n}}\times\boldsymbol{E}) \tag{18.30}$$

"Arrrgh, mateys! Shiver me timbers and avast!", you cry out in dismay. "This is easier? Nonsense!" Actually, though, when you think about it (so think about it) the first term is clearly (in the low velocity, low acceleration limits) the *usual static field*:

$$\boldsymbol{E} \approx \frac{e}{4\pi\epsilon_0}\frac{\hat{\boldsymbol{n}}}{R^2} \tag{18.31}$$

Interestingly, it has a “short” range and is isotropic.

The second term is proportional to the *acceleration* of the charge; both **E** and **B** are transverse and the fields drop off like R^{-1} and hence are “long range” but highly directional.

If you like, the first terms are the “near” and “intermediate” fields and the second is the complete “far” field; only the far field is produced by the acceleration of a charge. Only this field contributes to a net radiation of energy and momentum away from the charge.

With that (whew!) behind us we can proceed to discuss some important expressions. First of all, we need to obtain the power radiated by a moving charge.

18.1 Larmor's Formula

If one is far (enough) away from the an accelerating charge in the right direction, the field is given by primarily by the second (acceleration) term. This is the “usual” transverse EM field. If the particle is moving slowly with respect to c (so $\beta << 1$), then

$$\boldsymbol{E} = \frac{e}{4\pi\epsilon_0}\frac{1}{c}\left.\frac{\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\dot{\boldsymbol{\beta}})}{R}\right|_{\rm ret} \tag{18.32}$$

$$\boldsymbol{B} = \frac{e}{4\pi\epsilon_0}\frac{1}{c^2}\left.\frac{\hat{\boldsymbol{n}}\times\dot{\boldsymbol{\beta}}}{R}\right|_{\rm ret} \tag{18.33}$$

The energy flux is given by the (instantaneous) Poynting vector:

$$\begin{aligned}
\boldsymbol{S} &= \frac{1}{\mu_0}(\boldsymbol{E}\times\boldsymbol{B}) \\
&= \frac{e^2}{16\pi^2\epsilon_0\ R^2}\frac{1}{\mu_0\epsilon_0}\frac{1}{c^3}|\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\dot{\boldsymbol{\beta}})|^2\hat{\boldsymbol{n}} \\
&= \frac{e^2}{16\pi^2\epsilon_0\ R^2}\frac{1}{c^3}|\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times c^2\dot{\boldsymbol{\beta}})|^2\hat{\boldsymbol{n}} \\
&= \frac{e^2}{16\pi^2\epsilon_0\ R^2}\frac{1}{c^3}|\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\dot{\boldsymbol{v}})|^2\hat{\boldsymbol{n}}
\end{aligned} \tag{18.34}$$

As always, the power cross-section (energy per unit solid angle) is

$$\begin{aligned}
\frac{dP}{d\Omega} &= \boldsymbol{S}\cdot\hat{\boldsymbol{n}}R^2 \\
&= \frac{e^2}{16\pi^2\epsilon_0}\frac{1}{c^3}|\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\dot{\boldsymbol{v}})|^2 \\
&= \frac{e^2}{16\pi^2\epsilon_0}\frac{1}{c^3}|\dot{\boldsymbol{v}}|^2\sin^2(\Theta)
\end{aligned} \tag{18.35}$$

where Θ is the angle between $\hat{\boldsymbol{n}}$ and $\dot{\boldsymbol{v}}$.

Aha! we say. The characteristic $\sin^2\Theta$! Aha again! Inspecting the vector products, we see that the radiation is polarized in the plane of $\mathbf{n}, \dot{\mathbf{v}}$, perpendicular to $\mathbf{n}$. Finally, the integral over angles yields $8\pi/3$, so that

$$P = \frac{e^2}{6\pi\epsilon_0 c^3}\,|\dot{\mathbf{v}}|^2\,. \tag{18.36}$$

This is the **Larmor formula** for the power radiated from a nonrelativistic accelerated point charge. This has a covariant generalization that is valid for any velocity of charge. First we factor out an m^2 and convert this to momentum coordinates. Then we realize that the energy carried by this field (per unit time) is indeed related to the momentum by a factor of $1/c$ and convert the whole thing to 4–vector form. Last, we convert t into τ:

$$\begin{aligned}
P &= \frac{e^2}{6\pi\epsilon_0 c^3}\frac{1}{m^2}\left|\frac{d(m\boldsymbol{v})}{dt}\right|^2 \\
&= \frac{e^2}{6\pi\epsilon_0 m^2 c^3}\left|\frac{d(m\boldsymbol{v})}{\gamma d\tau}\right|^2 \\
&= \frac{e^2}{6\pi\epsilon_0 m^2 c^3}(1-\beta^2)\left|\frac{d\boldsymbol{p}}{d\tau}\right|^2 \\
&= \frac{e^2}{6\pi\epsilon_0 m^2 c^3}\left(\frac{d\boldsymbol{p}}{d\tau}\right)^2 - \left(\frac{1}{c^2}\frac{dE}{d\tau}\right)^2 \\
&= -\frac{e^2}{6\pi\epsilon_0 m^2 c^3}\left(\frac{dp_\alpha}{d\tau}\frac{dp^\alpha}{d\tau}\right)
\end{aligned} \tag{18.37}$$

This can be written one more way, (substituting $E = \gamma mc^2$ and $\mathbf{p} = \gamma m\mathbf{v}$ and using some vector identities) due to Liénard:

$$P = \frac{e^2}{6\pi\epsilon_0 c^3}\gamma^6[(\dot{\boldsymbol{\beta}})^2 - (\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}})^2] \tag{18.38}$$

We are all better people for knowing this.

Why, you may ask, is this torture necessary? Because quite a few of you will spend unreasonable amounts of your lives calculating things like radiative losses in accelerators. After all, if we could build GeV accelerators in a little bitty ten foot ring it would be a whole lot cheaper than 6 billion bucks, plus inflation. Unfortunately, nature says that if you try it the nasty thing will give off *synchrotron radiation*! Let us see that tanstaafl[1].

The radiated power is proportional to the acceleration. The work is proportional to the tangential force times the velocity. Light particles accelerate the most for a given tangential force and have the highest velocity for a given energy; radiative losses are thus the most important for those particles at all energies. We will evaluate the radiative power loss for an electron in a linear accelerator.

We begin with

$$P = \frac{e^2}{6\pi\epsilon_0 m^2 c^3}\left(\frac{dp}{dt}\right)^2 \tag{18.39}$$

where $-e$ is now really the charge on the electron. Since the accelerator is linear, we can find the force directly from the rate at which work is done on the electron (otherwise we would have to include the force bending it in a curved path, which does no work). It is related to the "gradient" of the total energy,

$$P = \frac{e^2}{6\pi\epsilon_0 m^2 c^3}\left(\frac{dE}{dx}\right)^2. \tag{18.40}$$

For linear acceleration we don't care what the actual energy of the particle is; we only care how that energy changes with distance.

We will turn this into a rate equation by using the chain rule:

$$P_{\rm rad} = \frac{e^2}{6\pi\epsilon_0 m^2 c^3}\frac{dE}{dx}\frac{dE}{dt}\frac{dt}{dx} \tag{18.41}$$

Thus the ratio of power radiated to power supplied by the accelerator $P_{\rm acc} = dE/dt$ is:

$$\frac{P_{\rm rad}}{P_{\rm acc}} = \frac{e^2}{6\pi\epsilon_0 m^2 c^3}\frac{1}{v}\frac{dE}{dx} \approx \frac{1}{6\pi\epsilon_0}\frac{e^2/mc^2}{mc^2}\frac{dE}{dx} \tag{18.42}$$

where the latter form is valid when the electron is travelling at $v \approx c$.

[1]There Ain't No Such Thing As A Free Lunch. No kidding.

This quantity will be less than one while the gain in energy in a distance $e^2/mc^2 = 2.82 \times 10^{-13}$ cm is of the order of $mc^2 = .5$ MeV. That would require a potential difference (or other force) on the order of 10^{14} MV/meter.

Maybe at the surface of a positron. Come to think of it, falling into a positron there comes a point where this is true and at that point the total mass energy of the pair *is* radiated away. But nowhere else. We can completely neglect radiative losses for linear acceleration simply because the forces required to produce the requisite changes in energy *when the particle is moving at nearly the speed of light* are ludicrously large. For a charged particle moving in a straight line, radiative losses are more important at low velocities. This is fortunate, or radios and the like with linear dipole antennas would not work!

However, it is incovenient to build linear accelerators. That is because a linear accelerator long enough to achieve reasonable energies for electrons starts (these days) at around 100–500 miles long. At that point, it is still not "straight" because the earth isn't flat and we don't bother tunnelling out a secant. Also, it seems sensible to let a charged particle fall many times through the "same" potential, which is possible only if the accelerator is circular. Unfortunately, we get into real trouble when the accelerator is not straight.

In a circular accelerator, there is a non–zero force proportional to its velocity *squared*, even when *little or no work is being done to accelerate the particle!* In fact, the centripetal force on the particle is

$$\left|\frac{d\mathbf{p}}{d\tau}\right| = \gamma\omega\,|\mathbf{p}| >> \frac{1}{c}\frac{dE}{d\tau} \tag{18.43}$$

all of which increase as the speed of the particle increases. If we completely neglect the radiative loss due to tangential acceleration (which is completely negligible once relativistic velocities have been reached) we see that

$$P = \frac{e^2}{6\pi\epsilon_0 m^2 c^3}\gamma^2\omega^2\,|\mathbf{p}|^2 = \frac{e^2 c}{6\pi\epsilon_0 r^2}\beta^4\gamma^4 \tag{18.44}$$

where we have used $\omega = (c\beta/r)$. The loss *per revolution* is obtained by multiplying by T (the period of a revolution). This yields

$$\Delta E = \frac{2\pi r}{c\beta}P = \frac{e^2}{3\epsilon_0 r}\beta^3\gamma^4 \tag{18.45}$$

which is still deadly if r is small and/or γ and β are large.

If one does some arithmetic (shudder), one can see that for high energy electrons (where $\beta \approx 1$), this is

$$\Delta E(\mathrm{MeV}) = 8.85 \times 10^{-2} \frac{[E(\mathrm{GeV})]^4}{r(\mathrm{meters})}. \tag{18.46}$$

At around 1 GeV, one needs roughly $1/(10r)$ of that energy gain *per cycle* in order to turn (heh, heh) a net profit. That is not so bad, but the power of 4 says that at 10 GeV, one needs a gain per cycle of $1000/r$ GeV (!) in order to turn a profit. Now, it is true that the bigger the radius the longer the circumference (linearly) and the longer the circumference the more work one can do with a given fixed potential in a cycle. So in terms of *force* this relation is not as bad as it seems. But it is bad enough, because you still have to do the work, which costs you the same no matter how hard you have to push to do it. Clearly even at 10 GeV, an orbit of radius ~ 100 meters or better is necessary. In electron–positron storage rings, work must be done at this general rate just to keep the particles moving.

Those of you who need to know can read section 14.3 on your own. The results are straightforward but algebraically tedious, and are of use only if you plan on studying accelerator design or neutron stars. Don't get me wrong. Nobel prizes have been won for accelerator design and may be again. Go for it.

Ditto for 14.4. This is highly readable and contains no algebra. In a nutshell, a particle moving in a synchrotron emits its radiation in its instantaneous direction of motion (which is indeed perpendicular to the acceleration). Since it moves in a circle, a stationary observer in the plane of motion sees short bursts of radiation at the characteristic frequency c/r. The length (in time) of the pulses is L/c in time, and thus will contain frequencies up to $c/L \sim (c/r)\gamma^3$ in a fourier decomposition of their "wave packet" where $L \approx r/(2\gamma^3)$ is the length of the pulse in space. For highly relativistic particles moving in big circles, the characteristic frequency can be many orders of magnitude smaller than the high frequency cut off, as in AM radio frequencies to X–rays or worse. Synchrotron radiation is a potential source of high frequency electromagnetic energy.

Of course, it isn't tunable or coherent (in fact, its highly incoherent since the spectrum is so wide!) and we'd love to use the same kind of trick to make coherent, tunable, high frequency light. Some of you probably *will* use the same kind of trick before you leave, since free electron lasers produce energy from a similar principle (although with a totally different spectrum!). Section 14.6 deals with the spectrum, and we will blow that off, too. Suffice it to say that it can be calculated, and you can learn how, if you need to. You really should remember that $\omega_c \approx \omega_0 \gamma^3$, and should take a peek at the distribution curves. These curves let one detect synchrotron radiation from cosmological sources. These sources are generally charged particles falling into dark stars, radiation belts around planets, sunspots, or anyplace else that relativistic electrons are strongly accelerated in a circular, or helical, path. Finally, we will neglect 14.5 too, which analyzes radiation emitted by particles moving in wierd ways. Jackson is encyclopaediac, but we needn't be.

We will come back into focus at section 14.7, Thomson Scattering of Radiation. This is scattering of radiation by charged particles and is closely related to Compton scattering. It is important, as it is a common phenomenon.

18.2 Thomson Scattering of Radiation

Suppose that a plane wave of monochromatic electromagnetic radiation is incident on a free particle of charge e and mass m. The particle will experience a force from this field, and will accelerate. As it accelerates, it will emit radiation in different directions, *dispersing* the incident beam.

For a non–relativistic particle accelerated by a force we can see that:

$$\frac{dP}{d\Omega} = \frac{e^2}{16\pi^2\epsilon_0}\frac{1}{c^3}\left|\hat{\boldsymbol{\epsilon}}^* \cdot \dot{\boldsymbol{v}}\right|^2 \tag{18.47}$$

(where $|\hat{\boldsymbol{\epsilon}}^* \cdot \dot{\boldsymbol{v}}|^2 = |\dot{\boldsymbol{v}}|^2 \sin^2\Theta$ for a particular polarization perpendicular to the plane of $\hat{\boldsymbol{n}}$ and $\dot{\boldsymbol{v}}$).

The (leading order) acceleration is due to the plane wave electric field with polarization $\hat{\boldsymbol{\epsilon}}_0$, wave vector $\boldsymbol{k}_0$, and Newton's Law:

$$\dot{\boldsymbol{v}} = \frac{e}{m} E_0 \hat{\boldsymbol{\epsilon}}_0 e^{i\boldsymbol{k}_0\cdot\boldsymbol{x} - \omega t} \tag{18.48}$$

If the charge moves much less than one wavelength during a cycle (true for all but the lightest particles and strongest fields) then

$$|\dot{\boldsymbol{v}}|_{\text{av}} = \frac{1}{2}\text{Re }(\dot{\boldsymbol{v}} \cdot \dot{\boldsymbol{v}}^*) \tag{18.49}$$

Thus the average power flux distribution is

$$\begin{aligned}\left(\frac{dP}{d\Omega}\right)_{\text{av}} &= \frac{c}{32\pi^2\epsilon_0}|E_0|^2\left(\frac{e^2}{mc^2}\right)^2|\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 \\ &= \left\{\frac{e^2}{4\pi\epsilon_0 mc^2}\right\}^2\frac{\epsilon_0 c E_0^2}{2}|\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 \end{aligned}\tag{18.50}$$

This is clearly of the same general form as the scattering expressions we described and derived earlier. Since the result contains E_0^2 it makes sense to divide out the incident intensity and thus obtain a *differential* cross section that works for all but the strongest fields. We thus divide out the time-averaged flux of the Poynting vector of the incident plane wave:

$$I = \frac{\epsilon_0 c E_0^2}{2} \tag{18.51}$$

hence

$$\frac{d\sigma}{d\Omega} = \left\{\frac{e^2}{4\pi\epsilon_0 mc^2}\right\}^2|\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 \tag{18.52}$$

If we let the plane wave be incident along the z axis, let $\hat{\boldsymbol{n}}$ form an angle θ with that axis, and pick two polarization directions in and perpendicular to the $(\hat{\boldsymbol{n}}, \hat{\boldsymbol{z}})$ plane (as before), and average over polarizations then this dot product yields:

$$\frac{d\sigma}{d\Omega} = \left\{\frac{e^2}{4\pi\epsilon_0 mc^2}\right\}^2\frac{1}{2}(1+\cos^2\theta). \tag{18.53}$$

as it did back in our earlier work on scattering, but now for a *point* particle.

This is the **Thomson formula** for scattering of radiation by free charge. It works for X–rays for electrons or γ–rays for protons. It does not work when the photon momentum and the recoil of the charged particle cannot be neglected. The integral of this,

$$\sigma_T = \frac{8\pi}{3}\left\{\frac{e^2}{4\pi\epsilon_0 mc^2}\right\}^2 \tag{18.54}$$

is called the **Thomson cross–section**. It is 0.665×10^{-29} m^2 for electrons.

The quantity in parentheses has the units of length. If the total "mass–energy" of the electron were due to its charge being concentrated in a ball, then this would be the close order of the radius of that ball; it is called the **classical electron radius**. This number crops up quite frequently, so you should remember it. What it tells us is that even *point* particles have a *finite* scattering cross-section that appears in this limit to be independent of the wavelength of the light scattered.

However, this is not really true if you recall the approximations made – this expression will fail if the wavelength is on the same order as the classical radius, which is precisely where pair production becomes a significant process quantum mechanically. In quantum mechanics, if the energy of the incident *photon* $\hbar\omega \approx mc^2$ for the *electron*, significant momentum is transferred to the electron by the collision and the energy of the scattered photon cannot be equal to the energy of the incident photon. Whatever a photon is ...

We can actually fix that without too much difficulty, deriving the Compton scattering formula (which takes over from Thomson in this limit). This formula adds a wavelength/angle dependence to Thomson's general result and yields the *Klien-Nishina* formula, but this is beyond our scope in this course to derive or discuss in further detail.

We are almost finished with our study of electrodynamics. Our final object of study will be to to try to address the following observation:

Accelerated charges radiate. Radiation accelerates charge. Energy must be conserved. These three things have not been consistently maintained in our treatments. We study one, then the other, and require the third to be true in only part of the dynamics.

What is missing is *radiation reaction.* As charges accelerate, they radiate. This radiation carries energy away from the system. This, then means that a *counterforce* must be exerted on the charges when we try to accelerate them that *damps* charge oscillations.

At last the folly of our ways is apparent. Our blind insistence that only *retarded* fields are meaningful (so that we can imagine the fields to be zero up to some time and then start moving a charge, which subsequently radiates) has left us with only one charge that can produce the field that produces the force that damps applied external forces — the charge itself that is radiating. No other charge produces a field that can act on this charge "in time". We have invented the most sublime of violations of Newton's laws – an object that lift's itself up by its own bootstraps, an *Aristotelian* object that might even be able to come to rest *on its own* in the absence of external forces.

Clearly we must investigate radiation reaction as a *self–force* acting on an electron due to its own radiation field, and see if it is possible to salvage anything like a Newtonian description of even classical dynamics. We already know that Larmor radiation plus stable atoms spells trouble for Newton, but Newton still *works* classically, doesn't it?

Let's take a look. Uh–oh, you say. Wasn't the, well, wasn't *everything* singular on a point charge? Won't we get infinities at every turn? How will we realize finite results from infinite fields, potentials, self-energies, and so on?

Yes! I cry with glee. *That's* the *problem.* Finally we will learn how to take a singular field, a singular charge, and infinite energy, and make a physically realized (almost) radiation reaction force out of it.

Chapter 19

Radiation Reaction

19.1 The Death of Classical Physics

Thus far we have learned how to solve two kinds of problems. *Either* the fields were assumed to be given, in which case the relativistic Lorentz force law yielded covariant equations of motion for a point charged massive particle interacting with these fields *or* the trajectory of a charged, point particle was given and the fields radiated by this particle were determined.

This, however, was clearly not enough, or at least was not consistent. That is because (as a few simple mental problems will show) each of these processes is only *half* of an interaction — a complete, consistent field theory would include the *self–consistent* interaction of a charged particle with the field in its vicinity, or better yet, the self-consistent interaction of *all* particles and fields. We need to be able to calculate the *total* field (including the radiated field) at the position of any given point charge. Some of that field is due to the charge itself and some is due to the field produced by the other charges. But we do not know how to do this, really, since the one will affect the other, and there are clearly infinities present.

This sort of problem can also lead to Newtonian paradoxes, paradoxes that smack of the resurrection of Aristotelian dynamics. To see this, let us assume (non–physically) that we have a Universe consisting of a single point charge orbiting around an uncharged gravitational mass (or some other force center that causes the charge to move in a bound orbit). In that case, the point charge must (according to the laws of electrodynamics that we have thus far deduced) *radiate energy and momentum into the electromagnetic field.*

As it accelerates, it must radiate. As it radiates, energy and momentum must be carried away from the point particle to "infinity". The particle must therefore decrease its total energy. If the particle is bound in an attractive, negative potential well, the only way that total energy can be conserved is if its total energy decreases. The particle must therefore spiral inwards the center, converting its potential energy into radiative energy in the field, until it reaches the potential minimum and comes to rest.

There is only one difficulty with this picture. There is only one charged particle in the Universe, and it is interacting with only one attractive center. *What acts to slow the particle down?*

This is a non–question, of course – a thought experiment designed to help us understand where our equations of motion and classical picture are incomplete or inconsistent. The real universe has *many* charged particles, and they are all constantly interacting with all the *other* charged particles that lie within the "event horizon" of an event relative to the time of the big bang, which is the set of the most distant events in space–time in the past and in the future that can interact with the current event on the world line of each particle. It is the edge of the "black hole" that surrounds us[1].

However, in our simplied Universe this question is very real. We have systematically rid ourselves of the fields of all the other particles, so now we must find a field *based on the particle itself* that yields the necessary "radiation reaction" force to balance the energy–momentum conservation equations. This approach will have many unsatisfactory aspects, but it works.

First, when will radiation reaction become important? When the energy radiated by a particle is a reasonable fraction of the total relevant energy E_0 of the particle under consideration. That is

$$E_{\mathrm{rad}} \sim \frac{2}{3c}\frac{e^2a^2T}{4\pi\epsilon_0 c^2} \tag{19.1}$$

[1]It is interesting to meditate upon the fact that *your* event horizon and *my* event horizon are not coincident, which leads in turn to an interesting problem with logical positivism.

where a is the total (e.g. centripetal) acceleration and T is the period of the orbit associated with E_0 or the time a uniform acceleration is applied. If $E_{\rm rad} << E_0$ then we can neglect radiation reaction.

As before, if a particle is uniformly (linearly) accelerated for a time τ_r, then we can *neglect* radiation reaction when

$$E_0 \sim m(a\tau_r)^2 \gg \frac{2}{3c}\frac{e^2a^2\tau_r}{4\pi\epsilon_0 c^2} \tag{19.2}$$

Radiation reaction is thus only *significant* when the opposite is true, when:

$$\begin{aligned} \tau_r &\sim \frac{2}{3c}\frac{e^2}{4\pi\epsilon_0 mc^2} \\ &\sim \frac{2}{3}r_e/c = \frac{2}{3}\tau_e \end{aligned} \tag{19.3}$$

Only if $\tau_r \sim \tau_e$ and a is large will radiation reaction be appreciable. For electrons this time is around 10^{-23} seconds. This was the situation we examined before for linear accelerators and electron–positron anihillation. Only in the latter case is radiation reaction likely.

The second case to consider is where the acceleration is centripetal. Then the potential and kinetic energy are commensurate in magnitude (virial theorem) and

$$E_0 \sim m\omega_0^2 d^2 \tag{19.4}$$

where $a \sim \omega_0^2 d$ and $\tau_r \sim 1/\omega_0$. As before, we can neglect radiation reaction if

$$m\omega_0^2 d^2 \gg \frac{2}{3c}\frac{e^2\omega_0^4 d^2}{4\pi\epsilon_0 c^2\omega_0} = \omega_0 d^2 \frac{2}{3c}\frac{e^2}{4\pi\epsilon_0 c^2} \tag{19.5}$$

Radiation reaction is thus again significant *per cycle* only if

$$\omega_0\tau_r \sim 1 \tag{19.6}$$

(ignoring factors of order one) where τ_r is given above – another way of saying the same thing. ω_0^{-1} is (within irrelevant factor of 2π and $\frac{2}{3}$) the time associated with the motion, so only if this timescale corresponds to $\tau_r \approx \tau_e$ will radiation reaction be significant.

So far, our results are just a restatement of those we obtained discussing Larmor radiation except that we are going to be more interested in electrons in atomic scale periodic orbits rather than accelerators. Electrons in an atomic orbit would be *constantly* accelerating, so any small loss per cycle is summed over many cycles. A bit of very simple order-of-magnitude arithmetic will show you that radiative power loss need not be negligible as a *rate* compared to *human* timescales when ω_0^{-1} is very small (e.g. order of 10^{-15} seconds for e.g. optical frequency radiation). Charged particles (especially electrons) that move in a circle at a high enough (angular) speed do indeed radiate a significant fraction of their energy *per second* when the loss is summed over many cycles. The loss per cycle may be small, but it adds up inexorably.

How do we evaluate this "radiation reaction force" that has no obvious physical source in the equations that remain? The easy way is: try to balance energy (and momentum etc) and add a *radiation reaction force* to account for the "missing energy". This was the approach taken by Abraham and Lorentz many moons ago.

19.2 Radiation Reaction and Energy Conservation

We know that

$$\boldsymbol{F}_{\text{tot}} = m\dot{\boldsymbol{v}} \tag{19.7}$$

is (nonrelativistic) Newton's 2nd Law for a charged particle being accelerated by a (for the moment, non–electromagnetic) given external force. The work energy theorem dictates how fast the particle can gain kinetic energy if this is the only force acting.

However, at the same time it is being acted on by the external force (and is accelerating), it is also *radiating power away* at the total rate:

$$\begin{aligned} P(t) &= \frac{2}{3c}\frac{e^2}{4\pi\epsilon_0 c^2}\dot{\boldsymbol{v}}^2 \\ &= \frac{2}{3}\frac{mr_e}{c}\dot{\boldsymbol{v}}^2 \\ &= m\tau_r\dot{\boldsymbol{v}}^2 \end{aligned} \tag{19.8}$$

(the Larmor formula). These are the two pieces we've thus far treated independently, neglecting the one to obtain the other.

However, in order for Newton's law to correctly lead to the conservation of energy, the work done by the external force must equal the increase in kinetic energy *plus* the energy radiated into the field. Energy conservation for this system states that:

$$W_{\rm ext} = \Delta E_e + \Delta E_f \tag{19.9}$$

or the total work done by the external force must equal the change in the total energy of the charged particle (electron) plus the energy that appears in the field. If we rearrange this to:

$$W_{\rm ext} - \Delta E_f = \Delta E_e \tag{19.10}$$

and consider the electron only, we are forced to conclude that there must be *another* force acting on the electron, one where the total work done by the force decreases the change in energy of the electron and places the energy into the radiated field. We call that force $\boldsymbol{F}_{\rm rad}$, the *radiation reaction force.*

Thus (rewriting Newton's second law in terms of this force):

$$\begin{aligned} \boldsymbol{F}_{\rm ext} + \boldsymbol{F}_{\rm rad} &= m\dot{\boldsymbol{v}} \\ \boldsymbol{F}_{\rm rad} &= m\dot{\boldsymbol{v}} - \boldsymbol{F}_{\rm ext} \end{aligned} \tag{19.11}$$

defines the *radiation reaction force* that must act on the particle in order for energy conservation to make sense. The reaction force has a number of necessary or desireable properties in order for us to not get into "trouble"[2].

- We would like energy to be conserved (as indicated above), so that the energy that appears in the radiation field is balanced by the work done by the radiation reaction force (relative to the total work done by an external force that makes the charge accelerate).

- We would like this force to vanish when the external force vanishes, so that particles do not spontaneously accelerate away to infinity without an external agent acting on them.

- We would like the radiated power to be proportional to e^2, since the power and its space derivatives is proporotional to e^2 and since the force magnitude should be dependent of the sign of the charge.

[2]Trouble such as particles capable of lifting themselves up by their own metaphorical bootstraps...

- Finally, we want the force to involve the "characteristic time" τ (whereever it needs a parameter with the dimensions of time) since no other time-scaled parameters are available.

Let's start with the first of these. We want the energy radiated by some "bound" charge (one undergoing periodic motion in some orbit, say) to equal the work done by the radiation reaction force in the previous equation. Let's start by examining just the reaction force and the radiated power, then, and set the total work done by the one to equal the total energy radiated in the other, over a suitable time interval:

$$\int_{t_1}^{t_2} \boldsymbol{F}_{\rm rad} \cdot \boldsymbol{v} dt = -\int_{t_1}^{t_2} P dt = -\int_{t_1}^{t_2} m\tau_r \dot{\boldsymbol{v}} \cdot \dot{\boldsymbol{v}} dt \tag{19.12}$$

for the relation between the rates, where the minus sign indicates that the energy is removed from the system. We can integrate the right hand side by parts to obtain

$$\int_{t_1}^{t_2} \boldsymbol{F}_{\rm rad} \cdot \boldsymbol{v} dt = \int_{t_1}^{t_2} m\tau_r \ddot{\boldsymbol{v}} \cdot \boldsymbol{v} dt - m\tau_r (\dot{\boldsymbol{v}} \cdot \boldsymbol{v}) \,\big|_{t_1}^{t_2} \tag{19.13}$$

Finally, the motion is "periodic" and we only want the result over a period; we can therefore pick the end points such that $\dot{\boldsymbol{v}} \cdot \boldsymbol{v} = 0$. Thus we get

$$\int_{t_1}^{t_2} (\boldsymbol{F}_{\rm rad} - m\tau_r \ddot{\boldsymbol{v}}) \cdot \boldsymbol{v} dt = 0. \tag{19.14}$$

One (sufficient but not necessary) way to ensure that this equation be satisfied is to let

$$\boldsymbol{F}_{\rm rad} = m\tau_r \ddot{\boldsymbol{v}} \tag{19.15}$$

This turns Newton's law (corrected for radiation reaction) into

$$\begin{aligned} \boldsymbol{F}_{\rm ext} &= m\dot{\boldsymbol{v}} - \boldsymbol{F}_{\rm rad} \\ &= m(\dot{\boldsymbol{v}} - \tau_r \ddot{\boldsymbol{v}}) \end{aligned} \tag{19.16}$$

This is called the **Abraham–Lorentz equation of motion** and the radiation reaction force is called the **Abraham–Lorentz force**. It can be made relativistic be converting to proper time as usual.

Note that this is not necessarily the *only* way to satisfy the integral constraint above. Another way to satisfy it is to require that the difference be orthogonal to $\boldsymbol{v}$. Even this is too specific, though. The only thing that is *required* is that the total integral be zero, and short of decomposing the velocity trajectory in an orthogonal system and perhaps using the calculus of variations, it is not possible to make positive statements about the *necessary* form of $\boldsymbol{F}_{\rm rad}$.

This "sufficient" solution is not without problems of its own, problems that seem unlikely to go away if we choose some other "sufficient" criterion. This is apparent from the observation that they *all* lead to an equation of motion that is *third order in time*. Now, it may not seem to you (yet) that that is a disaster, but it is.

Suppose that the external force is *zero* at some instant of time $t = 0$. Then

$$\dot{\boldsymbol{v}} \approx \tau \ddot{\boldsymbol{v}} \tag{19.17}$$

or

$$\dot{\boldsymbol{v}}(t) = \boldsymbol{a}_0 e^{t/\tau} \tag{19.18}$$

where $\boldsymbol{a}_0$ is the instantaneous acceleration of the particle at $t = 0$.

Recalling that $\boldsymbol{v} \cdot \dot{\boldsymbol{v}} = 0$ at t_1 and t_2, we see that this can only be true if $\boldsymbol{a}_0 = 0$ (or we can relax this condition and pick up an additional boundary condition and work much harder to arrive at the same conclusion). Dirac had a simply lovely time with the third order equation. Before attacking it, though, let us obtain a solution that doesn't have the problems associated with it in a different (more up-front) way.

Let us note that the radiation reaction force in almost all cases will be very small compared to the external force. The external force, in addition, will generally be "slowly varying", at least on a timescale compared to $\tau_r \approx 10^{-24}$ seconds. If we assume that $\boldsymbol{F}_{\rm ext}(t)$ is *smooth* (continuously differentiable in time), slowly varying, and small enough that $\boldsymbol{F}_{\rm rad} \ll \boldsymbol{F}_{\rm ext}$ we can use what amounts to perturbation theory to determine $\boldsymbol{F}_{\rm rad}$ and obtain a second order equation of motion.

Under these circumstances, we can assume that $\boldsymbol{F}_{\rm ext} \approx m\dot{\boldsymbol{v}}$, so that:

$$\begin{aligned} \boldsymbol{F}_{\rm ext} &= m(\dot{\boldsymbol{v}} - \tau_r \ddot{\boldsymbol{v}}) \\ &\approx m\dot{\boldsymbol{v}} - \tau_r \frac{d\boldsymbol{F}_{\rm ext}}{dt} \end{aligned} \tag{19.19}$$

or

$$
\begin{aligned}
m\dot{\boldsymbol{v}} &= \boldsymbol{F}_{\text{ext}} + \tau_r \frac{d\boldsymbol{F}_{\text{ext}}}{dt} \\
&= \boldsymbol{F}_{\text{ext}} + \tau_r \left\{ \frac{\partial}{\partial t} + (\boldsymbol{v} \cdot \boldsymbol{\nabla}) \right\} \boldsymbol{F}_{\text{ext}} \qquad (19.20)
\end{aligned}
$$

This latter equation has no runaway solutions or acausal behavior as long as $\boldsymbol{F}_{\text{ext}}$ is differentiable in space and time.

We will defer the discussion of the covariant, structure free generalization of the Abraham–Lorentz derivation until later. This is because it involves the use of the field stress tensor, as does Dirac's original paper — we will discuss them at the same time.

What are these runaway solutions of the first (Abraham-Lorentz) equation of motion? Could they return to plague us when the force is *not* small and turns on *quickly*? Let's see...

19.3 Integrodifferential Equations of Motion

We seek solutions to the third order AL equation of motion that evolve into the "natural" ones when the driving force is turned off. In other words, radiation reaction must, by hypothesis, only damp the system and not drive it. Clearly even this requirement makes no sense when time reversal symmetry is considered. Once we fall into the trap of choosing retarded interaction *only*, we are sunk and anything we do to fix it will be a band–aid.

Let us introduce an "integrating factor" into the equations of motion. If we assume (quite generally) that

$$\dot{\boldsymbol{v}}(t) = e^{t/\tau_r}\boldsymbol{u}(t) \qquad (19.21)$$

where $\boldsymbol{u}(t)$ is to be determined, then the equations of motion simplify to

$$m\dot{\boldsymbol{u}} = -\frac{1}{\tau_r}e^{-t/\tau}\boldsymbol{F}(t). \qquad (19.22)$$

We can formally integrate this second equation, obtaining

$$m\dot{\boldsymbol{v}}(t) = \frac{e^{t/\tau_r}}{\tau_r}\int_t^C e^{-t'/\tau_r}\boldsymbol{F}(t')dt' \qquad (19.23)$$

The constant of integration is determined by our requirement that *no runaway solutions exist!* Note well that it is a constraint that lives in the *future* of the particle. In order to use this to find $\boldsymbol{v}(t)$, we must know the force $\boldsymbol{F}(t)$ for some time (of order τ_r) in the future! After this, the integrand is "cut off" by the decaying exponential.

This suggests that we can extend the integral to $C = \infty$ without difficulty. In the limit $\tau_r \to 0$, we recover Newton's law, as we should. To see this, let

$$s = \frac{1}{\tau_r}(t' - t) \tag{19.24}$$

so that

$$m\dot{\boldsymbol{v}}(t) = \int_0^\infty e^{-s}\boldsymbol{F}(t + \tau_r s)ds. \tag{19.25}$$

The force is assumed to be slowly varying with respect to τ (or none of this makes sense, just as was the case above) so that a Taylor series expansion converges:

$$\boldsymbol{F}(t + \tau s) = \sum_{n=0}^{\infty} \frac{(\tau_r s)^2}{n!} a \frac{d^n \boldsymbol{F}(t)}{dt^n} \tag{19.26}$$

which, upon substitution and integration over s, yields

$$m\dot{\boldsymbol{v}} = \sum_{n=0}^{\infty} \tau_r^n \frac{d^n \boldsymbol{F}}{dt^n}. \tag{19.27}$$

In the limit $\tau \to 0$ only the lowest order term survives. This is Newton's law without radiation reaction. The higher order terms are successive radiative corrections and matter only to the extent that the *force* varies in time. Note that this force obeys a "Lenz's Law" sort of behavior; when the applied force is changed (say, increased) there is an additional "force" *in the direction of the change* that acts on the particle. A particle moving in a circle has a force that changes *direction* but not magnitude. This change is (think about it) tangent to the motion and in the opposite direction. It acts to *slow the charged particle down.* Hmmmmmm.

There are two extremely annoying aspects to this otherwise noble solution. First, as we have repeatedly noted, it requires a knowledge of $\boldsymbol{F}(t)$ in the *future* of the particle to obtain its acceleration *now*. Truthfully, this isn't really a problem – obviously this is absolutely equivalent to saying that $\boldsymbol{F}(t)$ can be expanded in a Taylor series (is an analytic function). Second, (and even worse) it *responds* to a force that is completely in its future with

Figure 19.1: $\boldsymbol{F}(t)$, $\dot{\boldsymbol{v}}(t)$ and $\boldsymbol{v}(t)$ on a timescale of τ_r. Note that the particle "preaccelerates" *before* "the force gets there", whatever that means.

an acceleration *now*. It "knows" that a force is going to act on it *before* that force gets there.

Mind you, not *long* before the force gets there. About 10^{-24} seconds before (for reasonable forces). Classically this is very bad, but quantum theory fuzzes physics over a much larger time scale. This is viewed by many physicists as an excuse for not working out a consistently causal classical theory. You can make up your own mind about that, but note well that even if the integrodifferential equation had involved *past* values of the force you should have been *equally* bothered – either one makes Newton's law nonlocal in time!

Note well that we've already seen (nonlocal) integrodifferential equations in time in a somewhat similar context! Remember our derivation of of dispersion relations, in particular Kramers-Kronig? We had a kernel there that effectively sampled times in the future or past of a system's motion. This worked because we could integrate over *frequencies* with a constraint of *analyticity* – our fields were presumed fourier decomposable. Fourier transforms are, of course, infinitely continuously differentiable as long as we avoid *sharp* changes like (pure) heaviside function forces or field changes, and yes, they explicity provide a knowledge of the quantities in the future *and* past of their current values.

I personally think that this is yet another aspect of the *mistake* made by requiring that our description of electrodynamics always proceed from the past into the future with a retarded interaction. As we have seen, this is silly – one could equally well use only advanced interactions or a mix of the two and the solutions obtained for a given boundary value problem will be *identical*, where the "boundary" is now a four-volume and hence requires *future* conditions to be specified as well as the *past* conditions on a spatial three-surface bounding the four-volume.

19.4 Radiation Damping of an Oscillating Charge

The most important application of the Abraham–Lorentz force law is the radiation reaction of bound electrons in atoms as they radiate. This is the problem originally studied by Lorentz, in the context of a classical oscillator, and yes, we are returning to our discussion of dispersion but now with a *physical model* for why we expect there to be a damping term instead of a strictly phenomenological one.

To simplify life, we consider a Lorentz "atom" to be an electron on a spring with constant $k = m\omega_0^2$; a one–dimensional classical oscillator with a resonant frequency ω_0. If the oscillator is displaced from equilibrium, it radiates energy away and is simultaneously damped. This is a classical analogue of the emission of a photon by a quantum atom, which is accompanied by the atom entering a lower energy level.

The equation of motion for the electron is (from the AL force law above, integrated as described for offset times):

$$\ddot{x}(t) + \omega_0^2 \int_0^\infty e^{-s} x(t + \tau s) ds = 0 \tag{19.28}$$

where we have used Hooke's law. If we try the usual song and dance (assume that $x(t) = x_0 e^{-\alpha t}$ we get the characteristic equation

$$x_0 e^{-\alpha t} \left(\alpha^2 + \omega_0^2 \int_0^\infty e^{-(1+\alpha\tau)s} ds \right) = 0. \tag{19.29}$$

In order for the integral to exist, it must damp at infinity, so $\mathrm{Re}(1+\alpha\tau) > 0$. In that case, we get:

$$\begin{aligned}
\alpha^2 + \frac{\omega_0^2}{-(1+\alpha\tau)} \int_0^\infty e^{-x} dx &= 0 \\
\alpha^2 + \frac{\omega_0^2}{(1+\alpha\tau)} &= 0 \\
\alpha^2(1+\alpha\tau) + \omega_0^2 &= 0 \\
\tau\alpha^3 + \alpha^2 + \omega_0^2 &= 0 \\
(\tau\alpha)^3 + (\tau\alpha)^2 + (\omega_0\tau)^2 &= 0 \\
z^3 + z^2 + \omega_0^2\tau^2 &= 0 \qquad (19.30)
\end{aligned}$$

where we've defined $z = \alpha\tau$.

This is the same cubic that would arise directly from the original AL equation of motion *but* the restriction on the integral eliminates the "runaway" solutions ($\alpha = -(1+\omega_0^2\tau^2)/\tau$) at the expense of introducing preaccelerated ones. There is no point in giving the physical roots in closed form here, but you should feel free to crank up e.g. Mathematica and take a look.

If $\omega_0\tau << 1$ (which is the physically relevant range), then the first order result is

$$\alpha = \frac{\Gamma}{2} \pm i(\omega_0 + \Delta\omega) \qquad (19.31)$$

whith

$$\Gamma = \omega_0^2\tau \qquad (19.32)$$

and

$$\Delta\omega = -\frac{5}{8}\omega_0^3\tau^2. \qquad (19.33)$$

The constant Γ is the *decay constant* and the $\Delta\omega$ is the *level shift*. Note that the radiative force *both* introduces damping *and* shifts the frequency, just like it does for a classical damped oscillator. If we evaluate the electric field radiated by such an oscillator (neglecting the transient signal at the beginning) we find that the energy radiated as a function of frequency is

$$\frac{dI(\omega)}{d\omega} = I_0 \frac{\Gamma}{2\pi} \frac{1}{(\omega - \omega_0 - \Delta\omega)^2 + (\Gamma/2)^2} \qquad (19.34)$$

which is the characteristic spectrum of a *broadened, shifted resonant line*.

Figure 19.2: A typical broadened and shifted resonant line due to radiation reaction.

This concludes our discussion of the *consequences* of radiation reaction. You will note that the derivations we have seen are not particularly satisfying or consistent. Now we will examine the "best" of the derivations (Dirac's and Wheeler and Feynman's) and try to make some sense of it all.

The following sections are alas still incomplete but will be added shortly.

19.5 Dirac's Derivation of Radiation Reaction

19.6 Wheeler and Feynman's Derivation of Radiation Reaction

19.7 My Own Field-Free Derivation of Radiation Reaction